

CRANFIELD UNIVERSITY

FLORIAN BAZANT-HEGEMARK

Towards automated classification
of clinical optical coherence tomography (OCT) data
obtained from dense tissues

CRANFIELD POSTGRADUATE MEDICAL SCHOOL

PhD THESIS

CRANFIELD UNIVERSITY

CRANFIELD POSTGRADUATE MEDICAL SCHOOL

PhD THESIS

Academic Year 2007-2008

FLORIAN BAZANT-HEGEMARK

Towards automated classification
of clinical optical coherence tomography (OCT) data
obtained from dense tissues

Supervisors:

Igor Meglinski

Nicholas Stone

2008

This thesis is submitted in partial fulfilment of the requirements
for the degree of Philosophiæ Doctor

© Cranfield University 2008. All rights reserved. No part of this publication may
be reproduced without the written permission of the copyright owner.

Abstract

Cervical cancer can be prevented if its precursors are recognised. Those lesions that justify preventive treatment are currently identified using methods that suffer from delayed results, false positives and subjective judgement.

Optical coherence tomography (OCT) is a novel imaging modality that provides high-resolution backscattering data similar to ultrasonography. It could potentially provide *in vivo* and real-time imaging from within the entire cervical epithelium, where cervical cancer predominantly develops.

In this study, we used a bench top OCT system with a 1310 nm light source. It employs fibre optics and operates in the time domain. A collection of 1387 images from 212 *ex vivo* tissue samples from 199 participants requiring a histopathologic examination of the cervix has been created. Images from this collection were assessed in respect to their benefit in providing markers or evidence of early developments representative of cervical cancer.

In our images, the contrast in dense tissue is weak and specific markers that could be associated with a higher cancer risk were difficult to establish. For two reasons it was decided to use an algorithm for classifying the images: 1) Modern OCT systems acquire gigabytes of data per second which cannot be assessed in a clinically meaningful time. 2) An unsupervised classification tool can provide an objective assessment.

There is no established method for evaluating OCT images of dense tissue. A classification algorithm was designed that uses Principal Components Analysis as means of data reduction and Linear Discriminant Analysis as a classification tool. This approach does not rely on clinical markers to be designated *a priori*.

The algorithm was applied to the clinical data set to separate samples with mild from severe risk of cancer progression. The performance after leave-one-patient-out cross-validation reaches 61.5 % (sensitivity = 66.7 %, specificity = 47.3 %, kappa = 0.52). These results are not convincing enough to let OCT replace current systems as clinical tools in cervical precancer assessment. Routes for improving results are suggested.

This thesis provides a novel, generic algorithm for rapidly classifying OCT data obtained from dense tissues.

Acknowledgements

At this point I have the desire to thank those many people who have encouraged me throughout this project. These include my parents and family, friends, colleagues, and the (four!) academic supervisors who were involved in this project. Above all, I would like to acknowledge my on site supervisor Dr. Nicholas Stone, and the members of the Biophotonics Research Group where I conducted most of this project. I wish to thank those colleagues who helped me in the beginning of the project, Dr. Sammy Cheung and Dr. Marco Bonesi. As this project involved me moving from the university campus to the hospital environment, I also had the opportunity to meet clinical staff, such as Prof. Hugh Barr, who is the most outstanding character to mention. Continuous enthusiasm, support and understanding were the most essential contribution for the completion of this project.

Table of contents

Abstract.....	i
Acknowledgements	ii
Table of contents	iii
List of tables	xi
List of figures	xiii
Notation	xxiii
1 The clinical need: improving cervical cancer management	1
1.1 The disease “family”: epithelial cancer	1
1.2 The specifics of cervical cancer.....	1
1.2.1 Anatomy of the cervix uteri.....	2
1.2.2 Aetiology of cervical precancer.....	3
1.2.3 Risk factors	4
1.3 “Precancer” vs. “Cancer”	6
1.4 Improving assessment of cervical precancer	7
1.4.1 Early changes.....	7
1.4.2 Current diagnostic criteria	9
1.4.3 Diagnostic challenge.....	11
1.4.4 Logistic challenge.....	12
1.4.5 Areas of improvement	13
2 The novel diagnostic tool: optical coherence tomography	15
2.1 Introducing OCT	15
2.1.1 Description of OCT	15
2.2 Time domain, frequency domain, and swept source OCT	16
2.3 Components of an OCT system.....	18
2.3.1 Light source	18
2.3.2 Light guide.....	20
2.3.3 Probe arm.....	20
2.3.4 Specimen	20
2.3.5 Reference arm.....	21
2.3.6 Detector	22

2.4	Contrast, penetration depth and resolution in OCT.....	23
2.4.1	Autochthonic contrast enhancers.....	24
2.4.2	Chemical agents.....	24
2.4.3	Engineered contrast enhancers	24
2.5	OCT in the clinical setting.....	25
3	The research question: OCT for cervical precancer?	27
3.1	Aims and objectives	27
3.1.1	Chapter 4: Review of OCT and competitor technologies.....	27
3.1.2	Chapter 5: Review of OCT data analysis	27
3.1.3	Chapter 6: Pre-processing OCT data.....	27
3.1.4	Chapter 7: Introducing PC-fed LDA	28
3.1.5	Chapter 8: Application to clinical study.....	28
3.1.6	Chapter 9: Discussion.....	28
3.1.7	Appendix chapters	28
3.2	Summary.....	29
4	Modalities review: learning from prior work	31
4.1	Abstract.....	31
4.2	Methods: Search strategy and selection criteria	32
4.3	Results: Emerging technologies for high-resolution mapping of the cervix..	33
4.3.1	Confocal microscopy.....	34
4.3.2	Optical coherence tomography.....	37
4.3.3	IR spectroscopy	40
4.3.4	Raman spectroscopy	43
4.4	Discussion.....	46
4.4.1	Continuing research under the prospect of HPV vaccines	47
4.4.2	Acknowledgements	48
5	Analysis review: enabling quick and objective assessment of numerous measurements	49
5.1	Abstract.....	49
5.2	Introduction	50
5.2.1	Challenges with OCT	50
5.3	Rationale for this review and limitations.....	51

5.3.1	Selection	52
5.4	Bulk optical properties	54
5.4.1	Quantitative measures	54
5.4.2	Blood glucose	54
5.4.3	Atherosclerotic lesions	54
5.4.4	Brain tissue	55
5.4.5	Epithelial neoplasia.....	55
5.5	Texture and speckle analysis	56
5.5.1	Speckle patterns	57
5.5.2	Macrophage density.....	58
5.5.3	Epithelial lesions and precancers.....	58
5.6	Segmentation	59
5.6.1	Layer thickness	60
5.6.2	Ophthalmic layers.....	61
5.6.3	Segmentation of dense tissue.....	62
5.7	Classification	63
5.7.1	Glaucoma.....	64
5.7.2	Skin cancer	65
5.7.3	Breast tissue.....	65
5.8	Discussion.....	66
5.8.1	Common patterns and suggestions	66
5.8.2	Outlook	67
5.8.3	Acknowledgements	69
6	Designing a classification algorithm	71
6.1	Introduction	71
6.2	Feature extraction	73
6.2.1	Noise reduction.....	74
6.2.2	Surface recognition.....	75
6.2.3	Refining the surface vector.....	76
6.2.4	Applying surface vector correction (“wrapping”).....	77
6.2.5	Benefit	79
6.3	Optimal parameters for classification.....	81

6.3.1	Intensity normalising	81
6.3.2	Cut off axial depth	81
6.3.3	Amount of principal components	82
6.3.4	Summary	83
7	Testing the classification algorithm	85
7.1	Abstract	85
7.2	Introduction	86
7.2.1	Optical coherence tomography	86
7.2.2	Research questions	87
7.3	Methods	87
7.3.1	Measurements	87
7.3.2	Pre-processing	88
7.3.3	Data analysis	89
7.4	Results	90
7.4.1	Part 1: Pre-processing	91
7.4.2	Part 2: Classification	94
7.5	Discussion	96
7.5.1	Pre-processing	96
7.5.2	Classification	99
7.5.3	Acknowledgements	100
8	Applying the algorithm to clinical data	101
8.1	Abstract	101
8.2	Introduction	102
8.2.1	Cervical cancer	102
8.2.2	Non-invasive probing	103
8.2.3	Investigation in this study	103
8.3	Materials and methods	103
8.3.1	The study protocol	103
8.3.2	Measurements	104
8.3.3	Data evaluation	105
8.4	Results	107
8.4.1	Recruitment	107

Table of contents

8.4.2	Images.....	108
8.4.3	Data evaluation.....	110
8.5	Discussion.....	116
8.5.1	Recruitment	116
8.5.2	Sample quality	117
8.5.3	Data evaluation.....	117
8.5.4	Results in context of literature.....	119
8.5.5	Potential for improvements	120
8.5.6	Communication to clinician	121
8.5.7	Acknowledgements	122
9	The meaning of it all: Discussion.....	123
9.1	A framework for understanding this thesis	123
9.1.1	General approach.....	123
9.1.2	Specificities in this thesis	124
9.2	The findings in detail.....	127
9.2.1	Modalities for assessing cervical precancer	127
9.2.2	OCT data evaluation.....	128
9.2.3	Principal components-fed linear discriminant analysis.....	128
9.2.4	PC-fed LDA for analysis of epithelial precancer	129
9.3	Outlook.....	130
9.4	Suggestions to continue from this work	131
9.5	Summary and final words.....	133
	References	135
	Appendix A - Defining “non-invasive” and “non-destructive”.....	149
	Appendix A.1 - Non-invasive.....	149
	Appendix A.2 - Minimally invasive.....	149
	Appendix A.3 - Invasive.....	150
	Appendix A.4 - Non-destructive, uncomfortable	150
	Appendix A.5 - Invasiveness in examinations of the cervix	150
	Appendix B - Optimal parameters for PC-fed LDA of dense tissue OCT data	151
	Appendix B.1 - Pre-processing.....	151
	Appendix B.2 - Results.....	153

Appendix B.2.1 - Intensity normalisation vs. histogram equalisation.....	153
Appendix B.2.2 - Effect of threshold value.....	154
Appendix B.2.3 - Amount of principal components	156
Appendix B.3 - Discussion and suggestions	157
Appendix B.3.1 - Threshold for binary mask.....	157
Appendix B.3.2 - Median correction of surface misalignment	157
Appendix B.3.3 - Axial cut-off.....	157
Appendix B.3.4 - Amount of PCs	158
Appendix B.3.5 - Outliers	158
Appendix C - Ethics files	159
Appendix C.1 - Study protocol.....	159
Appendix C.2 - Addendum to standard clinic invitation letter	171
Appendix C.3 - Study participant information letter.....	172
Appendix C.4 - Consent form	175
Appendix D - PC-fed LDA for OCT in dermatology.....	177
Appendix D.1 - Abstract.....	177
Appendix D.2 - Introduction	178
Appendix D.2.1 - Port-Wine Stains.....	178
Appendix D.2.2 - Optical coherence tomography.....	179
Appendix D.2.3 - Research questions	181
Appendix D.3 - Methods	181
Appendix D.3.1 - Synthesised images.....	181
Appendix D.3.2 - Measurements.....	183
Appendix D.3.3 - Pre-processing	183
Appendix D.3.4 - Data analysis.....	185
Appendix D.4 - Results	186
Appendix D.5 - Discussion.....	187
Appendix D.5.1 - Procedure and limitations	187
Appendix D.5.2 - Summary and conclusions.....	188
Appendix D.6 - Acknowledgements	189
Appendix D.7 - Additional references.....	189
Appendix E - Publications.....	191

Appendix E.1 - Peer reviewed.....	191
Appendix E.2 - Non peer-reviewed.....	191
Appendix E.3 - Book chapters / Uploads	192
Appendix E.4 - Enclosed copies.....	193
Appendix E.4.1 - Chapters in this thesis	193
Appendix E.4.2 - Other	193

List of tables

Table 1.1: Management of cervical precancer in the UK (National Health Service Cervical Screening Programme, NHSCSP) involves various levels. It will take several decades before HPV vaccination can make the NHSCSP redundant.	11
Table 2.1: Clinical features which render themselves to assessment with OCT.....	25
Table 3.1: Summary of the hypotheses investigated in this thesis, methods to test them and rejection criteria.	29
Table 4.1: Overview of modalities discussed in this review.	33
Table 5.1: Articles employing feature recognition or automated steps for analysis of OCT data of dense tissues. ROI...region of interest	53
Table 5.2: Overview of approaches for classification on the basis of texture.....	57
Table 5.3: Comparison of automated tissue layer recognition approaches.	60
Table 6.1: Overview of approaches for classification on the basis of texture.....	71
Table 6.2: Pre-processing parameters which will be considered in subsequent classification.	83
Table 7.1: Comparison for classification of 3 A-scan representations: The Fourier Transform, the Amplitude of the FT, and time domain data. The dataset consisted of 3 groups of 10 images each (9,000 A-scans). Scores are percentages correctly classified by the algorithm, after LOSO cross-validation.	92
Table 7.2: Training performance (TP) and cross-validation (CV) of a dataset consisting of 9 groups; 7 groups of 30 images each and 2 groups of 60 images each (330 representative measures). Scores are percentages correctly classified by the algorithm, before and after cross-validation. The highest scores are bold.	93
Table 7.3: Classification rates after LOGO cross-validation. Training models were created for pairs of similar images (2×30 images for a pair).....	95
Table 7.4: Classification rates after LOGO cross-validation. Training models were created for sets of pairs of similar images (2×30 images for a pair).	95
Table 8.1: Pre-processing and group composition for the 4 data evaluation arms.....	106
Table 8.2: Basis for sensitivity, specificity, and κ value calculation as used in the literature (left table), and as applied from group assignments in this study (right table).	107

Table 8.3: Composition of the 4 groups which are used to form the 10 models, and total numbers of selected/excluded samples and patients. The dashed box encircles the samples used for the “Entire” sets (in the following graphs designated by the symbols \diamond , \square), and the grey shaded box marks the samples used for the “LLETZ” sets (symbols \triangle , \times)..... 110

Table 8.4: Number of images and group sizes of the 10 TP models..... 112

Table 8.5: Number of patients, and therefore cross-validations, of the 10 LOPO CV models. The amount of images used for cross-validation is shown in Table 8.4. 114

Table 8.6: Comparison of highest κ values for “Entire” set and LLETZ set with literature..... 119

Table 8.7: Details to references in Table 8.6..... 119

Table 8.8: Maximum performances and dependencies on PCs..... 120

Table 9.1: Summary of the hypotheses investigated in this thesis, findings and conclusions. 133

Table B.1: Visual impact of histogram equalisation and median smoothing of the recognised surface. The circles shall help spotting some of the differences. The surface recognition was done by using a 0.95 quantile binary threshold filter. ... 152

Table D.1: Desired classification performance in designed model. Vessel combinations are shown in Figure D.1. 183

List of figures

Figure 1.1: Artist impression of the cervical canal (coronal cross-section) and frontal view (modified after [1]).	2
Figure 1.2: (A) artist impression of the uterus plus cervix (mid-sagittal cross-section) with zoom into the SCJ (arbitrary cross-section). The cervix measures a few centimetres in diameter. (B) 3 D artist impression of the SCJ (modified after [2] and [3]). The thickness of the layer of squamous epithelium is typically a few 100 μm .	3
Figure 1.3: Artist impressions of key stages in cervical cancer: (A) asymptomatic dysplasia, (B) asymptomatic and curable, non-invasive cancer, (C) symptomatic cancer with poor prognosis (modified after [1]).	3
Figure 1.4: Connections between several risk factors for acquiring HPV which may lead to cervical cancer. Three key factors are marked: The factor in a dashed line (antiviral treatment) is currently not an available option. Two factors with bold lines: † Persistence of a high risk strain is a target for cancer prevention. ‡ The vaccines' mechanism of protection is providing prophylactic immunisation against HPV 16/18.	4
Figure 1.5: The role of HPV in cervical precancer. This scheme is valid for the vast majority of cervical cancers. As long as the basal lamina has not been dissolved and penetrated, these lesions have a very good diagnosis for almost 100 % curative rate.	5
Figure 1.6: Top: Schematic of CIN stages during cancer progression. Presence of CIN2 justifies preventive treatment. Bottom: Successful management aims to prevent developing high-grade lesions and allows observation of low-grade lesions. The overlaid OCT images are not to scale (modified after [5]).	6
Figure 1.7: Scheme of CIN progression. (A) healthy squamous epithelium, (B) CIN1, (C) CIN1, (D) CIN2, (E) CIN2 progressed to cancer, (F,G) CIN3 progressed to cancer, (H) cancer transgressed basal membrane. It shall be emphasised that cancer may occur at any stage. The thickness of the epithelial layer is typically a few 100 μm or around 500 μm .	8

Figure 1.8: Sections for histopathologic assessment (Columns from left to right: Normal Cervix – Cervical Dysplasia – Invasive Cervical Carcinoma). In the source of this figure, several different molecular markers were compared with the standard H&E (row ‘a’). p16^{INK4a}, p15^{INK4b}, and p14^{ARF} (rows ‘b’-‘d’) are tumour suppressor proteins that play a role in cell cycle regulation. Their presence in dysplastic cells is explained by the fact that mechanisms of tumour suppression are being initiated without success. In the left column, the “brick layer” architecture of the normal epithelium is visible. Progressed CIN is displayed in the middle column and shows the increase in nuclear material, evident through the accumulation of nucleophile markers in the (row ‘a’), or suppressor markers (rows ‘b’-‘d’). The thickness of the epithelial layer is around 500 μm (modified after [14]). 10

Figure 1.9: Drawing of a LLETZ procedure which preserves the excised tissue and allows subsequent, histopathologic margin assessment. 13

Figure 2.1: Schematic working principle of OCT: Adjacent back-scattering profiles (“A-scans”) form a tomogram (“B-scan”). Sequential B-scans would form a volume. 16

Figure 2.2: Simple Michelson interferometer, and the general setup for an OCT system. The optical delay line (OD) is the reference arm; the object is placed under the probe arm. BS...beam splitter, DET...detector, LCS...low coherent source, OD...optical delay, Proc...processing, Disp...display, Arch...archiving, Eval...evaluation. 16

Figure 2.3: Differences of TD, FD, and SSOCT when obtaining the A-scan profile. The general layout is described in Figure 2.2. The relevant alterations are surrounded by a dashed rectangle. CAM...camera, DET...detector, DG...diffraction grating, LCS...low coherent source, SS...swept source. 17

Figure 2.4: Axial resolution – long and short coherence length. In OCT, a shorter coherence length, per convention defined at the full width at half maximum (FWHM), correlates with higher resolution. l_c ... coherence length..... 19

Figure 2.5: Left side: Lateral resolution – scheme of a focussed beam. b ...depth of focus, Δx ...lateral resolution, l_c ...axial resolution (decoupled). Right side: The highest resolution is only achieved at the focal point, where the coherence volume is minimal. This should lie below the tissue surface (modified after [47]). 19

Figure 2.6: Left: The general OCT setup as shown in Figure 2.2. Right: A more specific drawing of the probe arm (surrounded by dashed rectangle). It shows the beam guiding at the sample arm and the movement of the specimen using a precision stage. In either case, the movements are communicated with the processing unit. 20

Figure 2.7: Left: Layout and optical path of a rapid scanning optical delay line (RSOD). The scanning mirror is mounted on a galvanometer. Right: the RSOD implemented into an OCT system (modified after [47])...... 21

Figure 2.8: Trade-offs for information per time. FOV...field of view; SNR...signal-to-noise ratio; FPS...frames per second. 22

Figure 2.9: Left: The general OCT setup as shown in Figure 2.2. Right: heterodyne detection as used in this study. SLD...super luminescent diode..... 23

Figure 2.10: OCT, ultrasound, and confocal microscopy put into contrast in terms of their penetration depth and typical resolution (modified after [47]). 23

Figure 4.1: CM *en face* images of biopsy samples: Normal tissue at 50 μm below the surface (b), an intermediate layer (c), and close to the basement membrane (d), and of CIN3 tissue at the equivalent distances (f-h). (a) and (e) show an overlay of an automated segmentation algorithm. Reprinted with kind permission [78]. 35

Figure 4.2: OCT images of ectocervix with apparent layer-structure (left) and less vertically structured (right) ectocervix. The bar measures 1 mm..... 39

Figure 4.3: Left side: Mean spectra of different tissue types. These can serve as a reference for unknown spectra (multivariate models are more complex). Right side: Volumetric presentation of consecutive IR maps. The colour of each voxel represents the classification likelihood for a type of cervical tissue. The volume measures $1,400 \times 1,400 \times 16 \mu\text{m}^3$. Modified and reprinted with kind permission [128]. 43

Figure 5.1: Left image: B-scan of an oesophageal biopsy specimen, of about $2 \times 2 \text{ mm}^2$ size, obtained *ex vivo* by a 7.5 m resolution (in air) frequency domain system operating at 1300 nm (Michelson Diagnostics EX1301 OCT microscope, Michelson Diagnostics Ltd.). The faint tissue layers make it difficult to decide whether clear layers are amiss due to weak contrast or because of the clinical presentation of the disease. Right image: En face sub-surface section, viewed from top, of a volumetric data set composed of 500 such B-scans. The improved

representation as opposed to 2 D images is evident. The raw data requires several gigabytes of memory. The clinical requirement for this technique would be to contribute to diagnosis, and determine whether the patient should receive cancer treatment. N.B. the weak contrast of this Figure is not due to reproduction..... 50

Figure 5.2: Scheme of landmark developments for OCT. TD: Time domain OCT, FD: Frequency domain OCT; 3 D: Volumetric OCT. The present challenge is automated data evaluation in near real-time..... 68

Figure 6.1: Scheme of surface normalisation, i.e., normalising an OCT image to the surface. The 2 steps are: a) surface recognition and b) the wrapping algorithm.... 73

Figure 6.2: Noise reduction by cutting off the noise level Top row: original noisy image, 3 D intensity matrix, and mean reflection profile. The two red lines enclose an area above the surface which is considered as noise level. Middle row: The noise level determined from the top row plus $1.5 \times$ its standard deviation. Note that the profile shows the same graph, at a different scale. Bottom row: the cut-off has been applied and the dynamic range normalised. 74

Figure 6.3: A) De-noised B-scan of a biological sample (porcine oesophagus). B) White dots mark the point of highest intensity within one A-scan (overlay over dimmed intensity image). C) White dots mark maximum rate of change in intensity within one A-scan. B & C show that the maximum intensity and derivative do not necessarily correlate with the surface. D) Surface pixels as recognised from a binary threshold mask..... 75

Figure 6.4: Surface normalising: A) Original noisy OCT image (false colour log of intensity envelope data), representative for many *ex vivo* scans: The sample surface, although somewhat straight, is not exactly parallel to the scanning direction. B) Binary image using a 0.90 quantile threshold. C) Surface recognised from B. D) Denoised image after surface normalisation using the vector recognised from the binary image for wrapping..... 76

Figure 6.5: Reflection artefacts result in incorrectly detected surface pixels..... 77

Figure 6.6: For-loops vs. vectorisation. The for-loop processes each column subsequently. Vectorisation reorders a matrix in one step. 78

Figure 6.7: Scheme of a vectorisation. 80

Figure 6.8: Visualisation of histogram equalisation, which is suggested for improving visual contrast. A) Log of intensity normalised sub-volume B) Histogram equalised data. 81

Figure 6.9: Cutting off parts of the image which contain low SNR. The line on the surface corrected images marks the border between original, shifted data and padded matrix points. 82

Figure 7.1: Representative OCT images of vegetables (after de-noising, to be fed into surface recognition). A) Celery-longitudinal B) Celery-axial C) Leek-longitudinal D) Leek-axial E) Onion F) Potato G) Mushroom stem H) Mushroom top I) Lemon J) Lime. All B-scans in this publication have a width of 3 mm. The authors would like to emphasise the somewhat similar appearance of C&D, G&H, or I&J. 88

Figure 7.2: A) Mean A-scan of original data. B) Mean A-scan of surface normalised image. C) M/SD along row of A-scans; (—) surface normalised, (---) original. 91

Figure 7.3: PCA of 3 vegetables, 10 images each. The plots show the PCs of 9000 A-scans with the 2 highest scores (determined by ANOVA) out of 25. A) Surface normalised. B) FT of A-scans (no surface normalisation). Legend: ○...celery, Δ...onion, +...mushroom. 92

Figure 7.4: PCA of 3 vegetables, 10 images each. The data points show the 2 highest scoring PCs (ANOVA) of the 30 representative measures. A) Surface normalised. B) No surface normalisation. Legend: ○...celery, Δ...onion, +...mushroom. 93

Figure 7.5: Graphical representation of exemplary confusion matrices with A) low, and B) high misclassification rates. 94

Figure 7.6: Representative OCT images of porcine samples. A) Squamous lining of the oesophagus B) tongue C) tracheae D) broncheoli E) oesophageal tissue after application of 5 % acetic acid as contrast agent F) 5 day old oesophageal tissue. Each image covers 3 mm in width. 96

Figure 7.7: A) Surface normalised B-scans: The depth profile is homogeneous across the scan width and can be well described by a mean A-scan. B) Here structures result in different A-scans, e.g. A-scans at the two white lines will be different. In this situation, a standardised mean seems to better represent the B-scan. 98

Figure 8.1: Photograph of an entire cervix sample, amputated from the uterus after a hysterectomy and placed in a specimen tray under the probe arm of the OCT system. 104

Figure 8.2: Difficulties with matching an OCT tomogram (Top) with the equivalent histopathology section (Bottom). These OCT images were created from data taken with OCT instrumentation that is capable of volumetric imaging. Of this volume, the best possible match and orientation from a series of histological slices was selected. (This comparison is credible through the particular spatial arrangement of the three cysts that is not present anywhere else in the sample.) N.B., this instrumentation was not available at the time of the study and matches from the study may be worse. (A) Top: Reconstruction from volumetric OCT dataset; Bottom: Equivalent site from histopathological sample. (B) Attempt of introducing correction factors, for refractive index mismatch and tissue shrinkage from fixation process. 105

Figure 8.3: OCT images of good quality of the different CIN groups defined by histopathology. These images are representative for those images of good quality and full penetration depth. Each image has a width of 3 mm and a resolution of ~10 μm . Columns show: A) No sign of dysplasia, B) CIN1, C) CIN2, D) CIN3. From such images it is not straight forward to establish or recognise features which are relevant to disease progress. 109

Figure 8.4: Examples for images whose useful areas had to be manually selected, or were entirely excluded from the classification due to lack of penetration depth. 109

Figure 8.5: Examples for images which had to be excluded because of the difficulties with surface recognition due to orientation, curling, or wart structures..... 110

Figure 8.6: How to read the dependency of TPs on the number of PCs in the following figures: The comet tail plots are derived from whisker plots, stretching from maximum to minimum. A bold line marks the interquartile range, and the median value is marked separately. (A) shows little variation and hence little influence from the number of PCs. (B) shows a larger variation. (C) Here, the majority of PC numbers would yield good results. (D) shows a critical situation where the maximum TP can easily be confounded by a sub-ideal number of PCs. 111

- Figure 8.7: TPs of different classification models. Group designators: N...no sign of dysplasia, 1...CIN1, 2...CIN2, 3...CIN3. The percentages are the highest values achieved with the optimal number of PCs fed into LDA. These optimal numbers of PCs are shown in the data table below. E.g., N↔1 designates a model for separating healthy from CIN1. For this model, 34 PCs gave the highest score of 77.5 % to separate data from all specimen types (◇). The same model built from LLETZ and hysterectomy samples (△) achieved 80.7 % with 25 PCs. The comet tails of each data point show the range of TP achieved when feeding between 10 and 35 PCs..... 112
- Figure 8.8: κ values for the models shown in Figure 8.7. Shown are the 40 values with highest performance in each model, and the range of 1040 models as function of the number of fed PCs, as explained in Figure 8.6. The data table shows the optimal number of PCs for these κ values, and they are mostly, but not entirely, identical to the TPs. (* Agreement interpretation after [214]). 113
- Figure 8.9: LOPO CV of different classification models. Interpretation and legend are the same as in Figure 8.7. 114
- Figure 8.10: κ values for the models shown in Figure 8.9. Shown are the 40 values with highest performance in each model, and the range of 1040 models as function of the number of fed PCs, as explained in Figure 8.6. The data table shows the optimal number of PCs for these κ values. (* Agreement interpretation after [214]). 115
- Figure 8.11: Upper left: scatter plot of the 2 strongest PCs of the entire data set, consisting of four groups (no sign of dysplasia, CIN1, CIN2, CIN3). Each B-scan descriptor is represented by a dot. By selecting mean descriptors of a group we can visualise the axes in an intuitive way. The blue circles roughly correspond to the PC dimensions (north, south, east, west, and centre). The five profiles are mean spectra from within the blue circles and can serve as an indicator that the two strongest PCs describe steepness and the presence and appearance of a plateau region. These profile features do not seem to relate with CIN stages. 118
- Figure 8.12: Two dissimilar, already surface normalised scans from the same patient, allowing discrimination. Here, no representative measure has been taken, and each

dot represents an actual A-scan. Using a representative measure, these graphs would show 2 dots. 122

Figure 9.1: Example for observational finding: A series of experimental data points serves as basis to create a rule. 123

Figure 9.2: Example for methodological finding: Introducing a tool (here the logarithm) with the aim to facilitate interpretation..... 124

Figure 9.3: Data collection in this thesis. 124

Figure 9.4: The process of converting OCT data by reduction to a hyper-dimensional data point, as developed in this thesis. 125

Figure 9.5: Establishing a linear correlation from multi-dimensional data (here illustrated by 2 dimensions)..... 126

Figure 9.6: The aim of this thesis: using OCT to predict the risk of precancer progression. 126

Figure 9.7: A correlation for the cervical OCT image collection has been performed. This was made difficult by overlapping data clouds and subsequent weak separation..... 127

Figure B.1: A) Image with recognised surface (quantile 0.95 for binary mask). B) Applied surface. C) Refined surface (sliding median of 7 pixels). D) 10 % corrupted surface vector. 151

Figure B.2: Comparison of mean classification rates for various threshold levels without artificial misalignments. Upper values (-) from intensity normalised data. Lower values (×) histogram equalised. Intensity normalised data has better classification values..... 153

Figure B.3: Classification rates for the test image set for using different threshold levels for creating the binary image which is used for surface recognition. Please note that the dotted lines are for visualisation only and shall not indicate any trend..... 154

Figure B.4: Classification rates for the test image set for various amounts of misalignment and subsequent correction. Please note that the dotted lines are for visualisation only and shall not indicate any trend or interpolation..... 154

Figure B.5: Effect of median correction on classification rate, mean and standard deviation over different threshold levels. The difference between little (▲) and strong (■) misalignment could show that 1) Larger median windows are required

when strong misalignment is expected; 2) Median windows larger than 11 do not further increase the results; 3) These effects are consistent for different threshold levels..... 155

Figure B.6: Classification rate for different cut-off values. There is little variation for cutting off 100-60 % (red circle)..... 155

Figure B.7: Classification rates for different cut-off values..... 156

Figure B.8: Average classification time for different cut-off values. Times were measured using Matlab’s tic-toc functions. The small standard deviations (10 measurements) show high confidence..... 156

Figure B.9: Left: example of little impact of the amount of PCs. Right: example of significant impact of the amount of PCs and a performance drop with more than 27 PCs..... 156

Figure D.1: Synthesised test images for assessing the fundamental possibility of classifying PWS as a function of treatment parameters. Different skin types are represented as different grey scale gradients (8 of 12 gradients are shown). The images here are shown as fed into analysis, with random depth and size variations, and with artificially created speckle noise. The classification model is tested on its ability to classify into “Treatment option 1”, “Treatment option 2”, or “Untreatable” (1, 2, or X). The expected outcome is listed in Table D.1. Right graph: Mean A-scans of these 96 synthesised images..... 182

Figure D.2: Assigning data to test set or prediction set for cross-validation: Representation of 924 permutations for a selection of 6 out of 12, shown in four rows of 231 columns. Each column has 12 pixels (6 grey and 6 black). None of the 924 columns equal each other. 182

Figure D.3: Example OCT images of PWS, of the lower cheek (left image) and mid cheek (right image)..... 184

Figure D.4: Pre-processing for clinical images. (A) Original scan. (B) The recognised surface as overlay over the image (The image is darker for visualisation purposes only). (C) Screenshot of the program for manually selecting vessel A-scans. (D) Surface normalised and cropped image..... 184

Figure D.5: The 2 strongest loadings for LDA model of the synthesised images. Note that extrema correlate with vessels depths Figure D.1. 186

Figure D.6: The 2 strongest loadings for LDA model of tissue data..... 186

Notation

AFT-M.....	descriptor for reduced A-scans (mean of frequency content)
AFT-SD	descriptor for reduced A-scans (standard deviation of frequency content)
ANN	artificial neural network
ANOVA.....	analysis of variance
Arch	archiving system
BS	beam splitter
CAM.....	camera
CIN	cervical intraepithelial neoplasia
CIN1	mild cervical intraepithelial neoplasia
CIN2	moderate cervical intraepithelial neoplasia
CIN3	severe cervical intraepithelial neoplasia, synonymous with CIS
CIS.....	carcinoma in situ (i.e., basal membrane intact and not transgressed), synonymous with CIN3
CM.....	confocal microscopy
CV.....	cross-validation
DET	detector
DG	diffraction grating
Disp.....	display system
DNA	deoxyribonucleic acid
Eval.....	evaluation
FAD	flavin adenine dinucleotide, coenzyme, in oxidised (dehydrogenated) state
FD.....	frequency domain
FDA	US Food and Drug Administration
FOV	field of view
FPS	frames per second
FT	Fourier transform
FTIR	Fourier transform infrared spectroscopy
FWHM.....	full width at half maximum
GI.....	gastrointestinal
H&E.....	Haematoxylin and Eosin (contrast agents commonly used in histopathology of cancerous tissues)
HPV	Human papilloma virus. A number suffix designates a strain, e.g. HPV16

- IRinfrared - the part of the electromagnetic spectrum with a wavelength between 1 μm and 1 mm, which can be further separated to near, mid, and far infrared
here: synonymous with infrared spectroscopy.
- κ kappa, a chance-corrected estimator for classification performance
- l_c coherence length
- λ wavelength
- LBCliquid based cytology
- LCS.....low coherent source
- LDAlinear discriminant analysis
- LLETZ.....large loop excision of the transformation zone
- LOGOleave-one-group-out
- LOOleave-one-out
- LOPO.....leave-one-patient-out
- LOSO.....leave-one-scan-out
- M/SDmean divided by the standard deviation, a data reduction method
- MD.....Mahalanobis distance
- MMP.....matrix metalloproteases
- NADH.....Nicotinamide adenine dinucleotide, coenzyme, in reduced (mono hydrogenated) state
- NCBI.....National Center for Biotechnology Information
- NHSNational Health Service, (UK public health system, introduced in 1948)
- NHSCSPUK National Health Service Cervical Screening Programme (introduced in 1987)
- NIRnear infrared - the part of the electromagnetic spectrum with a wavelength between 0.75 μm and 2.5 μm , where photons, although not visible, partake in molecular interaction similar to visible light
- OCTOptical coherence tomography
- ODoptical delay
- Pap test.....Synonymous with smear test, a sampling method that employs scraping to obtain superficial, cytological specimens. Named after Dr. G Papanicolau. Also: Pap smear, etc.
- PCprincipal component
- PCAprincipal components analysis
- PC-fed LDA...principal components fed linear discriminant analysis
- Proc.....processor

RAM.....	random access memory
RKT	rotating kernel transformation
RNA.....	ribonucleic acid
ROI	region of interest
RS	Raman spectroscopy
RSOD	rapid scanning optical delay
SCC.....	squamous cervical carcinoma, also: squamous cell carcinoma
SCJ.....	squamocolumnar junction
SGLDM	spatial grey-level dependency matrix
SLD.....	super luminescent diode
SNR	signal-to-noise ratio
SN-MSD	descriptor for reduced A-scans (M/SD of the surface normalised A-scans)
SS.....	swept source
TD.....	time domain
TP	training performance
TZ	transformation zone

1 The clinical need: improving cervical cancer management

1.1 The disease “family”: epithelial cancer

Cervical cancer is a narrowly defined clinical entity: It is confined to the uterine cervix, and the vast majority of cases arise in the epithelial layer. In contrast to non-epithelial cancers like sarcomas or leukaemia, epithelial cancers – carcinomas – constitute the majority of cancers in adults, because:

- Epithelium is the common organ lining and ubiquitous in the body.
- It is a defined interface, and a main purpose is to provide protection.
- It is continuously exposed to oncogenic agents from the environment.
- Its stem cells retain the ability to proliferate.
- If errors in proliferation persist, cancerous growth can develop.
- Exposure to oncogenes can further impede the body’s repair mechanisms.

1.2 The specifics of cervical cancer

It is difficult to categorise “cancers”, and usually a definition is based on the organ or tissue type. However, for cervical cancer, there are characteristics which let it stand out from other cancers:

- The aetiology of the majority of cases is defined and linked to the inability to shed off high risk human papilloma virus (HPV) strains.
- This knowledge has enabled the development of prevention programmes and diagnostic tools¹.
- Although the cervix is an organ whose examination requires endoscopic approaches, it has good accessibility.

Decades of experience with cervical cancer prevention management have created a reference for prevention programs.

¹ Many of the findings on the biochemical and molecular aetiology components are younger than the Papanicolaou smear test. When this was introduced, the particularities of cervical cancer might not have been known.

1.2.1 Anatomy of the cervix uteri

The cervix surrounds the os of the uterus, which in Latin simply means: The neck surrounds the mouth of the womb (Figure 1.1). It regulates the ability to conceive, keeps the uterus closed during pregnancy, can dilate to a diameter of ten centimetres during parturition, and is almost permanently remodelling.

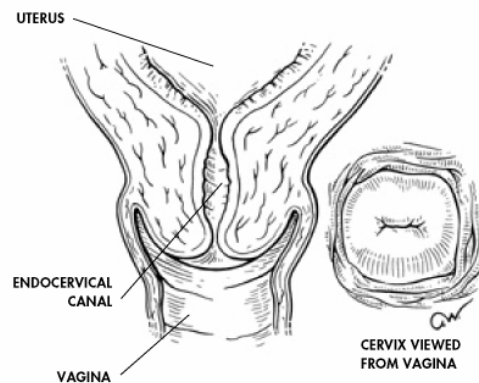


Figure 1.1: Artist impression of the cervical canal (coronal cross-section) and frontal view (modified after [1]).

Two types of epithelium are characteristic for the cervix (Figure 1.2):

- The inner canal, the endocervix, is lined with columnar epithelium, a single layer of elongated cylindrical cells (covering glands) which can secrete mucus. The viscosity of this mucus can vary and defines its permeability.
- The outer part, the ectocervix, has several layers of *squamous cells* which form a protective layer against the acidic environment of the vaginal fornix, in the manner of a regenerating shield. The squamous layers have a thickness of several hundred μm .

The visible, sharp border between the two types is called the squamocolumnar junction (SCJ) and is a prominent feature for visual (colposcopic) examination.

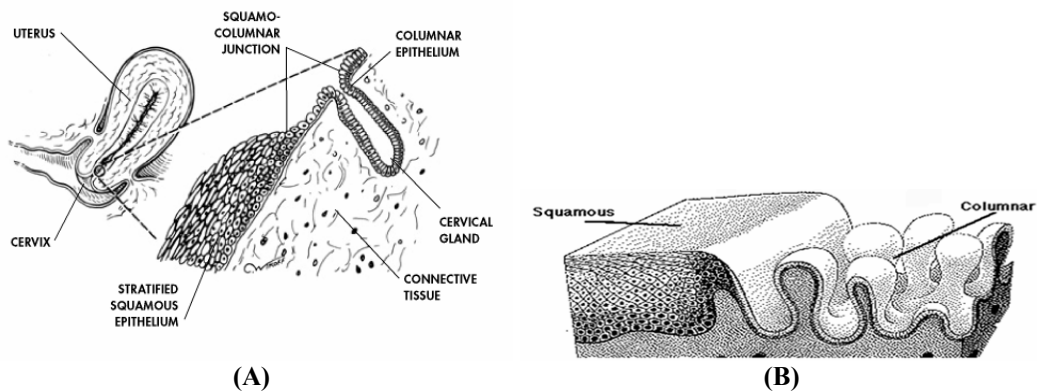


Figure 1.2: (A) artist impression of the uterus plus cervix (mid-sagittal cross-section) with zoom into the SCJ (arbitrary cross-section). The cervix measures a few centimetres in diameter. (B) 3D artist impression of the SCJ (modified after [2] and [3]). The thickness of the layer of squamous epithelium is typically a few 100 μm .

1.2.2 Aetiology of cervical precancer

Columnar epithelium can undergo metaplasia (tissue remodelling) to squamous epithelium. This is normal and happens in the so-called transformation zone (TZ). Inherently, those cells who are able to regenerate tissue and remodel are the susceptible ones for oncogenic agents. This agrees with our understanding of carcinomas.

The most prevalent form of cervical cancer is squamous cervical carcinoma. It is preceded by a long phase of preclinical disease. These early changes do not cause symptoms which would be noticeable by the patient. Precancerous dysplasia does not necessarily lead to cancer. With increasing severity of dysplasia the likelihood of this lesion to progress to cancer later in life rises (Figure 1.3).

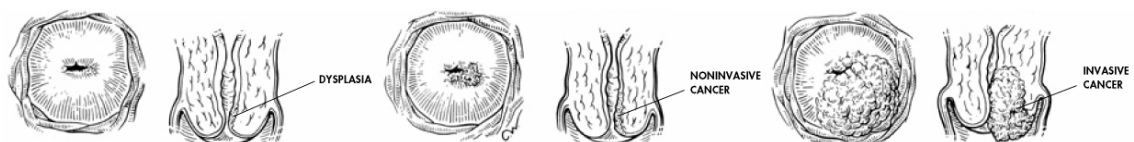


Figure 1.3: Artist impressions of key stages in cervical cancer: (A) asymptomatic dysplasia, (B) asymptomatic and curable, non-invasive cancer, (C) symptomatic cancer with poor prognosis (modified after [1]).

1.2.3 Risk factors

Knowing the risk factors is the key for prevention. HPV plays the central role in cervical cancer (Figure 1.4): most risk factors are linked to exposure, persistence, and immune response to such an infection. **It is not the HPV infection alone** that is causing cancer, as it is a common virus almost every person can carry at some point in her or his life.



Figure 1.4: Connections between several risk factors for acquiring HPV which may lead to cervical cancer. Three key factors are marked: The factor in a dashed line (antiviral treatment) is currently not an available option. Two factors with bold lines: † Persistence of a high risk strain is a target for cancer prevention. ‡ The vaccines' mechanism of protection is providing prophylactic immunisation against HPV 16/18.

Figure 1.4 shows risk factors for development of CIN. There are further risk factors for the progression from CIN to invasive disease (multiple pregnancies, “western” life style, and infections with sexually transmissible diseases, such as Chlamydia and viral infections). It is however likely those other risk factors are related to a life style that can be associated with a higher risk for a HPV infection, and that they are not necessarily directly causing cancer.

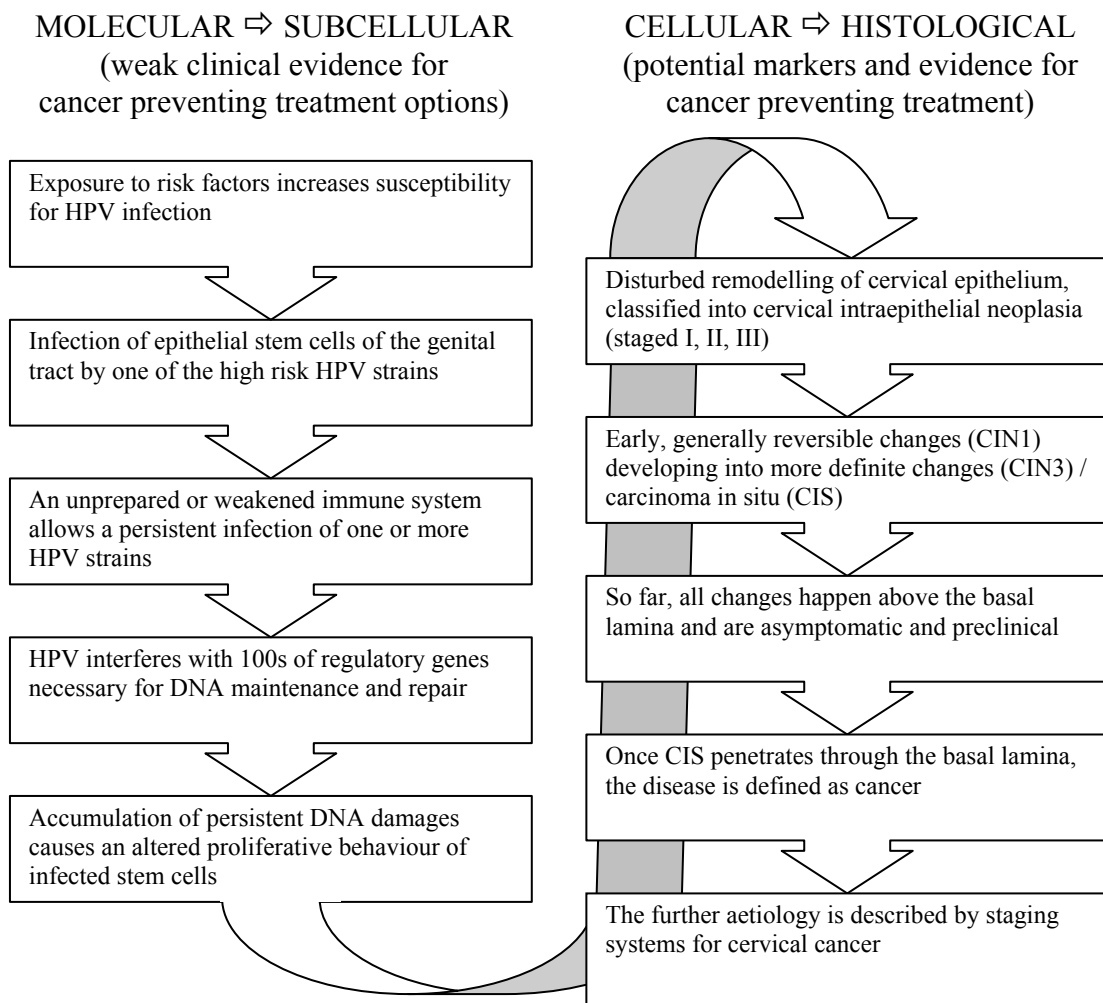


Figure 1.5: The role of HPV in cervical precancer. This scheme is valid for the vast majority of cervical cancers. As long as the basal lamina has not been dissolved and penetrated, these lesions have a very good diagnosis for almost 100 % curative rate.

Smoking, alcohol or drug abuse are believed to act as a trigger: A plausible mechanism is that in cells which would undergo apoptosis, this salvage mechanism is seized by the

HPV infection. A detailed review on various risk factors is provided elsewhere [4]. Figure 1.5 shows a cascade of steps leading to cervical cancer.

1.3 “Precancer” vs. “Cancer”

In order to understand what cervical screening is aiming at, it is important to distinguish precancer from cancer.

The key for all cancer management is prevention. Healing cancer is often impossible and usually involves a combination of chemical destruction of tissue and surgical removal of organs. There is a trade-off between the chance of cure and the harm from the clinical intervention.

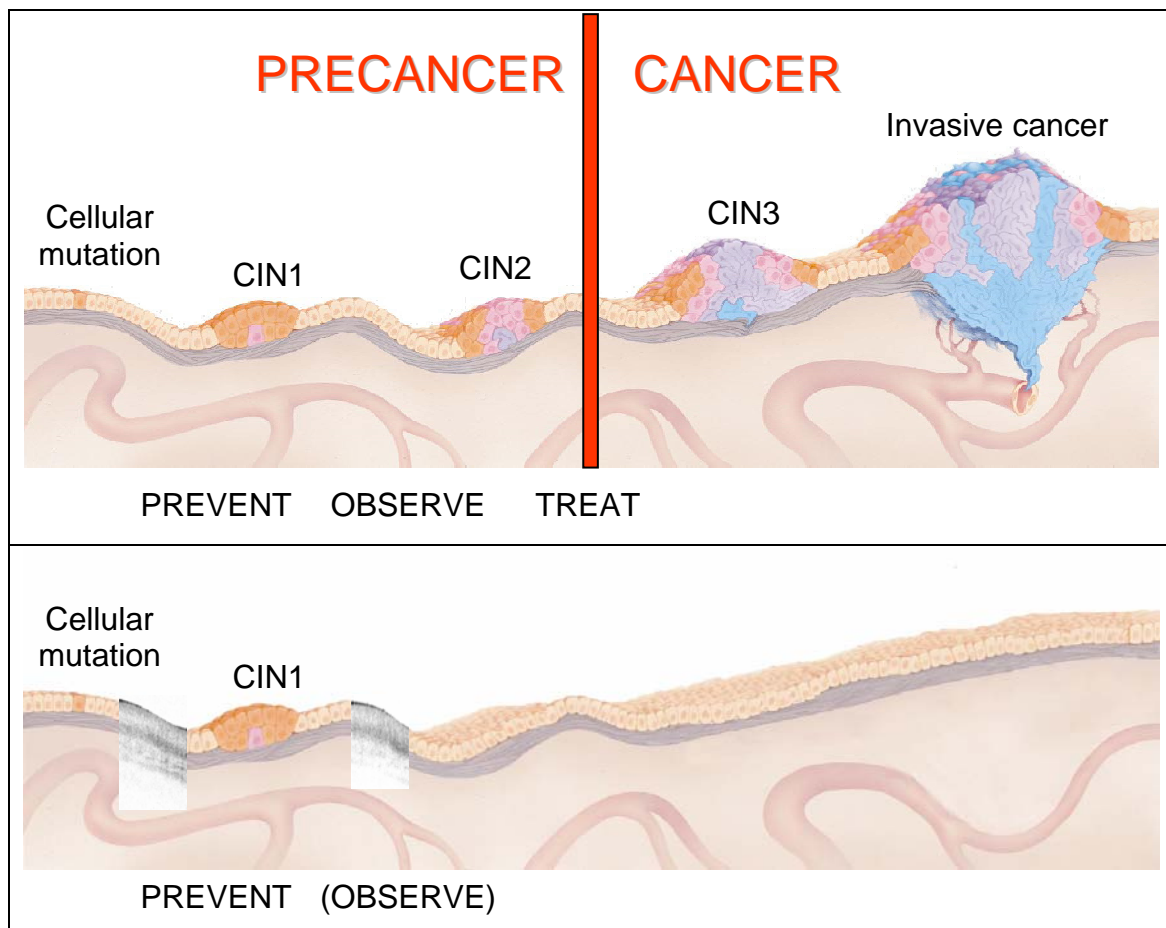


Figure 1.6: Top: Schematic of CIN stages during cancer progression. Presence of CIN2 justifies preventive treatment. Bottom: Successful management aims to prevent developing high-grade lesions and allows observation of low-grade lesions. The overlaid OCT images are not to scale (modified after [5]).

Cancer management aims therefore to shift from treating to preventing. For most cancers, risk factors have been identified or are subject of ongoing research. Since

cancer is not a deterministic disease, such risk factors can only contribute to diagnosis and prognosis (Figure 1.6).

The problem is the identification of the threshold between treatment and surveillance to justify treatment: Where shall one put the red line in Figure 1.6?

Quite a few properties of cervical cancer can help justifying why an investigation into the merit of a technology such as optical coherence tomography as early diagnostic tool is worth being undertaken.

1.4 Improving assessment of cervical precancer

Management of cervical cancer is archetypical: In those countries which have established programs for screening, early detection, and preventive treatment, cancer rates have dropped [6]. In the UK, the Cervical Screening Programme of the National Health Service (NHSCSP), which was introduced in 1998, is estimated to save 4,500 lives a year [7]. This is possible because squamous cervical carcinoma (SCC), the most prevalent form of cervical cancer², is preceded by a long phase of pre-clinical alterations, which are commonly referred to as cervical intraepithelial neoplasia (CIN), and graded into mild (CIN1), moderate (CIN2), or severe (CIN3) [8]. The following paragraphs give a short summary of the present understanding of CIN and provide the common ground for understanding the mechanisms of precancer detection.

1.4.1 Early changes

What are the characteristics of the different steps in precancer development?

Histological level: Epithelial cells originate from stem cells which are lying amongst the basal layer of cells on a supporting molecular network of collagen. The underlying connective tissue is called the basal lamina. The sophisticated regulatory system in the healthy body achieves a smooth gradient of cell maturity from the stem cell level towards the outside throughout the thickness of the epithelium (several 100 μm). The

² It is difficult to provide an accurate percentage. Depending on socio-economic factors and the screening situation, prevalence can be 98-99 %. Cervical screening programmes have indeed resulted in a reduction of SCC, and a recent publication on the UK situation in 2007 indicates that the SCC ratio has dropped to 70-80 % (Shafi 2007 [8]). This is an indicator for the success of screening programs.

development from a freshly divided unit to a mature cell can roughly be correlated with the distance from the basal layer (Figure 1.7). CIN is correlated with subcellular alterations which cause aberrations from this scheme.

Cytological level: Mature cells have a reduced nucleus-to-cytoplasm ratio and are flat (“squamous”), thereby forming a protective layer. This goes in hand with biochemical changes: glycogen accumulation due to a cease in energy demand and depletion of metabolites with short half life, e.g. RNA or phosphorylated molecules. If there are deviations in cell shape and sub cellular architecture (dyskaryosis), and if there is metabolic activity in surface layers, this indicates problems in the proliferation at the basal level. However, such changes are not exclusive to precancer, benign conditions (e.g. inflammatory processes) can show similar deviations.

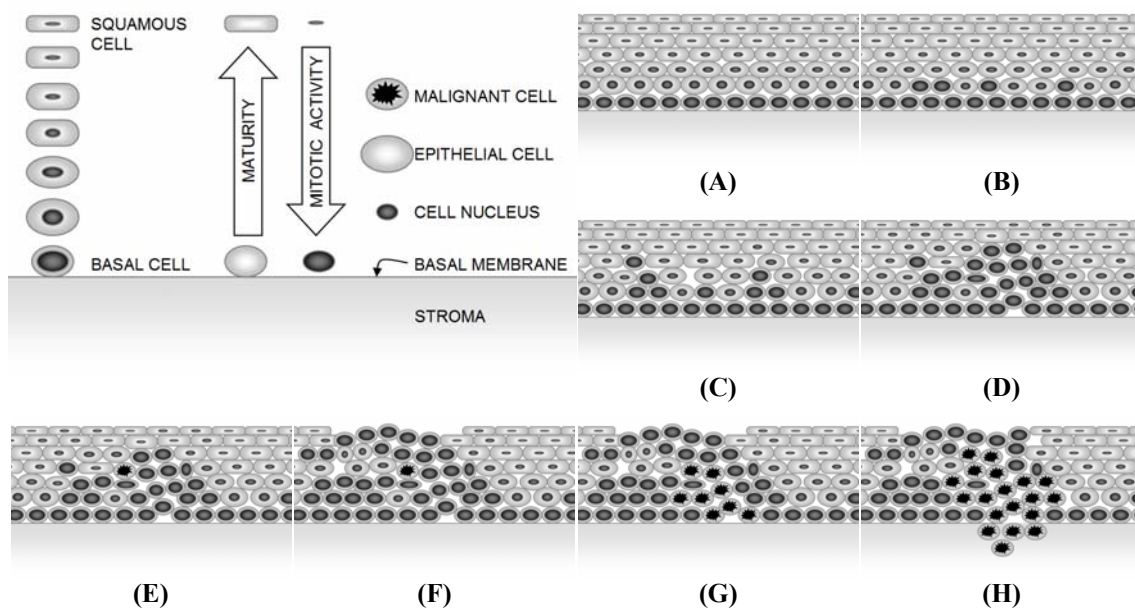


Figure 1.7: Scheme of CIN progression. (A) healthy squamous epithelium, (B) CIN1, (C) CIN1, (D) CIN2, (E) CIN2 progressed to cancer, (F,G) CIN3 progressed to cancer, (H) cancer transgressed basal membrane. It shall be emphasised that cancer may occur at any stage. The thickness of the epithelial layer is typically a few 100 μm or around 500 μm .

The lower boundary: Growth of cells below the basal lamina is believed to be facilitated if this layer has been decomposed by proteins, the matrix metalloproteases (MMP). Viewed under the microscope, this can be noticed when the usually sharp border between epithelium and underlying tissue is interrupted (Figure 1.7H). Expression levels of MMPs allow correlation with CIN staging [9], and the ability to

transgress through tissue boundaries is a feature for determining severity. Cancer prevention aims at catching cells before they gain motility [10].

1.4.2 Current diagnostic criteria

The diagnostic criteria for this are continually revised and duly published, e.g. by the British Society for Colposcopy and Cervical Pathology [11], or its American counterpart which provides an online practice recommendation [12]. The general criteria for a diagnosis are similar to other types of cancer: cellular immaturity, cellular disorganization, nuclear abnormalities, and increased mitotic activity [13]. These criteria may be timeless, but the diagnostic arsenal for establishing the presence of such criteria is subject to extensive research. It is noteworthy that – seemingly due to the difficulties in agreeing on defined markers – the current classification system in the UK is three-tiered (CIN1-3), whereas sources in the US use a two-tiered system which classifies into low-grade and high-grade lesions. Common markers to highlight nuclear material in tissue are the contrast agents hematoxylin and eosin (H&E). The exact mechanisms for these stains are not understood in their entirety. Haematoxylin is a basic/cationic agent that binds predominantly to basophilic molecules – the phosphoryl groups of nucleic acids, and also to carboxyl groups of e.g. acid proteins. It shows up in shades of violet/dark blue. Eosin is an acid agent and stains basic proteins, which are found in collagen or cytoplasm. Under the microscope it appears light red or pink. There are many different stains available, amongst which “H&E” is a dominating stain for establishing criteria for cancer, such as nucleus-to-cytoplasm ratio in tissue, under the microscope. It remains, however, unspecific to the exact types of molecules. The optical diagnostic tool investigated in this thesis is supposed to exhibit a similar(ly unspecific) contrast for nuclear material and interfaces, without the need for stains. It will therefore not aim to compete with immunohistologic stains, which provide the possibility of acting as antibodies to antigens that have been recognised as specific markers with high affinity, e.g. for the presence of a specific high-risk HPV strain. These have the potential to not only allow a better diagnosis, but maybe allow a prognosis on the progression of the precancerous lesion. An example for such a comparison is given in Figure 1.8.

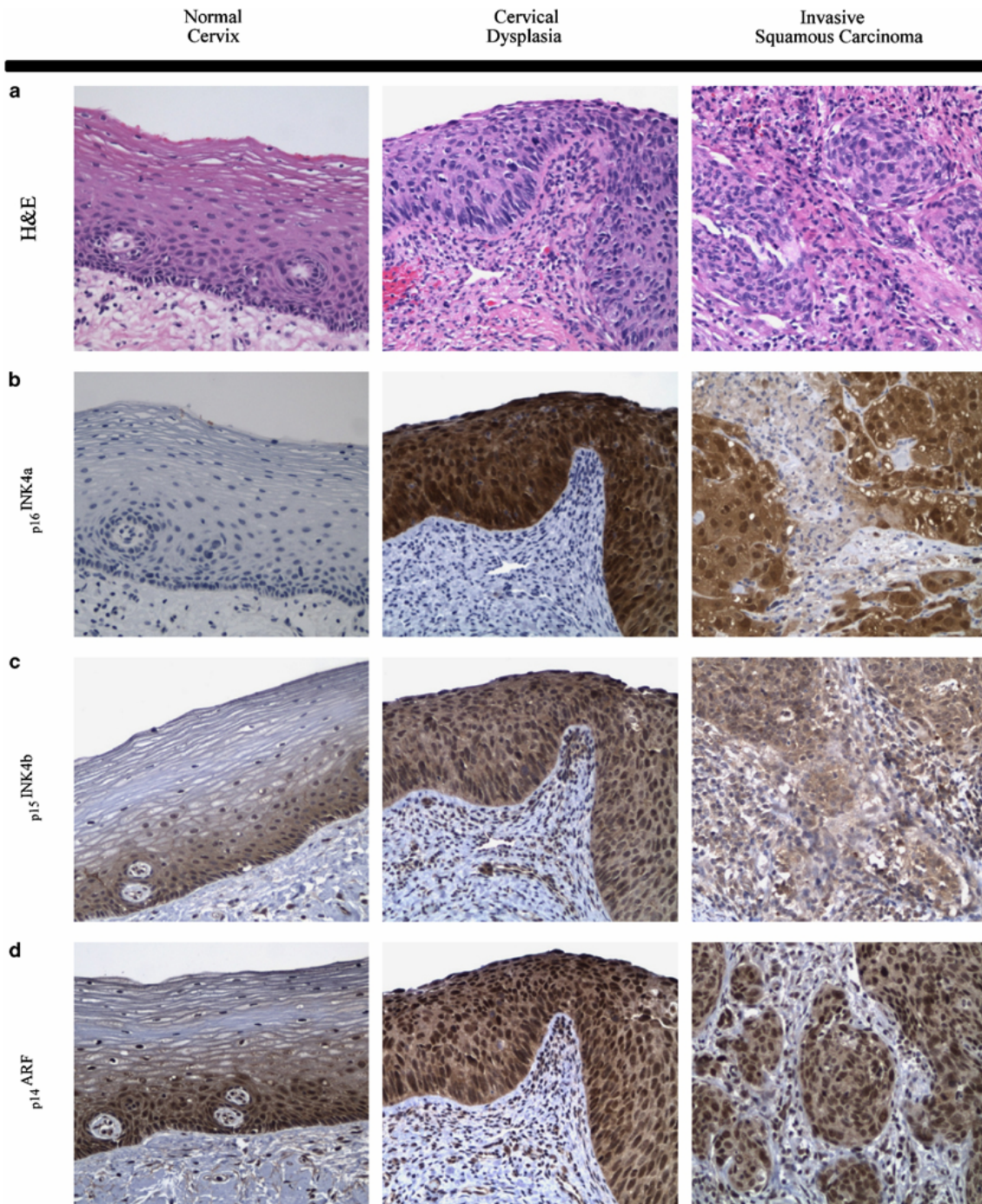


Figure 1.8: Sections for histopathologic assessment (Columns from left to right: Normal Cervix – Cervical Dysplasia – Invasive Cervical Carcinoma). In the source of this figure, several different molecular markers were compared with the standard H&E (row ‘a’). $p16^{INK4a}$, $p15^{INK4b}$, and $p14^{ARF}$ (rows ‘b’-‘d’) are tumour suppressor proteins that play a role in cell cycle regulation. Their presence in dysplastic cells is explained by the fact that mechanisms of tumour suppression are being initiated without success. In the left column, the “brick layer” architecture of the normal epithelium is visible. Progressed CIN is displayed in the middle column and shows the increase in nuclear material, evident through the accumulation of nucleophile markers in the (row ‘a’), or suppressor markers (rows ‘b’-‘d’). The thickness of the epithelial layer is around 500 μm (modified after [14]).

1.4.3 Diagnostic challenge

Progression of CIN1 to more severe stages is not definite: Rather, CIN stages serve as a marker for a likelihood of progressing to cancer, and a decreased likelihood of reverting back to normal without intervention (Figure 1.7E-G). Certain HPV strains have been shown to be involved and to be a necessity for cancer progression. A large body of evidence supports that the cancer risk is a consequence of the body's inability to develop immunity against HPV, thereby failing to rid itself of the virus. This has led to the development of HPV vaccines.

The key issue in CIN management today remains: Justifying preventive treatment, without missing a lesion and without over treating healthy patients, in as few diagnostic procedures as possible. Diagnosis in modern screening programs, as summarised in Table 1.1, is initiated by cytological examination of exfoliated cells (superficial), colposcopy (superficial), and histopathologic assessment of biopsies (sub-surface) [15]. Findings of moderate or severe dyskaryosis on a smear or consecutive events of mild abnormalities prompt a referral to colposcopy [16]. Current findings support that women should be invited for colposcopy even after only one mild event [17].

Table 1.1: Management of cervical precancer in the UK (National Health Service Cervical Screening Programme, NHSCSP) involves various levels. It will take several decades before HPV vaccination can make the NHSCSP redundant.

Level	Approach / Technology	Responsibility
Prevention	Public awareness HPV vaccination	Several bodies, media, NHS individual
Screening	Cytology ("Pap test")	Provided via NHSCSP
Early detection	Visual assessment (colposcopy) Histopathology	Part of NHSCSP, Gynaecology
Treatment	Removal of precancerous lesion	Clinician with informed consent of patient
Monitoring	Shorter intervals for suspicious lesion(s), and after treatment	Follow-up as part of NHSCSP

Of the adequate tests in 2002/3, the follow-up results for around 96.6 % were negative or borderline for disease. Further surveillance was justified in 2.1 % of cases with mild dyskaryosis. In 1.1 % of cases with moderate or severe dyskaryosis the protocol recommended preventive treatment [18]. Whilst normal and progressed CIN stages are distinguishable, false positives and false negatives are higher for distinguishing CIN1

from CIN2, and are also higher for differential diagnosis of other benign conditions (benign metaplasia and inflammation etc.) [19]-[22]. Women with mild dyskaryosis who are immediately seen in colposcopy have a higher rate of CIN confirmed by histopathology compared to women who are invited 2 years later, suggesting that regression can happen without intervention [17],[23]. However, larger intervals increase the risk of missing a developing high-grade lesion; therefore present management has to subject every mild lesion to further diagnostic procedures. Immediate colposcopy referral has a better rate of women attending the follow-up [19]. As a consequence, today's clinical practice has to accept repeated tests, anxiety of patients, and potential over treatment of a lesion [24].

1.4.4 Logistic challenge

In the UK, of 4.4 million women who are annually invited for screening, only about 3.6 million actually attend [18]. About 10 % of these tests are inadequate for evaluation, resulting in over 4 million tests performed per year, at an average cost of £38 per test, or £157 million a year [7], for an estimated 4,500 lives saved due to cervical cancer as primary cause of mortality.

It takes time to communicate these results to the screening participants, thereby causing anxiety and treatment delay: 48 % of the women are informed within 4 weeks and 74 % within 6 weeks. These figures of 2006/7 are “the highest level recorded” [25] and mean that any woman waiting up to one month for her result still has to consider herself to be better than the average. The treatment delay is tackled in the UK, by managing moderate and severe with higher priority. This short listing is however based on the subjective impression of the clinician at the time of the sampling, and inherently means a delay for women who do not seem to require urgent treatment. If a biopsy is then required for definitive diagnosis, this result reaches 25 % of women within 2 weeks and 38 % of the women within 4 weeks. This correlates with a recent study on 600 women in England and Wales which states that the median time to initiate treatment is 50 days [26]. These figures show a potential for new diagnostic tools to improve the issues of time, cost, and anxiety which are linked to precancer screening.

1.4.5 Areas of improvement

Therefore there are attempts to provide a clinical tool that is able to provide information about the tissue in near real-time, where necessary repeated over a longer surveillance time, and in a non-destructive fashion. The ultimate aim is replacing biopsies in order to monitor suspicious lesions and schedule preventive treatment [19]. More realistic is the aim of reducing and guiding better targeting of biopsies, and improving the prognostic value of the findings. This would allow a reduction of cost to healthcare and patient distress, as participants can be expected to prefer a “see-and-treat” protocol to a “diagnose-and-defer” approach [27]-[29].

A further consideration for researching new diagnostic tools is that treatment today is limited to methods which leave the tissue suitable for margin assessment, as this is relevant for prognosis and follow-up (Figure 1.9): Margin involvement is correlated with higher recurrence rates [30]-[34], specifically at the endocervical margin [35]. This margin assessment by histopathology is based on few sections (sampling a sample) and bears the possibility of misclassifying [36]. Destructive treatment may have less pregnancy-related morbidity than excision treatment [37]. If tissue can be characterised by a non-destructive tool, this provides destructive treatment options such as laser ablation, cryo surgery, or photodynamic therapy [38]-[41].

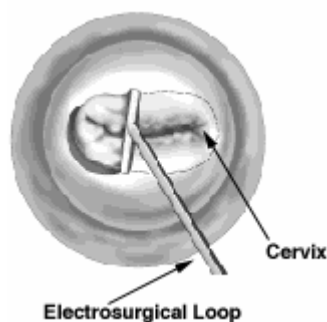


Figure 1.9: Drawing of a LLETZ procedure which preserves the excised tissue and allows subsequent, histopathologic margin assessment.

These outlined clinical needs have driven the direction of this PhD project. The following chapter outlines a possible technique to achieve these goods. Optical coherence tomography has been explored throughout this body of work, as a tool for real-time tissue imaging and potential pathological analysis in the cervix.

2 The novel diagnostic tool: optical coherence tomography

2.1 Introducing OCT

There are over 15 years worth of research publications, numerous textbooks and encyclopaedic descriptions dedicated to optical coherence tomography (OCT). In view of this it is difficult to capture an essence of this technology as part of an introduction for a thesis. In an as concise as possible way, a description of the OCT system that is relevant for the development of this thesis shall be presented. This shall provide a framework for the understanding of this technology in those aspects that are relevant for the scope of this work. Some other developments of OCT shall only be touched on, so that the system used in this work can be put into context.

2.1.1 Description of OCT³

OCT is an imaging technology based on low coherence interferometry, first described by Huang *et al.* in 1991 [42]. OCT is often compared with ultrasound by its virtue of probing for the intensity of backscattered light (“echo”) from within a sample. Agreeing with this association, the classic OCT scanning protocol consists of adjacent 1 D A-scans, which are joined to form a 2 D image or a 3 D volume (Figure 2.1).

In contrast to ultrasound, the “echo” of the incident light cannot be detected directly. The incident light beam is characterised by short “pulses” of coherent light, which are crucial for detection: Interference of this coherent part with a reference beam allows localising the path length and hence the location of the scattering with high precision.

Typical OCT systems provide a resolution of 10 μm ; “ultra-high” resolution is in the magnitude of 1 μm . When using near infrared light, penetration depths in soft tissue are in the magnitude of 1-2 millimetres. The mode of contrast in OCT is based on variations of the refractive index within a specimen. Specifically layers of different types of tissue can show up prominently on OCT images. Since the mid 1990s OCT has therefore been

³ Modified from a manuscript that was written as contribution to: Barr H, Hutchings J, Kendall C, Bazant-Hegemark F, Stone N. Endoscopic approaches to the treatment of early malignancy and their relationship to clinical outcome. Chapter 10 in: *UK Key Advances In Clinical Practice Series, Colorectal Cancer 4th edn*, 2008 (in press).

suggested to be promising for real-time optical biopsy. State of the art probe systems have dimensions of a few mm and can be readily used in endoscopy.

OCT data acquisition

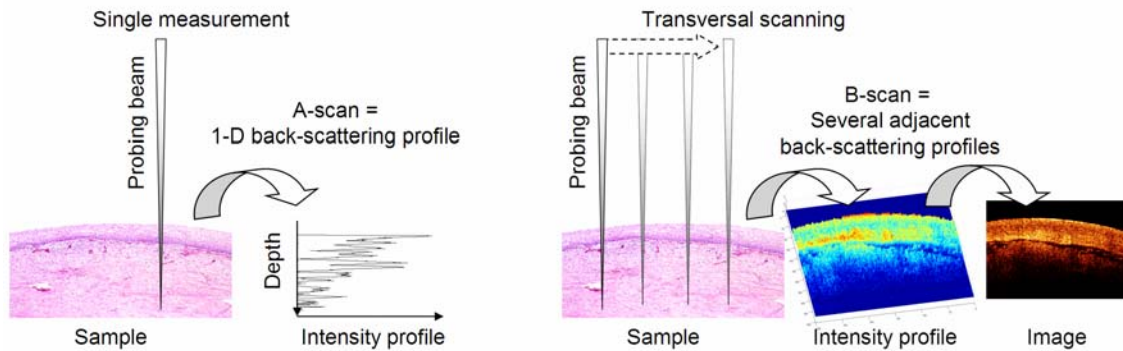


Figure 2.1: Schematic working principle of OCT: Adjacent back-scattering profiles (“A-scans”) form a tomogram (“B-scan”). Sequential B-scans would form a volume.

2.2 Time domain, frequency domain, and swept source OCT

Interference is a fundamental concept for OCT. Figure 2.2 shall serve to show the similarities between the Michelson interferometer and an OCT system.

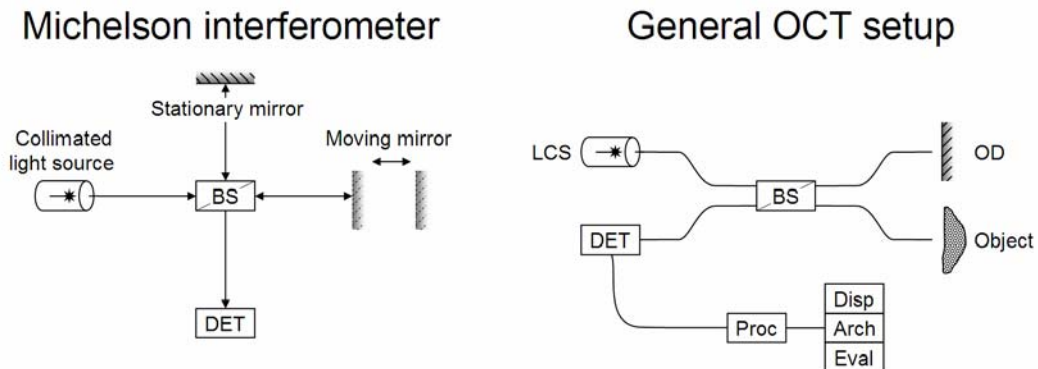


Figure 2.2: Simple Michelson interferometer, and the general setup for an OCT system. The optical delay line (OD) is the reference arm; the object is placed under the probe arm. BS...beam splitter, DET...detector, LCS...low coherent source, OD...optical delay, Proc...processing, Disp...display, Arch...archiving, Eval...evaluation.

Time domain (TD), frequency domain (FD), and swept source (SS) systems differ in the way the A-scan profile is obtained. In TDOCT, the reference mirror is moved to obtain

a sequence of reflection events along the A-scan. In FDOCT, this movement is avoided and the beam is separated into the spectral components. The Fourier transform (FT) of the obtained signal yields the intensity envelope of the A-scan. Another way to obtain the signal at different frequencies is to sweep the wavelength of the light source (SSOCT). The differences in the setup of these three methods are highlighted in Figure 2.3. Practical differences concern SNR, scanning speed, mechanical movement, cost, and linearity of the signal. For the purpose of this thesis, these differences are not significant. A recent review on detailed aspects of TDOCT, FDOCT, and SSOCT can be found elsewhere [43].

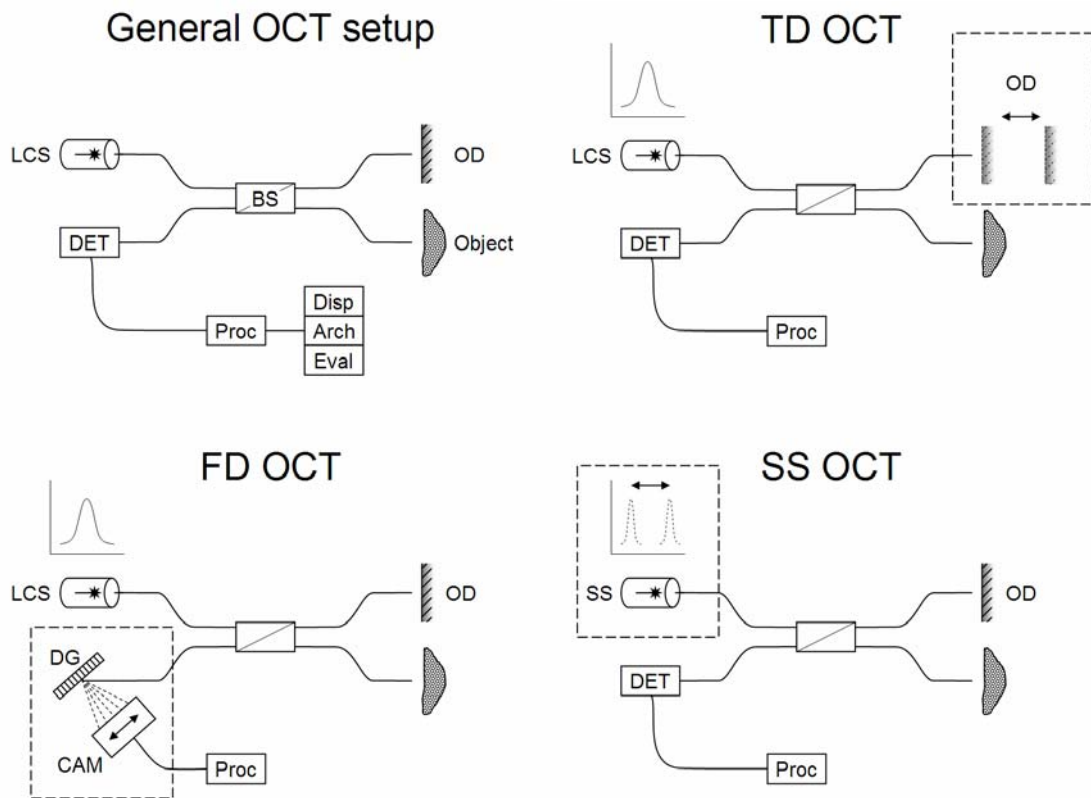


Figure 2.3: Differences of TD, FD, and SSOCT when obtaining the A-scan profile. The general layout is described in Figure 2.2. The relevant alterations are surrounded by a dashed rectangle. CAM...camera, DET...detector, DG...diffraction grating, LCS...low coherent source, SS...swept source.

The configurations in Figure 2.3 sequentially measure A-scans as a base unit. A different approach is full field OCT, where a camera detector allows parallel detection of A-scans.

2.3 Components of an OCT system

For describing a system, it seems appropriate to follow the same path in which light will travel: Starting from the light source, via sample and reference arms, and to the detector.

2.3.1 Light source

Coherent light is necessary for interferometric detection and defines the axial resolution. In order to allow a high-resolution mapping, the coherence has to be partial. Such a partial coherence is commonly described as a light pulse. The central wavelength of the coherence window is the nominal wavelength of an OCT system. The width of the side lobes defines the axial resolution and the shape of the side lobes defines the quality of the resolution (Equation 2.1 and Figure 2.4)⁴. The ideal coherence function would have a very narrow rectangular envelope where interference is possible.

$$l_c = \frac{2 \ln 2}{\pi} \left(\frac{\lambda^2}{\Delta \lambda} \right) = 0.44 \left(\frac{\lambda^2}{\Delta \lambda} \right) \quad (\text{Equ. 2.1})$$

**l_c ... coherence length, λ ...centre wavelength of the light source,
 $\Delta \lambda$... l_c at the full width at half maximum (FWHM) of the autocorrelation function.
 (after [44] and [45])**

⁴ For instance, the light source in this study (Superlum SLD-561) has a central wavelength of 1.3 μm with a FWHM ranging from 1270 to 1330 nm. The nominal resolution according to Equation 2.1 calculates as:
 $l_c = 0.44 \times (1300^2 / 60) [\text{nm}^2/\text{nm}] = 12393 \text{ nm} \cong 12.4 \mu\text{m}$.

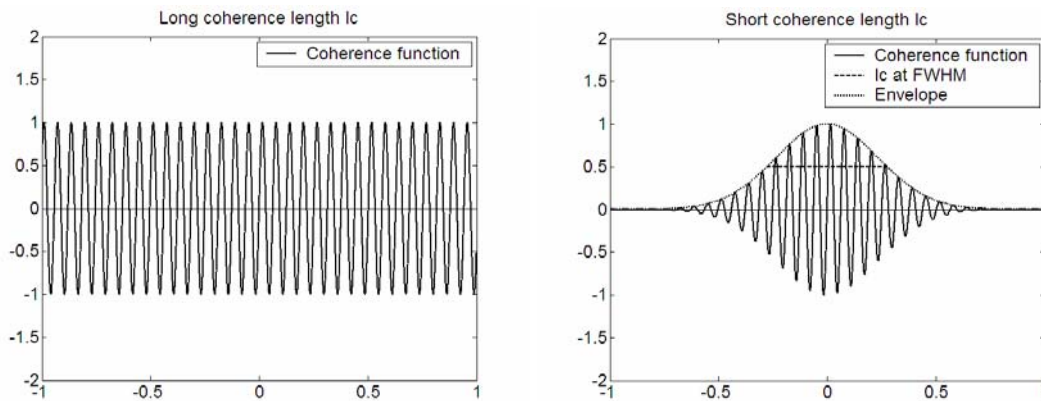


Figure 2.4: Axial resolution – long and short coherence length. In OCT, a shorter coherence length, per convention defined at the full width at half maximum (FWHM), correlates with higher resolution. l_c ... coherence length.

The lateral (transverse) is resolution decoupled from the axial resolution, and depends on the beam shape (Equations 2.2-2.3 and Figure 2.5):

$$\Delta x = \frac{4\lambda}{\pi} \left(\frac{f}{d} \right) \quad (\text{Equ. 2.2})$$

Δx ...lateral resolution, λ ...centre wavelength of the light source, f ...focal length, d ...spot size on the objective lens (after [44] and [45])

$$b = \frac{\pi \Delta x^2}{2\lambda} \quad (\text{Equ. 2.3})$$

b ...depth of focus (after [44] and [45])

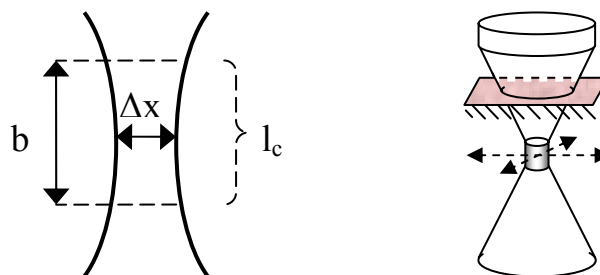


Figure 2.5: Left side: Lateral resolution – scheme of a focussed beam. b ...depth of focus, Δx ...lateral resolution, l_c ...axial resolution (decoupled). Right side: The highest resolution is only achieved at the focal point, where the coherence volume is minimal. This should lie below the tissue surface (modified after [47]).

2.3.2 Light guide

In practice, two configurations exist: bulk optics or fibre optics. Bulk optics have the disadvantage of their size requirements, but allow precision adjustments and have fewer problems with refractive index mismatch. Fibres are sensitive to polarisation but allow reaching remote sites. Specifically for endoscopic applications they are therefore mandatory.

2.3.3 Probe arm

By analogy to older ranging technologies, the backbone of an OCT dataset is the A-scan. This is a 1 D back-scattering profile. The mechanical design of the probe arm defines how the light beam is guided across the sample for a 2 D or 3 D image. For *ex vivo* systems, precision stages can move the sample relative to the light beam. A common solution employable for *in vivo* measurements is to guide the light beam using a galvanometric turning mirror (Figure 2.6).

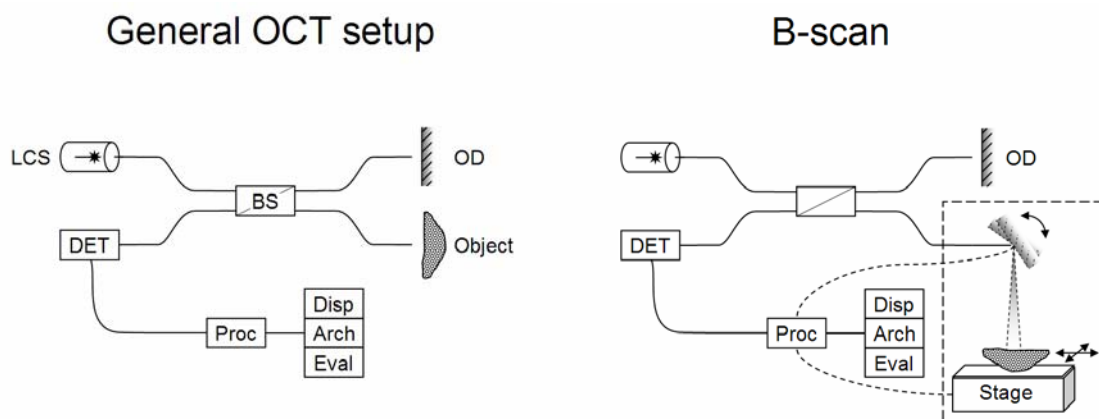


Figure 2.6: Left: The general OCT setup as shown in Figure 2.2. Right: A more specific drawing of the probe arm (surrounded by dashed rectangle). It shows the beam guiding at the sample arm and the movement of the specimen using a precision stage. In either case, the movements are communicated with the processing unit.

2.3.4 Specimen

Prior knowledge of the specimen's optical properties might help in the choice of the central wavelength. With higher wavelengths (up to near infrared), higher penetration can be achieved. In biological samples, the water absorption characteristics however set

a limit to the wavelength that can be utilised. This has coined the term “diagnostic window” [46].

2.3.5 Reference arm

This component is integral to OCT systems. The light that is reflected from a specimen cannot be detected directly and the signal registration requires interference. The mirror used for this purpose is part of the reference arm. In the original time domain setup, it is this mirror that is moved in order to allow mapping over several depth points. The speed of such mirror movements is mechanically limited. Rapid scanning optical delay (RSOD) lines use galvanometer-mounted turning mirrors (Figure 2.7). It is imperative that the optical path of the reference and the sampling arm have equal length. This has to consider any path the light travels, and the optical density of the medium. For instance, the path of a RSOD as in Figure 2.7 measures $2 \times$ from the fibre end to the diffraction grating, $8 \times$ the focal distance and $2 \times$ from the diffraction grating to the double pass mirror.

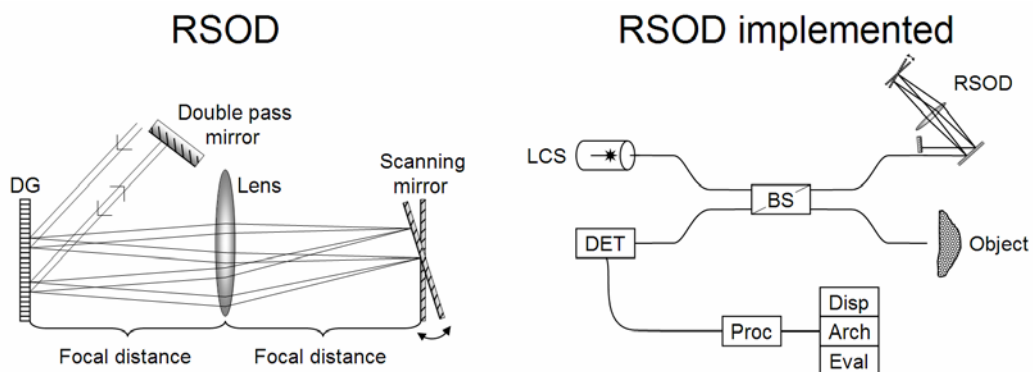


Figure 2.7: Left: Layout and optical path of a rapid scanning optical delay line (RSOD). The scanning mirror is mounted on a galvanometer. Right: the RSOD implemented into an OCT system (modified after [47]).

2.3.6 Detector

The intensity at the detector, for a standard Michelson Interferometer, is given by Equation 2.4:

$$I_0 \approx \frac{1}{4}|E_r|^2 + \frac{1}{4}|E_s|^2 + \frac{1}{2}E_r E_s \cos\left(2\frac{2\pi}{\lambda}\Delta l\right) \quad (\text{Equ. 2.4})$$

I_0 ...intensity at the detector, E_r ...field of the reference, E_s ...field of the signal,
 $\Delta l = l_r - l_s$...length difference of signal arm and reference arm (after [44])

The SNR can be formulated by standard techniques from optical communications theory and reflects that a higher optical power is required when imaging with higher speed or resolution (Equation 2.5 and Figure 2.8):

$$\text{SNR} = 10 \log\left(\frac{\eta P}{2h\nu \text{NEB}}\right) \quad (\text{Equ. 2.5})$$

η ...detector quantum efficiency, P ...detected power, $h\nu$...photon energy,
 NEB ...noise equivalent bandwidth of detection (after [45])

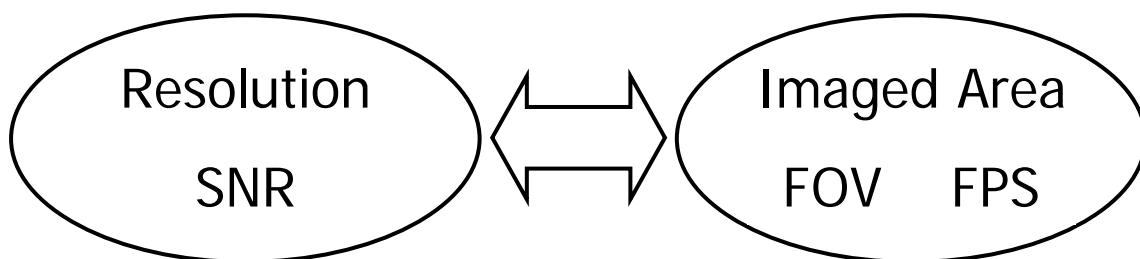


Figure 2.8: Trade-offs for information per time. FOV...field of view; SNR...signal-to-noise ratio; FPS...frames per second.

Time domain systems allow heterodyne detection: Part of the light is used as baseline for noise reduction of the signal [43]. For this, the detector registers the signal with two photodiodes (Figure 2.9).

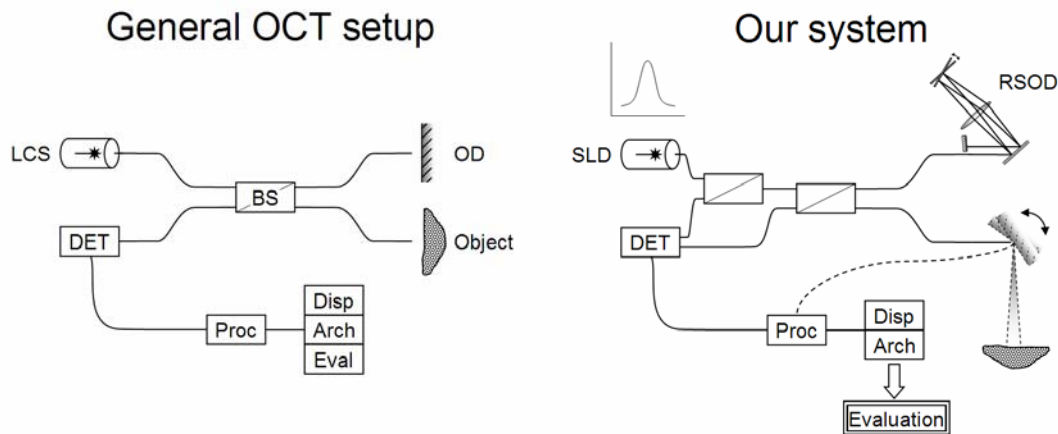


Figure 2.9: Left: The general OCT setup as shown in Figure 2.2. Right: heterodyne detection as used in this study. SLD...super luminescent diode.

2.4 Contrast, penetration depth and resolution in OCT

The mode of interaction that is registered from a scan is the depth-dependent back-scattering. Here it might be worthwhile to put OCT in relation to other technologies that acquire back-scattering profiles (confocal microscopy (CM) and ultrasonography): According to Figure 2.10, OCT nicely fills the gap between the two in regard to resolution and penetration depth.

Contrast may be enhanced by either reducing the refractive index mismatch of clinically insignificant tissue matrix, or by increasing the mismatch of the clinical markers.

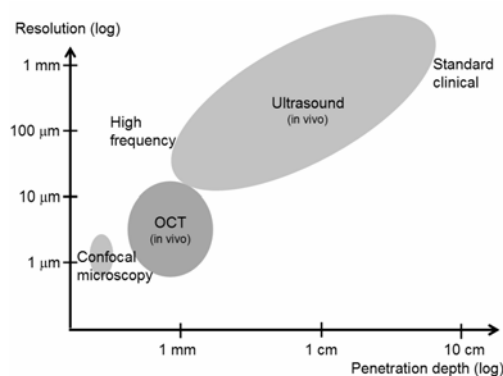


Figure 2.10: OCT, ultrasound, and confocal microscopy put into contrast in terms of their penetration depth and typical resolution (modified after [47]).

In practice, this is not straightforward as application of agents raises issues such as tissue interaction, time dependency, protocol adherence and the dynamic range of the

contrast enhancement. Most critical for OCT, introducing agents creates an invasive procedure.

2.4.1 Autochthonic contrast enhancers

Certain conditions or organs could have native or acquired contrast agents, which would avoid issues associated with foreign agent. E.g., the assessment of lung tissue may be facilitated by the abundant carbon depositions in smokers' lungs [48]. In functional OCT, the non-structural measurement may give stronger contrast: The reported penetration depths in blood vessel Doppler measurements are generally deeper than for standard OCT imaging. The information in these cases is not structure, but high (spatial) resolution velocity profiles.

2.4.2 Chemical agents

Contrast agents aim at altering the usual scattering behaviour. This is either achieved by altering scattering or by reducing the refractive index mismatch. Organic chemicals can alter the optical clearance behaviour in tissue [49],[50]. Along with increased penetration depth, the feasibility to pick up optically different features in less deep regions has been reported.

During colposcopy, acetic acid and iodine are standard contrast agents. Acetic acid has been displayed as influential in fluorescence spectroscopy and confocal microscopy [51],[52]. This might turn out to be relevant for clinical OCT measurements.

2.4.3 Engineered contrast enhancers

More sophisticated contrast agents that have been investigated are oil filled microspheres [53], and plasmon-resonant gold nano-cages or nano-shells which have a high scattering cross-section [54],[55]. Their signal strength increases exponentially with higher concentration. Application of dyes which are absorbing parts of the infrared spectrum have also shown to give notable contrast [56].

The so far mentioned contrast agents are changing the refractive pattern in a passive way. The magnetomotive contrast displayed by Oldenburg *et al.* makes use of agents which are known from magnetic resonance imaging, such as superparamagnetic iron oxide nanoparticles [57]. Their motion, which can be induced by external magnetic fields, shows up on the OCT image.

2.5 OCT in the clinical setting

We now understand OCT as a method for distance ranging, based on interferometric detection that allows mapping of reflection events with an accuracy of a few micrometres. For dense tissue, only areas just about a millimetre below the surface have a sufficient signal-to-noise ratio (SNR) to allow imaging. It is however non-destructive and quick. Table 2.1 gives a brief overview where OCT can be expected to be of benefit.

Table 2.1: Clinical features which render themselves to assessment with OCT

Clinical category	Clinical feature	When can this be expected
Localisation	Within penetration depth	Clinical markers positioned few mm in (dense) tissue or a few cm in clear medium
Superficial presentation	Sub-surface	Does not have to be evident on surface
Above resolution	>10 μm	The disorder provides markers of this magnitude
Topical assessment	In reach of fibre optics	Endoscopic applications
Contrast	Difference in refractive index	Architectural changes represent disorder or correlate with its state

The essence of Table 2.1 is: It is necessary to identify a marker that becomes relevant for the staging of the disease (diagnosis) or its likely progression (prognosis). This marker should be within the range of the imaging modality and correlated with the staging. It cannot be excluded that nuclear material with higher refractive index than surrounding tissue has a contrast similar to the H&E stain, and would therefore show up on an OCT image. This is the supporting opinion for investigating the benefit of OCT in epithelial malignancies.

3 The research question: OCT for cervical precancer?

The previous chapters have served as introduction for two topics: The clinical need for improving the situation in precancer management of the cervix, and the possibility to improve this situation with OCT. It was therefore the aim of the project to assess the possibility of OCT in the clinical setting. For the purpose of this thesis, this project is broken down into aims and objectives:

3.1 Aims and objectives

The aim of this thesis is to characterise the potential of OCT in cervical precancer management. At this point it has to be reminded that the existing management is already successful in reducing the incidence and that even a promising result might not be sufficient for improving the current situation. Findings might however make their way to the management of other diseases which are not currently covered by a screening program.

The following chapters are organised in the way the objectives were tackled:

3.1.1 Chapter 4: Review of OCT and competitor technologies

There are other optical high-resolution non-destructive measurement methods which are being investigated for their future role in disease management or characterisation. This chapter shall explore the mechanism of detecting cancer markers and position OCT amongst its competitors.

3.1.2 Chapter 5: Review of OCT data analysis

We will establish that OCT instrumentation has progressed to a point where it accumulates large amounts of data. There will undoubtedly be a need for means of assisting data evaluation. Publications on automated evaluation of OCT data are explored for their potential in the context of this work.

3.1.3 Chapter 6: Pre-processing OCT data

For the subsequent classification it is crucial to perform pre-processing of data. Although this normalising is the basis for a successful discrimination model, it is an independent step that is explained in this Chapter.

3.1.4 Chapter 7: Introducing PC-fed LDA

Structural features in OCT images of dense tissue are not always pronounced and may be too weak for recognising cancer markers. It would be desirable to have an algorithm at hand that does not require prior knowledge of the disease marker. A classification algorithm is designed and explored on non-clinical samples.

3.1.5 Chapter 8: Application to clinical study

A collection of clinical OCT images is created. Using standard histopathology as reference method, OCT images are classified using the proposed algorithm. The results of this classification shall serve as an indicator of the possible performance of OCT in the clinical setting.

3.1.6 Chapter 9: Discussion

The intention of the last chapter is to put the achievements in this work into context of the aims and objectives and shall highlight the value of the findings.

3.1.7 Appendix chapters

Further work has been undertaken during this project, but is not necessarily essential for following the thought processes. However, this work is related to the developing of the algorithm and a potentially successful application in the medical field. These parts have therefore been put into the appendix for reference:

- Appendix B: Optimal parameters for PC-fed LDA
- Appendix C: Ethics files
- Appendix D: Portability: OCT for prognosis of Port-Wine Stains

3.2 Summary

A listing of the questions investigated in this thesis is given in Table 3.1:

Table 3.1: Summary of the hypotheses investigated in this thesis, methods to test them and rejection criteria.

Research question	Expected answer	Method	Aim
Can OCT be used as staging tool for pre-cancerous lesions in the cervix?	“Well it could be used, but...” type of answer.	Split into further hypotheses.	Define a role of OCT for the clinical setting in the near future.
Does OCT perform different than other high-resolution imaging tools?	Some difference in speed, contrast, invasiveness.	Undertake review to put OCT into contrast with competitor technologies.	Define advantages of OCT over other technologies.
Can OCT data evaluation in dense tissue be automated?	If a certain feature can be defined, an algorithm can be programmed to classify it.	Undertake review of data evaluation in OCT dense tissue images.	Define conditions under which these algorithms work.
Is linear discriminant analysis a suitable tool for OCT data classification?	Depends on appropriate pre-processing and variable normalisation.	Use test images to evaluate algorithm.	Investigate whether a classification has a correct rate (arbitrary) of >70 %.
Does automated analysis of cervical OCT images have a clinical benefit?	It should be applicable to identified markers.	Acquire image collection and apply developed algorithm.	Investigate whether the classification is worse than standard clinical histopathology scores in literature.

4 Modalities review: learning from prior work

Optical micrometer resolution scanning scanning for non-invasive grading of cervical precancer

Florian Bazant-Hegemark^{1,2}, Mike D Read³, and Nicholas Stone^{1,2}

1) Cranfield Health, Cranfield University at Silsoe, Bedfordshire MK45 4DT, UK

2) Biophotonics Research Group, Gloucestershire Royal Hospital, Great Western Road, Gloucester GL1 3NN, UK

3) Women's Health Directorate, Gloucestershire Royal Hospital, Great Western Road, Gloucester GL1 3NN, UK

Modified[†] from a manuscript that was published in
Technology in Cancer Research and Treatment, 7(6), 483-496, (2008)

4.1 Abstract

Background: Management of cervical precancer is archetypical for other cancer prevention programmes but has to consider diagnostic and logistic challenges. Numerous optical tools are emerging for non-destructive and rapid imaging for early diagnosis of precancerous lesions of the cervix. Some of these have reached a pre-commercial stage, but high-resolution mapping tools are not yet introduced in clinical settings. **Methods:** The NCBI PubMed webpage was searched using the keywords “CIN diagnosis” and the combinations of “cervix {confocal, optical coherence tomography, ftir, infrared, Raman, vibrational, spectroscopy}” and similar keywords. Suitable titles were identified and their relevant references followed. **Results:** Challenges in precancer management are discussed. Tools capable of non-destructive high-resolution mapping in a clinical environment were identified: confocal microscopy, optical coherence tomography, IR spectroscopy, and Raman spectroscopy. A critical review of the likely performance of these methods as diagnostic tool is provided. The rationale for carrying out research under the prospect of the HPV vaccine is given.

[†] Modifications include: British English spelling; cross-referencing; removing redundancies; and a greater level of detail.

Keywords: Optical Diagnostics, Cervical precancer

4.2 Methods: Search strategy and selection criteria⁵

This review aims to cover optical tools for staging of cervical precancer, capable of high-resolution imaging or mapping. Following an initial research, titles (not Mesh-terms) of the NCBI PubMed data base [58] were queried for “CIN diagnosis” and the results back to the year 1990 were sieved for titles that seemed relevant for non-destructive grading. Following this initial sifting, the database was queried for “cervix confocal”, “cervix optical coherence tomography”, “cervix ftir”, “cervix infrared”, “cervix Raman”, “cervix vibrational”, and “cervix spectroscopy”, and their relevant references followed. These high spatial resolution mapping technologies were selected because they are being developed and/or subject to active research for their benefit in staging cervical precancer. The criterion was whether there was an amount of clinical studies on human tissue available that had the dedicated purpose of investigating markers for adding benefit to current management of cervical precancer. Technologies that may have potential for use in the cervix but where such studies have not been published were omitted. Although small animal studies can be helpful in investigating techniques, there was a focus on studies of human tissue. The distinction between *ex vivo* and *in vivo* is made where appropriate. This choice shall not be understood as exclusive in the sense that other technologies are not equally legitimate, but should be seen as a self-imposed limitation on what this manuscript aims to discuss. Most recent studies were accepted as such when no forward references were produced from the ISI Web of Knowledge webpage [59]. Literature before 1990 was not part of the initial query but included if recognised as relevant in reference lists. This report covers the key publications concerning high spatial resolution cervical precancer mapping since 1975.

The emphasis was on high spatial resolution imaging at the micrometer scale, in contrast to wide-field imaging where some systems at pre-commercial stage are subject to advanced trials. The data provided by such systems is increasing rapidly and it seems appropriate to await a separate review of their performance in due course. In this

⁵ The introduction of this manuscript has been removed in this thesis for the sake of brevity.

context we would like to refer to earlier reviews on optical diagnosis of the cervix [6],[60]-[62], and on other reviews on molecular markers [63],[64] and methods of detecting HPV [65],[66].

4.3 Results: Emerging technologies for high-resolution mapping of the cervix

Emerging technologies are leading to the possibility of a real-time, minimally invasive assessment of early cervical disease. This section discusses confocal microscopy (CM), OCT, infrared (IR) spectroscopy, and Raman spectroscopy (RS), as these four are likely contenders for high-resolution mapping. Table 4.1 shows that they provide information on structure which correlates with histology and cytology, or on molecular composition which can be explained with knowledge of biochemical processes.

Table 4.1: Overview of modalities discussed in this review.

Modality	Confocal microscopy	Optical coherence tomography	Infrared spectroscopy	Raman spectroscopy
Probing for	Subcellular structure	Histological structure	Molecular composition, biochemical patterns	Molecular composition, biochemical patterns
Contrast	Backscattering and refraction	Change of refractive index	Peak intensities: absorption as function of molecule bonds	Raman signals as function of molecule flexibility
Axial resolution*	~1 μm	~5 μm	-	-
Lateral resolution*	~3 μm	~5 μm	Application / stage dependent	Application / stage dependent
Penetration depth*	~300 μm	~1 mm	-	-
Field of view	Microscopic area	~5 mm	Microscopic area	<i>In vivo</i> : point measure Microscopic area
Main imaging mode	<i>En face</i> , volumes from stacks	B-scan, volumes from adjacent B-scans	Point measurements and imaging	Point measurements, scan modes for mapping
Investigated precancer / proliferation features	Sub-cellular structure, nucleus/cytoplasm ratio	Histological structure, intact basement membrane	Characteristics of biochemical processes: glycogen, methylation, protein phosphorylation, DNA hydration, glycogen/phosphate ratio, RNA/DNA ratio	Peak intensities and ratios of collagen, elastin, nucleic acids, phospholipids, glucose-1-phosphate

* for typically applied wavelengths in cervical tissue

The differences in cervical epithelial cells which are surrogates for determining the malignant potential of a lesion are sound: size and shape of cells will alter the way light is reflected and scattered, different concentrations of cell components will result in

different amount of absorbance and transmission, and different amounts of fluorescent biochemical constituents would show semi-quantitative emission patterns.

Let us remind the reader that colposcopy itself is an assessment of the visual appearance of different types of epithelium, in order to obtain a combination of optical diagnostic markers: A typical examination refers to interaction of white light as seen under a microscope, and makes use of contrast agents for proteins (acetic acid) and glycogen (iodine). The visual impression contributes to the prognosis of CIN progression. However, the correct interpretation is based on the expert opinion of the clinician. In addition, aceto-whitening and iodine staining are not 100 % correlated with the cancer risk.

4.3.1 Confocal microscopy

A confocal microscope is generally described as a microscope which collects light that is coming from the focal plane only. The principle of CM was patented in 1957. Initial problems with the then available light sources were sample heating, focussing, and data transfer [67]. Mapping within a plane is achieved using galvanometer motors [68], altering the focal plane provides different slices for 3 D stacks. Today, lasers allow resolutions of around 1 μm , thereby providing images as shown in Figure 4.1.

CM visualises *en face* planes of microscopic structures: Studies on smear samples showed that subcellular structures, such as mitotic figures and chromatin patterns, can be resolved and hence contribute to diagnosis [69],[70]. The light for detection is created by either reflection or fluorescence, which opens the possibilities of this technology for functional imaging and double selectivity of target volume [71]. Fluorescence CM allows mapping fluorophoric metabolites (NADH, FAD) within subcellular structures and distinguishes between cytoplasmic and peripheral fluorescence [72]. Fluorescence has been shown to distinguish basal cells from parabasal, intermediate, and superficial cells [73].

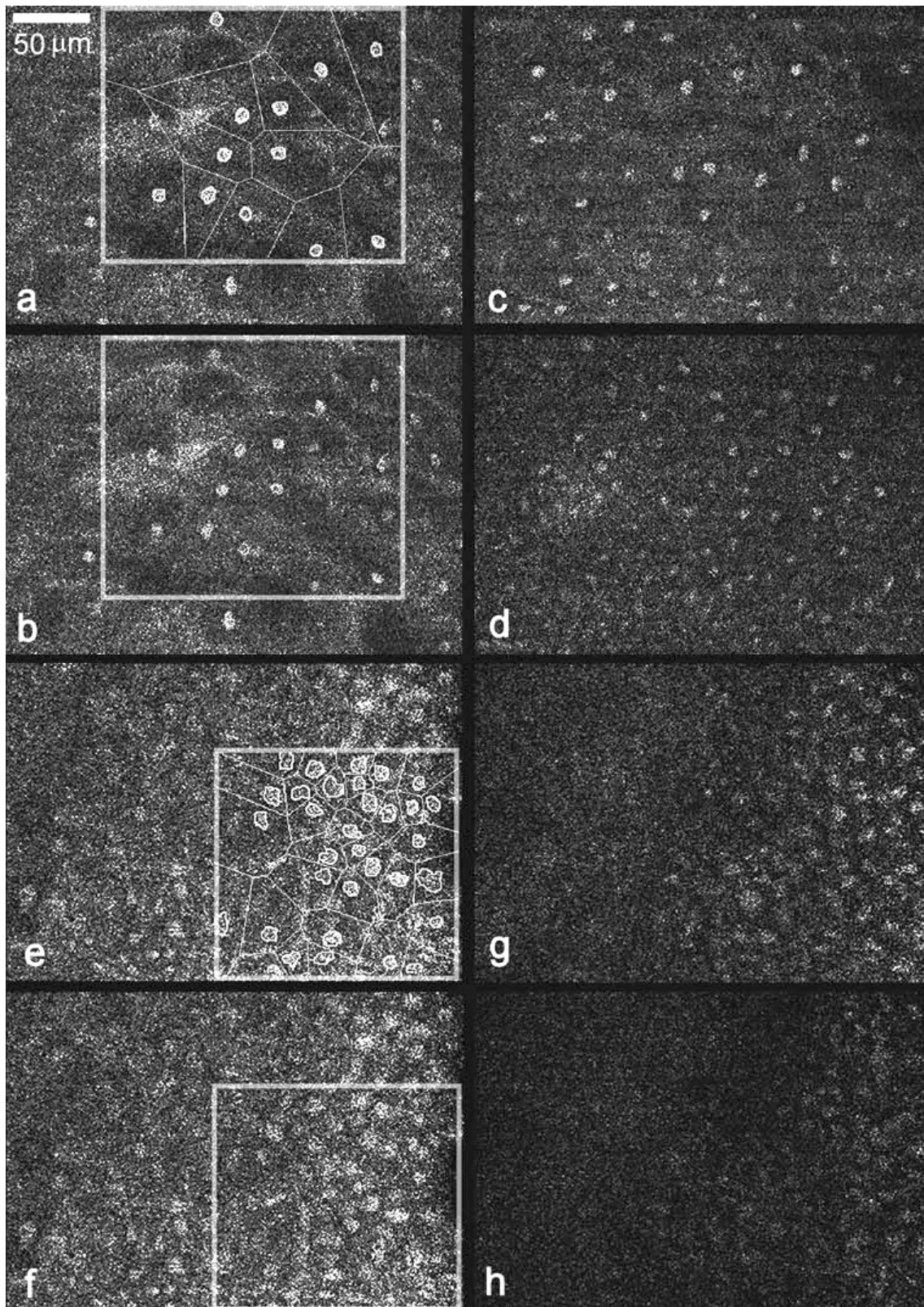


Figure 4.1: CM *en face* images of biopsy samples: Normal tissue at 50 μm below the surface (b), an intermediate layer (c), and close to the basement membrane (d), and of CIN3 tissue at the equivalent distances (f-h). (a) and (e) show an overlay of an automated segmentation algorithm. Reprinted with kind permission [78].

As the CM allows description of cellular DNA appearance, risk factors for cancer progression such as aneuploidy, morphological changes, nucleus-to-cytoplasm ratio, or irregularity of cell spacing, can be assessed [74],[72]. Simple chemicals can serve as contrast agent, e.g. acetic acid, which is a standard agent for colposcopy, improves the contrast for cell nuclei seconds after application [74],[75]. However, contrast agents can cause tissue swelling and complicate focussing [52].

The field of view for CM images is in the magnitude of a few 100 μm , mapping of larger areas would require sophisticated stitching software. Although tissue surfaces are not completely smooth it is possible to image cells near the surface and near the basement membrane, the ratio between the two can indicate the CIN stage [76]. The penetration depth (“working distance”) of modern probes is up to 250 μm , depending on the amount of scatterers in the tissue above. This can be significant in CIN2/3 biopsies; the basal layer cannot always be visualised [77]. However, as growth at the basement membrane would be difficult to distinguish from normal, 50 μm may be a sufficient depth for detecting progressed dysplastic features [78].

CM has been used for *ex vivo* sections in order to assess DNA aneuploidy as risk marker and also as measure for treatment outcome. Tissue sections for CM need to be considerably thicker (15-200 μm) than slices for routine histopathologic examination (~ 10 μm) [72],[74],[77]. Typical systems are capable of imaging a 515×660 μm^2 area with 2 μm lateral resolution at 15 frames per second [79], or a 600 μm field of view with 1.3 μm lateral resolution at 12 frames per second [80], or at 6.5 frames per second [78]. Multiphoton confocal microscopes, which may achieve much higher resolutions, use slightly higher resolutions in what seems to be a trade-off between spatial image resolution and field of view, as discussed in Figure 2.8. Studies of human cervical tissue have not been published at the time of writing. A study on murine cervical tissue with a multiphoton microscope imaged 512×512 pixel for a field of approximately 150×150 μm [81], and a study on a cricetine cheek pouch model using a multiphoton laser scanning microscope imaged 512×512 pixel for a 283×283 μm field [82].

Although the technical realisation of an *in vivo* system has been demonstrated [80],[83],[84], there are also recent *ex vivo* studies on cervical tissue [78]. For volumetric imaging stepper motors have to ensure accurate positioning. Studies which

abstain from tissue embedding have to cope with handling problems, when trying to image the full epithelial thickness: In a recent study only 9 out of 52 biopsies could be evaluated successfully, 43 were too fragile or had insufficient contrast [78]. However, handling of cervical biopsies can be tricky in general: The small size and fragile nature of biopsies can result in about 50 % of samples not being available for CM imaging [77]. Whether volumetric imaging is advantageous, has to be assessed: There are significant differences of nuclear volume, surface area and spherical shape factor between normal and abnormal cells; these are more pronounced on volumetric images [85]. Another study found that for classification, standardising to epithelial thickness for sub-surface 2 D slices gives better classification than from volumetric images, such a protocol seems therefore better for *in vivo* imaging [78]. Volumetric data can however aide automated segmentation; in the case of failed auto detection [86], adjacent planes can be used for correction [78]. The potential of volumetric CM seems to be best realised where sample stability is not an issue. Recently, the possibility of adapting conventional wide field microscopes to reach both a contrast and a section thickness comparable to CM has been displayed on a cervical cancer cell line [87], at lower cost than CM.

The majority of studies on the benefit of CM for cervical precancer assessment concerns establishing clinical features or system development. To the best of the authors' knowledge, there is only one study subjecting findings from CM to standard histopathologic assessment. It included 19 participants and allowed detecting the distinction of high-grade dysplasia from normal tissue with 91 % specificity and 100 % sensitivity based on nucleus-to-cytoplasm ratio [88]. Studies for the relevant classification of CIN1 vs. CIN2/3 are anticipated.

4.3.2 Optical coherence tomography

Studies of cervical tissues use near infrared light sources in the range of 800-1500 nm and powers which are harmless for skin [45],[89]. The contrast for structural features is determined by changes in the refractive index. Contrast agents for OCT are an active research topic [90] but have not impacted on studies of the cervix. In contrast to the other modalities, conventional OCT does not acquire *en face* images (although such scanning regimens exist [91]). The depth (axial) resolution is decoupled from the lateral

resolution and both are a crucial specification. Typical resolutions are in the magnitude of 5 μm laterally and axially in tissue. The penetration depth in dense tissue is around 1 mm. Strong absorbers in human tissue, e.g. melanin (visible range) and water (infrared range) limit the choice of wavelength. Typically applied are near infrared wavelengths where absorption curves have a minimum. Light sources for OCT have to fulfil stringent criteria in respect to partial coherence, in order to allow interferometric detection. The quality of a beams “pulse” directly determines the axial resolution. Ideal sources are not available for every wavelength and large efforts of OCT research concern laser sources. However, the advantage of the interferometric detection setup is that it is practically not confounded by ambient light.

Research systems today can measure 10^6 to 10^7 A-scans per second, which allows several volumes per second and so-called 4 D OCT [92]. Cervical studies with such systems are however still awaiting their publication. Issues with OCT are the low contrast and the amount of data generated, which stretch modern data storage systems and will require automated data evaluation.

The non-invasive, real-time OCT was a diagnostic revolution for ophthalmology. First images displaying endoscopic OCT assessment of cervical malignancy have been made in 1997 and 1998 [93],[94]. Because the resolution of OCT is smaller than epithelial cells, it should be possible to image cervical pathologies in real-time. A first *ex vivo* study on tissue of the female genital tract included 32 cervical specimens [95]. This study concluded that an intact basement membrane can be depicted as a feature of healthy epithelium, and included 21 healthy and 11 CIN3 or carcinoma specimens. Escobar *et al.* performed an *in vivo* study on 50 patients, and included the diagnostically more relevant samples, from CIN1, CIN2/3, and inflamed sites [96],[98]. It allowed concluding that the thickness of squamous epithelium is varying for healthy samples, and confirmed healthy epithelium as a marker for an intact basement membrane. This membrane itself is too thin for depiction on OCT images. Instead, the visible marker is a distinct boundary between squamous epithelium and underlying stroma. Zuluaga *et al.* reported a correlation in the backscattering intensity of normal and abnormal tissue [99]. Findings of a study with data from 212 patients resulted in the suggestion to use OCT for confirming positive colposcopic examinations prior to treatment, in low resource

settings [100]. All studies used 2 D images, such as shown in Figure 4.2, whereas state-of-the-art OCT is volumetric.

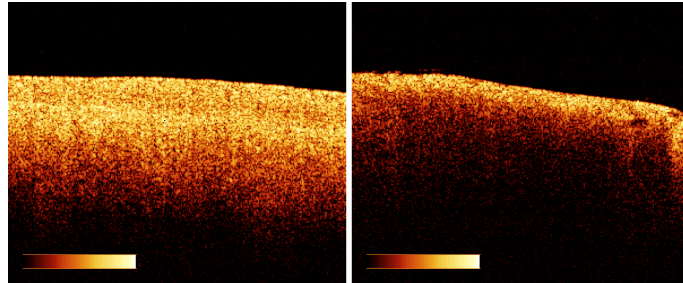


Figure 4.2: OCT images of ectocervix with apparent layer-structure (left) and less vertically structured (right) ectocervix. The bar measures 1 mm.

Imaging of the gastrointestinal (GI) tract has shown that 3 D OCT gives a much better representation of the microscopic organisation of tissue structures [101]. However, further markers which can be definitely correlated with CIN progression remain to be established. In ophthalmology, OCT development could be targeted to successfully resolve the retinal layers. In contrast, dividing the cervical epithelium into basal, parabasal, intermediate, and superficial layers is a stratification based on relative cellular appearance; it is not based on a boundary of the organism as such. In absence of clear structural markers, individual A-scans may serve as measure for optical properties. An algorithm looking at scattering coefficients has been investigated, but seems to have stringent requirements in respect to tissue layering [102]. No algorithm has been evaluated for larger sample size of clinically relevant groups.

In vivo probe design for volumetric data faces the challenge of guiding the laser beam accurately in 2 dimensions. For cardiovascular imaging and the oesophagus, helical scanning protocols exist to acquire tubular volumes from within [103]. This could provide useful for imaging of the endocervical canal. Such setups have the focal point at a distance relative to the helix axis, and have to assume a tissue surface with this cylindrical shape. Where this is not the case, e.g. at folds, data will therefore be of reduced quality or missing. This might be relevant when glandular carcinoma develops exactly in such folds.

Fluorescence and OCT

An approach to overcome the limited contrast for biochemical processes is to combine OCT with fluorescence, and devices for such measurements have been built [104]. The diameter of the endoscopic device is 2 mm. Technical challenges arise because OCT requires a focused beam, whereas the optical setup for fluorescence imaging is designed to provide homogeneous, diffuse illumination. *In vivo* measurements on a mouse colon take 45 seconds for a 6 mm stretch. The two modalities are not completely overlapping, a 1300 nm OCT beam allows a ~ 15 μm resolution and penetration depth of about 1 mm, and the 325 nm excitation source for fluorescence penetrates approximately 200 μm into tissue. Housing both modalities in one system allows a reliable co-registration of measurements. This addresses a major issue for modalities at this level of resolution. Measurements for cervical tissue have not been published, but fluorescence-guided OCT has been attempted in other organs.

4.3.3 IR spectroscopy

IR spectroscopy measures wavelength-dependent absorption properties by probing vibrating energy levels of functional groups and intermolecular interactions. This is possible in molecules with an IR active dipole. Fourier transform infrared (FTIR) spectroscopy is the standard method today but still often referred to as “FT”IR to emphasise the method of spectral data collection using FT interferometry rather than dispersion gratings. For medical applications, near- or mid infrared light (between 0.8 μm and 2 μm) sources give the best contrast. In contrast to analysis of pure substances, spectra from biological tissue represent a mixture of all *different* molecules in the interrogated micro-volume.

The typical mode of operation is absorption spectroscopy. In order to obtain a sufficient SNR, transmission measurements have to be taken from thin tissue sections or thin cell layers. Initial publications are therefore on exfoliated cells [105]. A review with an emphasis on different IR spectroscopy instrumentations for cervical cancer screening has been published recently [106]. Wong *et al.* measured spectra from 156 samples, categorising into normal, dysplastic, and malignant [105]. They found that cancerous progression in mature epithelial cells, but also normal proliferation, go in hand with glycogen reduction, increased hydrogen bonding from phosphodiester groups, and a loss

of hydrogen bonding in alcoholic groups of amino acids. Such features are intuitive as they represent energy demand and mitotic activity, and have been reconfirmed by subsequent studies on cytological smears [107]-[109], cancer cell lines [110],[111], and tissue [112]-[115].

While peak ratios of glucose/phosphate (decrease) and RNA/DNA (increase) seem to correlate well with proliferation of cervical epithelial cells, detecting a malignant candidate is more difficult: An experienced spectroscopist might be able to interpret the shape of a spectrum or peak ratios, but this is not ideal for the diagnosis of precancer. Chemometrics can help giving a better estimation for classification [106],[107],[114],[116],[117]. The investigation of IR spectroscopy for precancer assessment coincides with the evolution of personal computers which facilitate such analysis. Computers allow objective and reproducible classification methods that are based on complex calculations, practically remove the need for staff training, and allow a statistical estimation in border areas. An appropriately trained classification model can separate inflammation, endocervical cells, and bacterial infection [107],[118]. As with conventional cytology, debris, red blood cells, polymorph cells, endothelial cells and other components can confound the analysis [110],[119].

Some of these findings deserve a closer look for their diagnostic value in addition to precancer staging. The type of cells and their maturity state is of great importance, and the spectral signatures of immature, benign cell types can obscure abnormal cells [118],[120],[121]. Sampling plays an important role and issues concern areas such as the origin of a sample [120], the homogeneity within the sampling volume [116], cell handling [107], or cleaning protocols [122]. Spectral signatures of cells change after two hours of being suspended in saline solution at room temperature [119]. Cells would therefore have to be stored at a defined low temperature, which for a large scale program would pose a logistical challenge. Essentially, a protocol needs to ensure that it is valid to assume that: a proliferative signature within the sampling volume can be considered as a diagnostic factor for CIN. Still, in a study on 301 cytological samples, IR spectroscopy classified better than cytology (98.6 % sensitivity and 98.8 % specificity vs. 86.6 % and 90.5 %) [108]. The main potential for improvement derived from a study on 2000 cells of 22 women seems to be that cytologically or

morphologically normal epithelial cells from abnormal smears have an abnormal biochemical signature [112].

In search for sampling types, lavage specimens have been investigated [123]. The issues with sample collection and the performance were not convincing enough to replace smears; yet it was suggested as an alternative where smears are not possible.

Although IR spectroscopy is still being investigated for assisting cytology [124], it is not restricted to smear evaluation. In combination with a stage, tissue mapping is possible. IR systems can be designed for true imaging, in the sense that separate pixels are measured in parallel, rather than scanned sequentially. In order to be able to assume a homogeneous sampling volume, this is done on a few micrometre thin tissue sections. Typical systems take about 20 minutes for $500 \times 500 \mu\text{m}^2$ area (IR microscopy or microspectroscopy). Preparation of thin tissue sections, which presents a main cause for delay in today's routine, requires an effort quite comparable to conventional histopathology: obtaining the biopsy, embedding or freezing, slicing, transferring to slide, fixing and staining [109],[113]. Spectral differences in tissue are less pronounced than in exfoliated cells, but still allow good classification [121]. On the other hand, mapping overrides some of the difficulties of correctly locating a cell in its histological environment. It allows characterisation of the different cell and tissue types in cervical epithelium, such as cervical stroma, epithelium, inflammation, blood vessels, mucus, and abnormalities thereof [109],[113],[125],[126]. The key advantage of IR spectroscopy is the objective evaluation to assist the histopathologist. The false positive rates can be as low as 9.04 % for CIN1 and 0.01 % for both CIN2 and CIN3 [127], or 0.04 % [119], and can predict cancer in biopsies with 95 % accuracy [114].

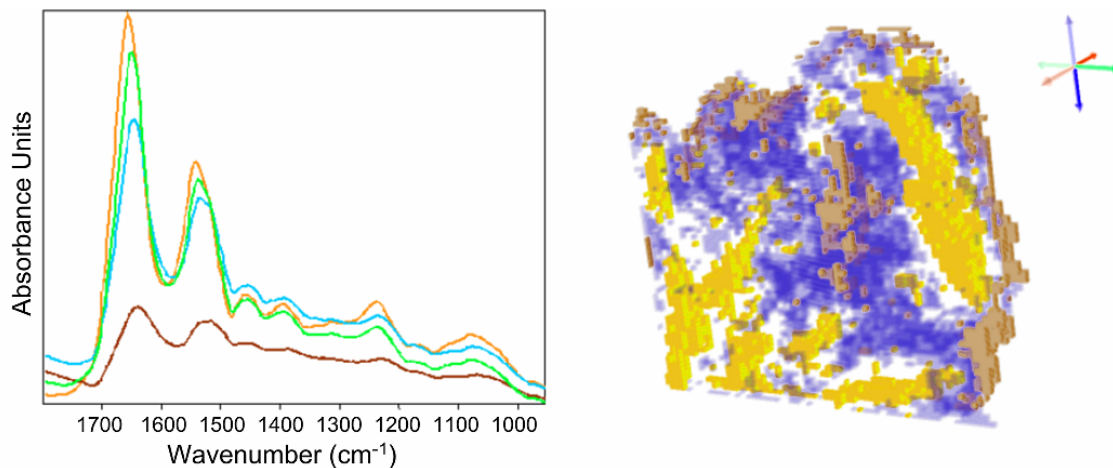


Figure 4.3: Left side: Mean spectra of different tissue types. These can serve as a reference for unknown spectra (multivariate models are more complex). Right side: Volumetric presentation of consecutive IR maps. The colour of each voxel represents the classification likelihood for a type of cervical tissue. The volume measures $1,400 \times 1,400 \times 16 \mu\text{m}^3$. Modified and reprinted with kind permission [128].

Wood *et al.* have taken this approach into the third dimension: By stacking maps from several consecutive tissue sections they created a 3 D stack (Figure 4.3). The whole process from taking the measurement to multi-slice analysis and volume rendering lasts less than 1 hour for a four-section model covering $1,400 \times 1,400 \times 16 \mu\text{m}^3$ [128]. Automation for cytological or histopathologic examination faces veritable challenges and requires a standardised sampling protocol. However, as a research tool it already has provided valuable evidence for understanding the biochemistry of precancer; and it can be imagined as an adjunct for *ex vivo* staging.

4.3.4 Raman spectroscopy

Raman spectroscopy (RS) is a spectroscopic method probing for a certain type of inelastic scattered light, the Raman-scattering. Raman signals are a function of the incident wavelength and light sources used are lasers operating at a single wavelength. Although the Raman Effect was discovered in 1928, it required the availability of appropriate laser sources to be investigated in the medical research field. RS differs from IR spectroscopy: The latter probes for absorption caused by changes in dipole moment. RS characterises molecules according to their change in polarizability, and aims to measure deviations from elastic scatter. Despite of these differences, evaluation and

presentation of Raman data are similar to IR spectroscopy (Figure 4.3). RS and IR spectroscopy together are commonly referred to as vibrational spectroscopy [129].

Raman signals from tissue are weak and occur, for applied wavelengths and powers, statistically only one time for 10^{10} incident photons. Other effects easily cover a Raman signal, e.g. fluorescence is about a million stronger [130]. A significant disturbance is fluorescence from biological tissues and from fibre optics (e.g. silica), and filter designs for contribution of optics are a crucial component of *in vivo* Raman systems [131]. The near infrared range around 780 nm is typically used for fibre optics probes; but also *ex vivo* studies using ultraviolet light (254 nm) exist [132]. Fluorescence which occurs at the same wavelength range as the Raman signal cannot be filtered out using normal filter design, and may require baseline subtraction [133]. Cross talk is also the reason why 2 D Raman data are mapped, rather than imaged. This makes RS a much slower technology than IR spectroscopy. Novel approaches which are stream-lining rather than scanning have a huge impact on scanning speed [134]. Systems employing time resolved technology to “time-gate” out the fluorescence signal require sophisticated hardware which can occupy a large room [135]. Although near IR light would penetrate deeper into tissue, it is estimated that a good Raman SNR from cervical tissue is only obtained from epithelial layers [136].

Raman active molecules and chemical groups allow characterising the presence, quantity, and composition of tissue [137]. Suggestions for using this technology to distinguish abnormal and normal tissue of the gynaecologic tract were mentioned in 1992 [138]. Results from a system using a 254 nm light source on cell suspensions, allow distinction of normal from malignant cells due to peak ratios from different nucleic acid contents [132]. Recent studies use IR lasers. *In vitro* tissue samples (36 biopsies of 18 patients) could be classified as precancerous or normal with 82 % sensitivity and 92 % specificity [139].

Results can be obtained from IR-inactive molecules and can be quite complementary to IR spectroscopy. However, the methods for statistical analysis of spectra are similar [140]. Utzinger *et al.* used peak ratios of cervical tissue spectra to distinguish normal from dysplastic areas [141]. Diagnostically significant molecules were demonstrated to be DNA, phospholipids, and collagen [141]. Analysis of characteristic peaks and fingerprint regions agrees with the biochemical picture provided by IR spectroscopy

[105]: basal cells show characteristic bands of nucleic acids, epithelial cells show high levels of glycogen in normal cells. Decreased levels of glycogen and increased levels of DNA and amino acids in epithelial layers are a feature of cancerous cells [136],[142],[143]. A study on 3 cell lines (keratinocytes representing normal, HPV 16 – infected keratinocytes, and a cervical cancer cell line), was able to distinguish pairs of these lines with 70-100 % sensitivity and 70-90 % specificity, i.e. also those cells regarded as normal but expressing a viral protein from normal ones [144].

In contrast to IR spectroscopy, *in vivo* probe systems which can probe for the biochemical changes exist and have been used for RS [145]. *In vivo* results differ from *ex vivo* measurements, but are still able to discriminate healthy or squamous dysplastic tissue [142]. Even spectra from formalin fixed discriminate well [146]. However such spectra are not comparable with fresh tissue, in fact RS is so sensitive to embedding resins that it proved useful for characterising the efficacy of dewaxing agents [147]. Hence Raman pilot studies should be performed in an environment the actual system is intended for: fresh tissue for *in vivo* applications [148], or on fixed cells for cytological measurements [144]. Unlike IR spectroscopy, RS can be used for both *in vivo* and *ex vivo* measurements of biochemical constituents. RS requires reduced ambient light to minimise spectral contribution. Taking measurements in a dark room or using a covering cloth can address this issue [141]. Recent *in vivo* probes require few seconds for measuring a (point-) spectrum and can be considered real-time [130],[143],[149]. *In vivo* mapping seems unlikely due to these time requirements, and exactly locating the measurement on the cervix is necessary. Even although motion artefacts could potentially be addressed by image registration methods [150], the length of the procedure would still have to be kept on an acceptable level for patients.

The available tissue studies restrict themselves to classification between normal and high-grade or cancerous samples: Krishna *et al.* investigated 150 biopsies, distinguishing normal vs. malignant tissue nearly 100 % correct [136]. Lyng *et al.* subjected histological samples from 40 patients to classification [142]. Raman signals from CIN samples lie between normal and cancerous tissue, and can be distinguished with nearly 100 % sensitivity and specificity. CIN was not further staged and the clinically relevant distinction between CIN1 and CIN2/3 was not addressed in this paper. The latest *in vivo* study on data from 66 patients showed that RS has 89 %

sensitivity and 81 % specificity to distinguish high-grade dysplasia from benign tissue, higher than the colposcopy for these cases (87 % and 72 %, respectively) [143]. Here the classification performance against low-grade dysplasia was attempted; because of the small sample size (6 spectra) in that group this model cannot be considered strong. The introduction of new statistical algorithms for classifying spectral variations of disease groups might help improving the classification performance and is an active research field [151].

Recently, the technical feasibility of depth probing Raman signals has been displayed repeatedly [135],[152]. The possibility of probing biochemical processes opens another field: While the other modalities give a diagnostic snapshot at the time of the measurement, RS could potentially provide a prognosis [153].

4.4 Discussion

Markers for an increased proliferative behaviour are intuitive, yet it remains difficult to directly conclude an increased risk of cancer progression from them. Better understanding even a rather defined entity as cervical cancer is challenging, and even more would be developing an appropriate *a priori* model for cervical cancer. For a complex system like human tissue, data evaluation requires multivariate statistics and data reduction techniques. Inevitably, a reference in the form of databanks has to be established in large clinical trials.

Combining reflectance spectroscopy and fluorescence imaging has shown that non-destructive testing of epithelium is possible and such systems are close to commercialisation. Based on available trial data, these systems seem to increase the sensitivity for CIN2/3 on the expense of a reduced specificity for CIN1. However, their advantage lies in the objective judgement which does not require the immediate expertise of a trained colposcopist. In this context, when regarding a spectral peak as suitable for discrimination, it is not the raw spectrum, but the classification model based on such spectra, which allows good discrimination. This has to be kept in mind, and the choice of statistical classification might well influence the performance.

Although systems utilising reflectance spectroscopy and fluorescence are near commercialisation, they are not 100 % definite in diagnosis [154],[155]. They are therefore intended as adjunct for colposcopy, in order to guide biopsies. For better

classifying disease and underlying mechanisms, it might however be beneficial to directly measure biochemical constituents and their relative concentration, or subcellular and sub-surface structure. High-resolution modalities which are being actively developed might be able to address these issues. For high-resolution *in vivo* measurements, it is important to realise how large the required area would be: Studies show that the radial location of a lesion on the cervix is random [156], and evaluation of the 2002 screening protocol of the International Federation for Cervical Pathology and Colposcopy showed that classifying the extent of a lesion (minor or major) and the localisation of a lesion in respect to the TZ - which is changing for each patient - are important for correct classification at the stage of colposcopy [21]. Therefore, tools aiming to replace colposcopy need to image the entire cervix.

Although some technologies deliver astonishing data, the clinical value for managing cervical precancer might not be sufficient. However, even if new findings do not justify replacing the conventional management, diagnostic tools can be valuable in scenarios where a cervical management program has not been introduced, or serve as a basis for other organs where screening programs are not yet established.

4.4.1 Continuing research under the prospect of HPV vaccines

During this project, the development and promising results of clinical studies on HPV vaccines received much media attention. Although approval by the FDA is only of immediate relevance in the USA, it indicates a milestone in the actual market availability. Therefore, with the availability of FDA-approved vaccines which are aimed against HPV, there is now a prospect of having a new means of tackling this disease: A vaccination against cancer. These vaccines can be expected to reduce cervical cancer linked HPV 16/18, the two most common genotypes in precancerous lesions which account for an estimated 70-75 % of all cases [4],[157]. However, the aetiology of cervical cancer can not be reduced to infection by these two strains. The vaccine is prophylactic, which requires an individual not to have been exposed to these HPV strains prior to vaccination [158].

It remains to be observed whether the currently available HPV vaccines will eradicate cervical cancer. The “ESGO” statement of the European Society of Gynaecological Oncology reiterates that carefully conducted research and long time monitoring will

remain necessary before HPV vaccination schemes may lead to a change in present cervical cancer prevention programmes [159]. Neither HPV nor CIN are the actual disease which causes worry. It is the higher risk of progression to cancer that is of concern. Disease management, including screening, superficial assessment, and sub-surface examinations as part of precancer prevention will remain necessary.

4.4.2 Acknowledgements

This work is funded by the Pump Prime Fund of Cranfield University and Gloucestershire Hospitals NHS Foundation Trust.

Dr. Nicholas Stone holds a Career Scientist Fellowship which is funded by the UK National Institute of Health Research.

5 Analysis review: enabling quick and objective assessment of numerous measurements

Towards automated classification of clinical optical coherence tomography data of dense tissues

Florian Bazant-Hegemark^{1,2} and Nicholas Stone^{1,2}

1) Cranfield Health, Cranfield University at Silsoe, Bedfordshire MK45 4DT, UK

2) Biophotonics Research Group, Gloucestershire Royal Hospital, Great Western Road, Gloucester GL1 3NN, UK

Modified[†] from a manuscript that was accepted for publication in
Lasers in Medical Science

5.1 Abstract

The native contrast of optical coherence tomography (OCT) data in dense tissues can pose a challenge for clinical decision making. Automated data evaluation is one way of enhancing the clinical utility of measurements. Methods for extracting information from structural OCT data are appraised here. A-scan analysis allows characterisation of layer thickness and scattering parameters, whereas image analysis renders itself to segmentation, texture and speckle analysis. All fully automated approaches combine pre-processing, feature registration, data reduction, and classification. Pre-processing requires de-noising, feature recognition, normalisation & refining. In the current literature, image exclusion criteria, initial parameters, or manual input are common requirements. The interest of the presented methods lies in the prospect of objective, quick, and/or post-acquisition processing. There is a potential to improve clinical decision making based on automated processing of OCT data.

Keywords: Biomedical image processing, Biomedical optical imaging, Image analysis, Image classification, Medical diagnosis

[†] Modifications include: British English spelling; cross-referencing; removing redundancies; and a greater level of detail.

5.2 Introduction

Specifically in the medical field, the non-destructive imaging provided by OCT is attractive (Chapter 2.5). Through rapid high-resolution imaging, detailed structural information becomes available – in vast datasets.

Structural information is often intuitive to view, as shown in Figure 5.1. Strictly speaking however, an OCT image might not be relevant in itself. For clinical diagnosis, a new repertoire of image interpretation skills and artefact knowledge is necessary, which constitutes a learning threshold for the intended operator. There are therefore efforts to provide more than a structural image alone.

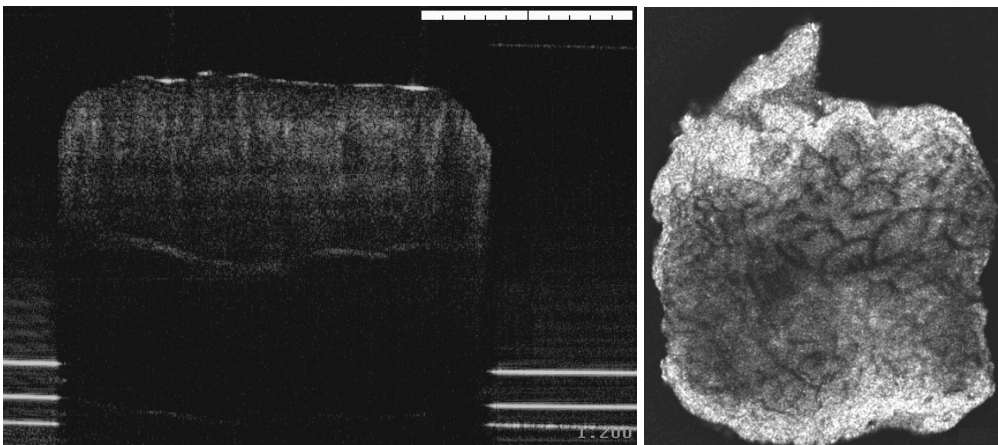


Figure 5.1: Left image: B-scan of an oesophageal biopsy specimen, of about $2 \times 2 \text{ mm}^2$ size, obtained ex vivo by a 7.5 m resolution (in air) frequency domain system operating at 1300 nm (Michelson Diagnostics EX1301 OCT microscope, Michelson Diagnostics Ltd.). The faint tissue layers make it difficult to decide whether clear layers are amiss due to weak contrast or because of the clinical presentation of the disease. Right image: En face sub-surface section, viewed from top, of a volumetric data set composed of 500 such B-scans. The improved representation as opposed to 2 D images is evident. The raw data requires several gigabytes of memory. The clinical requirement for this technique would be to contribute to diagnosis, and determine whether the patient should receive cancer treatment. N.B. the weak contrast of this Figure is not due to reproduction.

5.2.1 Challenges with OCT

Contrast

In dense tissue, characteristic limitations of OCT are the penetration depth in the magnitude of 1 mm and the mode of contrast, created by the variations of the refractive index. This native contrast can be weak, especially when compared to other non-destructive high-resolution methods. E.g., fluorescence imaging or Raman spectroscopy are able to pick up single molecules or alterations of protein structures

[160],[161]. OCT has advantages over these techniques: It is able to maintain the high-resolution over a large area; it is able to provide depth-resolved high-resolution *in vivo*; and the requirements of OCT instrumentation subsequently mean that OCT does not rely on imaging conditions such as reduced ambient light.

Since it would be attractive to have the combination of a strong marker and the wide-field, high-resolution, depth-resolved, rapid data acquisition, there are efforts to engineer multimodal systems [162],[163]. Equally important, attention is brought to improving clinical contrast in OCT data: Contrast agents have the potential to provide a clear signal for clinical decision making and have become quite a large research field [56],[90]. In addition to the structural characteristics, OCT data can provide functional information. Functional OCT imaging includes Doppler OCT, polarisation sensitive OCT, second-harmonic OCT, or spectroscopic OCT. These carry the potential for providing physiological information about flow properties [164],[165], birefringence [166],[167], neural activity [168],[169], or water content and oxygenation states [170],[171].

Automated data evaluation

A further approach to enhance clinical utility is to analyze OCT data by computational methods. This area has the potential to provide unsupervised, reproducible and objective measures, and to provide both quick and/or post-acquisition classification. When engaging ourselves in this field, we found that research groups tend to use different approaches and the emphasis is often on the image acquisition rather than the data analysis. With OCT being a novel technique, there is still a tendency to treat the OCT image as a result rather than as a step towards a clinician decision following detailed analysis. Subsequently, some papers even describe their analysis in the results section. We wanted to investigate whether there is a common pattern or ideal algorithm which could be suggested for further research.

5.3 Rationale for this review and limitations

The topic of automating OCT data evaluation has to be seen in context with other contrast enhancing approaches. Algorithms can fail and a clear marker, be it from contrast enhancing or functional data, will facilitate processing. In this manuscript, we want to concentrate on automated algorithms for retrieving information from

conventional “intensity” data. We therefore refer the reader to the above mentioned literature for more detail on functional OCT and contrast enhancing. Even when limiting ourselves to structural data of dense tissues, we certainly cannot claim to provide an exhaustive summary of every attempt so far. However, with this manuscript we would like to provide a starting point from where further analysis attempts can be systematically explored, and combined with other image enhancing methods for OCT.

5.3.1 Selection

Table 5.1 gives an overview over the approaches discussed in this manuscript. In general, automation requires recognition of relevant features. Identifying such a feature has to be mastered by an automation algorithm, which requires calculation time and has the potential to introduce errors. The grouping in Table 5.1 is somewhat arbitrary, e.g., most of the texture based approaches are subsequently classified. However, Table 5.1 shows for instance that A-scan analysis renders itself to measuring layer thickness or for deriving optical tissue parameters, while B-scans permit speckle and texture analysis.

Table 5.1: Articles employing feature recognition or automated steps for analysis of OCT data of dense tissues. ROI...region of interest

	First author and year	Specimen type	Sample size for analysis	Objective	Remark
Optical properties	Esenaliev 2001, [174] Kuranov 2006, [179]	phantom, animal tissue	-, 8 pigs	monitor blood glucose levels	automation not mentioned but seems feasible
	Levitz 2004, [176]	arteries	151 regions, 14 images	differentiate normal vs. atherosclerotic	ROI and initial parameters required
	Turchin 2005, [102]	phantom, epithelium	19 images, 1 patient	classify normal vs. pre-malignant	ROI and initial parameters required
	Jeon 2006, [177]	rat brain tissue	19 ROIs of 3 images	guide deep brain stimulation	ROI marked manually
	Ramrath 2007, [178]	rat brain tissue	-	guide deep brain stimulation	feature correlates with distinct spikes, requires initial training
	Iftimia 2005, [204] Goldberg 2008, [205]	breast tissue	3 biopsy samples 158 scans, 58 patients	guide fine needle aspiration biopsy	feasibility study on 3 biopsy samples, low coherence interferometry
Texture / speckle analysis (B-scan)	Gossage 2003, 2006, [181], [185]	Brodatz images, animal tissue, phantoms	9 test, 40 tissue images	proof of principle (classify specimen type)	equidistance probe – specimen ⇒ only for homogeneous bulk solutions
	MacNeill 2004, [187]	blood vessels	160 images, 49 patients	macrophage population as predictor for plaque rupture	unsuitable images excluded manually
	Qi 2006, [188]	upper GI tract	106 images, 13 patients	classify pre-malignancy	unsuitable images excluded manually
	Chen 2008, [189]	upper GI tract	100 images, 3 patients	provide measure for system resolution	ROI marked manually
	Lingley-Papadopoulos 2008, [190]	bladder	182 images, 21 patients	precancer staging	
Segmentation (A-scans)	Gamblicher 2007, [193]	epithelium	114 images	epithelial thickness	ROI marked manually, unsuitable images excluded manually
	Rogowska 2002, 2003, [196],[191]	rabbit cartilage	-	cartilage thickness	-
	Weissman 2004, [192]	forearm, lower leg	12 patients	epithelial thickness	-
	Korde 2006, [194]	epithelium (forearm and upper inner arm)	764 images, 112 patients incl. 448 forearm images	detect actinic keratosis	unsuitable images excluded manually 2 methods
	Hori 2006, [198]	infundibula of cheek, forehead, forearm	5 volunteers	density of infundibula and role as acne risk contributors	surface recognition for flat interfaces
	Bonnema 2008, [200]	cell cultures on blood vessel mimic (BVM)	12 BVM in total, longitudinal study	monitor stent endothelialization	exploits characteristic reflection and shadowing of metallic material
Classification	Zysk 2006, [206]	breast tissue	4015 A-scans, 4 patients	classify normal – malignant – adipose	bubbly appearance of adipose tissue
	Jørgenson 2007, [203]	skin	78 lesions, 34 patients	classify non-melanoma skin cancers	features measured manually
	Cheng 2006, [201]	fake/normal fingerprints	10-20 scans per finger, 8 fingers plus dummies	detect forgeries	no mention about layer selection

5.4 Bulk optical properties

Quantitative measurements from OCT data are a first step for establishing parameters and features for automation.

5.4.1 Quantitative measures

Examples of quantitative measurements from OCT data are work on the scattering properties and refractive index of skin [172] and tissue [173], monitoring blood glucose levels [174], differentiating atherosclerotic arteries [175],[176] and brain tissues [177] [178], or the determination of optical parameters of cervical tissue [102].

5.4.2 Blood glucose

Monitoring blood glucose levels is a challenging topic. Esenaliev *et al.* report an inverse association between the slope of an averaged A-scan and the blood glucose concentration [174]. With the sample properties being related to OCT data, it is possible to determine and classify them automatically. In the original paper, the way of determining the slope has not been described in detail. In a more recent publication a strict imaging protocol ensures tight contact of a sapphire glass window as part of the probe tip with the tissue surface. This allows setting the maximum peak of an A-scan as the approximate surface point, which is a prerequisite for the slope determination [179].

5.4.3 Atherosclerotic lesions

Levitz *et al.* presented an algorithm for determining the scattering coefficient, μ_s , and the anisotropy factor, g_{eff} , from regions of interest [176]. For the measurement, the average A-scan of a region of interest (ROI) area was calculated, and further smoothed by a median filter. In order to allow averaging, the biological specimens in this study were placed under a glass surface. The ROI had to be selected manually for this algorithm. This algorithm was tested for differentiation between normal and atherosclerotic arteries. 151 ROIs from 14 images could be classified into four clinically relevant disease states. An algorithm presented by van der Meer *et al.* in a proof-of-principle study on 20 lesions of 13 samples showed that the light attenuation coefficient μ_t differs for lipid-rich regions, media, calcifications, and thrombi in ROIs [175]. There is no mention how the ROIs were selected and presumably this was done manually. An

interesting aspect is that the approach allows a localised determination of μ_t and visualisation as overlay over the B-scan, thereby providing a differentiation that would not be obvious from the intensity image.

5.4.4 Brain tissue

Jeon *et al.* used the maximum reflectance intensity and attenuation rate of the mean of 5 A-scans to determine the different properties of different brain tissue types of rat brain. Distinguishing white and grey matter might contribute to guiding deep brain stimulation probes in treatment scenarios [177]. The study was performed on 19 manually selected ROIs of 3 images in total. However, four tissue types (hippocampus, primary somatosensory cortex barrel field, external and internal capsule region, and optic tract area) were shown to have different optical properties. A different approach was suggested by Ramrath *et al.*: using a three step procedure, areas within an A-scan that contain white matter are identified [178]. The essential step to recognise a white brain matter peak amongst speckle noise is to transform the A-scan using a nonlinear-energy-operator and a shift-and-multiply operator. Both are discrete counterparts of more complex operators and provide candidates for spikes, i.e. unlikely intensity peaks, amongst noisy data [180]. The identified spikes are validated by matching the attenuation coefficient μ_t within a window after this spike to a reference value for white brain matter. The reference coefficient needs to be provided, e.g. determined manually, but there is no requirement for a pre-defined surface or a ROI. Both approaches were demonstrated on images from small animal tissue. The size difference to a human organ might pose a challenge.

The above mentioned publications have to be seen as examples for a wealth of research communications treating the determination of optical properties using OCT. The crucial point of automation seems to be how to determine the ROIs, or to have prior knowledge about a pronounced feature. Databases of such properties may well serve as reference for automated tissue characterisation.

5.4.5 Epithelial neoplasia

Turchin *et al.* assessed the scattering behaviour of phantoms and cervical tissue with progressing malignancy using various parameters including total scattering coefficient,

variance of a small-angle scattering phase function, and the probability of backscattering [102].

They noted that for the classification of epithelial images, layer disruption is not well associated with progression of precancerous lesions. Continuous layers can be seen in healthy and dysplastic tissue alike, and can be absent in leukoplakia, metaplasia, and neoplasia. They proposed an algorithm which reconstructs scattering properties of the sample. Three parameters, the total scattering coefficient, μ_s , the variance of a small-angle scattering phase function, γ^2 , and the probability of backscattering, p_b , together describe the signal as a function of depth. For a single-layer model, these parameters allow classification in the absence of layers. The calculation for many layers then becomes complex, but for 2-layer samples the computational effort appears manageable. A restriction is that there is an assumption that layers are parallel to the surface. In an approximation, epithelial tissue can be regarded as two parallel layers. Turchin *et al.* created single-layer and 2-layer models. Proof-of-principle was displayed on tissue phantoms consisting of polystyrene beads (0.14, 1, and 4.75 μm) suspended in water or glycerol. In addition to the tissue phantoms, the paper mentions that medical data from two clinical centres has been used to evaluate the algorithm. However, only the results for 19 images from one patient are shown.

Their paper did not specify how images were selected for either the single-layer and 2-layer model, how well the algorithm works for multiple patients or on samples where the precancerous lesions were unknown prior to collection of the OCT image. Such a proof-of-principle is however an important step in the realisation of a clinical classification tool.

5.5 Texture and speckle analysis

The information of OCT images does not exclusively consist of evident structure. Speckle noise carries characteristics which can be exploited for classification and even semi-quantitative analysis [181],[182]. This allows determination of properties of homogeneous bulk solutions of particles smaller than the resolution in the absence of intuitive structural information. Table 5.2 gives an overview of methods which classify OCT data on the basis of B-scan texture.

Table 5.2: Overview of approaches for classification on the basis of texture

	Gossage <i>et al.</i> [185]	MacNeill <i>et al.</i> [187]	Qi <i>et al.</i> [188]	Lingley-Papadopoulos <i>et al.</i> [190]
Objective	Classify 4 animal tissue types, tissue phantoms and texture images	Macrophage density as measure for plaque vulnerability	Classify pre-malignant oesophageal lesions	Classify precancer of the bladder
Valid images	40	160 of 225	106 of 405	182
Pre-processing	<ul style="list-style-type: none"> • Logarithm • Normalisation • Histogram equalisation 	<ul style="list-style-type: none"> • Median filter • Bimodal segmentation 	<ul style="list-style-type: none"> • Manual cap removal • Binary threshold • Centre- symmetric autocorrelation 	<ul style="list-style-type: none"> • Bimodal histogram to determine noise background • Binary threshold
Feature(s)	Spatial grey-level dependency matrix, 2D Fourier transform	Normalised standard deviation	Textural features (intensity differences of centre pixel to its 8 neighbour pixels)	74 texture features reduced to 18 independent features
Classification	Mahalanobis distance	Statistical, ANOVA	“Computer aided diagnosis”	Discriminant function and decision tree

5.5.1 Speckle patterns

In early work, Hillman *et al.* have attempted to establish a correlation of speckle with the concentration of scatterers (polystyrene microspheres) within the probe volume [182]. The reference for the axial depth of a signal was the bottom of the sample – glass slide interface.

Work by Gossage *et al.* was based on the fact that collagen does not show structural features on conventional OCT because of speckle [181]. Speckle noise however does carry information from sub-resolution features. The two representative measures for an image were 1) the spatial grey-level dependency matrix (SGLDM) and 2) a derivative of the complex 2D Fourier transform of a selected image region. The SGLDM in short is a measure of the likelihood of a pixel of a defined grey-level to be in the vicinity of another pixel of a defined grey-level. This then can be expressed by various parameters, e.g. energy, entropy, correlation, local homogeneity, or inertia.

“Brodatz” images are photographs of textures and commonly used in texture analysis [183], examples are widely available in the internet, e.g. [184]. Nine Brodatz test images of specified texture and 10 images each from 4 types of animal tissue (skin, fat, normal lung, abnormal lung) were randomly split into training and classification sets. A Bayesian classification model selected the 3 best out of 24 total parameters. They were then classified according to their Mahalanobis distance. Mean classification rates range

from 37.6 % to 94.8 % for a 3-group-tissue-model, from 64.0 % to 98.5 % for 2-group-tissue-models, and from 97 % to 100 % for the test images. In a more recent study, Gossage *et al.* showed the ability to classify phantoms consisting of particles of sub-resolution size with the same method [185].

5.5.2 Macrophage density

Macrophages play a role in inflammation and tissue degradation and are investigated for their role as contributors to atherosclerotic plaque rupture. Macrophage density could be an indicator for vulnerable plaques. Macrophages have a size of 20 – 50 μm and produce a distinct scattering signal in OCT images. MacNeill *et al.* automatically measured the macrophage density from OCT images [187]. The study consisted of 225 images from 49 patients. 65 of the images had insufficient quality for analysis and hence were manually excluded, leaving 160 images for analysis. The automated segmentation was performed by separating a bimodal histogram. The normalised standard deviation of the signal intensity was calculated from this segmented area and found to correlate with macrophage density, thereby serving as a potential predictor for unstable plaques. Whilst the automated algorithm works on selected images, an operator would still have to select suitable images. As the authors themselves state, this could be a source of bias.

5.5.3 Epithelial lesions and precancers

Qi *et al.* have used centre-symmetric autocorrelation on endoscopic OCT images to automatically recognise clinically relevant features of Barrett's oesophagus, which is a precancerous condition of rising importance in the Western world [188]. Their data consisted of 405 images from 13 patients, however only 106 of these images were used for the study. The presented algorithm required manual removal of the probe cap from the image. The image was then globally thresholded with a binary mask to remove noise areas. This pre-processed image served as the substrate for the extraction of six statistical textural descriptors for the distribution of pixels, and pixels of similar value on an image. These six parameters, as well as 2 principal components of them, were analyzed for their correlation with disease status using a "computer aided diagnosis" which is not further specified. The best results were obtained from the local intensity variation (of 1 pixel and surrounding 8 pixels), and the between-pair intensity variation.

The scores from the principal components analysis (PCA) gave poorer results than the original six parameters.

A publication by Chen *et al.* used 100 images, of normal tissue and Barrett's oesophagus, imaged with both standard resolution and ultra-high-resolution (UHR) OCT, using the discriminatory power of textural features between these disease groups, in order to establish whether there is a statistical measure to quantify the difference between normal and UHR OCT images [189]. The textural features were calculated from ROIs whose selection was not further specified. Features were reduced to 2 principal components which were used for determining 2 linear discriminants. The scores show that UHR images provide a better basis to discriminate normal from dysplastic oesophageal epithelium. These data were not cross-validated and therefore provide a distance measure, not a prediction on the performance in clinical staging.

Lingley-Papadopoulos *et al.* used texture features to assess a progressing lack in structure which may be indicative of bladder cancer [190]. In their study, 18 independent texture features, identified from a pool of 74 features, were used to discriminate 182 images of precancerous stages of the bladder, and achieved a sensitivity of 92 % and a specificity of 62 %. These studies have in common that the textural features are classified using a discriminatory function. More attention on classification is given later in this manuscript.

5.6 Segmentation

The vast majority of literature cited so far relies on manual selection of a ROI or some knowledge about a reference point from where to calculate optical tissue coefficients. Work towards automated segmentation and thickness measurements has been undertaken for cartilage thickness [191], epithelial thickness [192],[193], and sub-surface layer assessment [194], thereby providing a clinically relevant measure rather than a set of images. Table 5.3 shows a comparison of their approaches.

Table 5.3: Comparison of automated tissue layer recognition approaches.

Pre-processing	Gamblicher <i>et al.</i> [193]	Rogowska <i>et al.</i> [191]	Weissman <i>et al.</i> [192]	Korde <i>et al.</i> [194]
Feature enhancement	Averaging	Rotating Kernel Transformation	Factor analysis de-noising Tilt / slant image	Median filter
Feature recognition	Peak – valley distance, or area	Sobel edge detection	Shapelet convolution (several shapes)	Sobel edge detection Skeletonization
Correction	-	Edge linking cost function	Least median of squares	Linear interpolation, Rule for multiple candidates, Exclusion of top 10 (artefact-rich) pixel

5.6.1 Layer thickness

Gamblicher *et al.* report two ways for determining epithelial thickness [193]: 1) Averaging of several A-scans, and taking the first peak – valley distance as thickness measure. This requires a second peak to be present, and they report that 40 % of their A-scans have to be excluded from analysis for this reason. This approach is automated but has poor correlation with histopathology, hence was deemed unacceptable for clinical application by the authors. 2) Manually encircling the epithelium, and determining the average thickness geometrically by dividing the area by its length. This approach is not automated but provided that the epithelium can be marked by a thresholding approach similar to the macrophage approach, would carry potential for automation.

Rogowska's *et al.* work can be regarded as pioneering for dense tissue segmentation from OCT images [191]. They have introduced and explored the Rotating Kernel Transformation (RKT) filter for speckle treatment [195]. The RKT filter emphasises straight lines within the intensity image and facilitates characterisation of the surface and lower boundary of a cartilage layer [196]. It is however computationally expensive; edge linking methods can correct for errors. The edge linking used in this approach is semi-automated as it requires start and end nodes to be set manually.

The approach of Weissman *et al.* aims to recognise the dermis-epidermis junction by convolution with various shapelets (functions based on a Gaussian curve) [192]. Since the contrast of epidermal OCT data was not suitable for this surface recognition, transformed tilt and slant image representations were used instead. Several length scales of the Gaussian shape ensured good convolution results for varying shapes. This pool of

convolution results then allows exclusion of outliers, and the optimal candidate for the position of the junction is determined by a least median of squares function. Since several shapes can be adapted for convolution, this approach is not restricted to tissue interfaces and 2 D scans.

While both approaches are fully automated, it seems interesting that they were not implemented in the more recent works of Gamblicher, or compared with each other. This indicates that, although the methods seem to work well in the way they are presented by the authors, they are not straightforward to implement.

Korde *et al.* have worked towards differentiating skin lesions from actinic keratosis, a precursor of squamous cell carcinoma [194]. Their algorithm utilises parameters obtained from edge detection and employs a surface recognition algorithm. The slope of the mean A-scan was taken as representative for the “average skin attenuation”. The very surface pixels are excluded from this fit as they generally contain strong reflection features. A second structural feature was obtained by using a vertical Sobel edge detection which emphasises horizontal lines. Short edge lines were likely to stem from noise and were therefore excluded. The amount of remaining pixels was defined as “horizontal edge detection”; this could be characterised as a ratio of cohesive layer boundaries per image area. Whereas the “average skin attenuation” did not show a significant correlation with skin lesions, the “horizontal edge detection” had a sensitivity and specificity of 73 % and 65 % for recognising actinic keratosis.

5.6.2 Ophthalmic layers

In ophthalmology, the many interfaces between air, corneal structure, vitreous humour, and retinal layers provide strong OCT contrast. The eye is an ideal organ for OCT imaging owing to its optical properties and layered structures which are thin enough to allow full penetration and are yet wide enough for full resolving. Being able to observe such layers has revolutionised ophthalmology. By providing a means of automating layer recognition and providing a means of objective classification it is possible to enhance the clinical benefit of these features in the process of clinical decision making. Clinically important features are corneal thickness, retinal damage, damage of the fovea or optical nerve, and hence efforts have been undertaken to characterise data automatically.

The success of OCT in ophthalmology might be best illustrated by the progress in automated analysis, which has no equivalent in dense tissue measurements: As an example, we want to cite the work of Fernandez *et al.* who have demonstrated that 7 layers of the retina can be readily characterised from ophthalmic images and that the retinal layers can be registered from the anterior and posterior side [197]. Data processing follows an extensive protocol. A threshold cut-off is used for noise reduction followed by a complex diffusion filter for speckle suppression. From this image the structure-coherence matrix of the image is obtained. The internal limiting membrane is determined by the first inner peak, and the retinal pigment epithelium is determined by the maximum outer peak. Further peaks are correlated with the remaining layers. Outliers are corrected by linear interpolation. This algorithm worked reliably for data from 72 healthy subjects. It can fail in pathologies where initial assumptions on layer composition are not met. However it allows automated generation of 2 D thickness maps of separate retinal layers, and therefore also longitudinal and observational studies. The authors want to point out that this study has to be seen as an example for segmentation algorithms for retinal layers, and the analysis and comparison of them would probably merit a separate review.

5.6.3 Segmentation of dense tissue

Hori *et al.* automated the registration of the infundibula in skin, which are being investigated for their role in the development of acne [198]. The surface recognition was performed by assuming the maximum peak as the surface. This was corrected by a median filter and a so-called “Olympic” exclusion (which excludes a set amount of minimum and maximum values regardless of them being outliers or not). The epidermal junction is defined as the first intensity minimum between the first two maxima of an A-scan.

Both values, surface and epidermal junction, were smoothed with a polynomial function to provide a thickness map of the epidermis. Whilst compared to the more sophisticated shapelet approach this seems to yield rough results, Hori *et al.* claim that the difference between their uncorrected points and the fitted curve was within the magnitude of their system resolution [198]. Such a finding opens the question of whether a – computationally expensive – correction would be necessary in first place.

After establishing the epidermal thickness, a band (volume) below the epidermis was defined. This band was then treated as an *en face* image, a “shadowgram” of this image is characterised by thresholding. From the resulting binary image, infundibula were then characterised as circular structures and extracted using the Danielsson distance mapping algorithm [199]. Distance maps, or distance transforms, are representations of Euclidean distances in a different parametric ordinate system. This facilitates finding set geometries, in this case circles on a 2 D image.

The algorithm is able to deliver epidermal thickness parameters of the imaged volume as well as distribution of infundibula. The surface of such a volume covered $4 \times 4 \text{ mm}^2$, however no information about the speed of the fully automated algorithm was given. Simplifying the determination of the epidermal thickness in contrast to Weissman’s shapelet analysis [192] seems to work well on the volumetric data.

Bonnema *et al.* have used a protocol to determine stent endothelialization on blood vessel mimics [200]. The clinical interest is to monitor the coverage by endothelial cells under different growth conditions. An algorithm thus has to distinguish superficial stent, overgrown (sub-surface) stent, and tissue without a stent mesh. A strut cross-section provides a round shape with a diameter of a few pixels on a B-scan. Characteristic reflection and shadow artefacts from the strut can be exploited, although the protocol remains an extensive set of rules. Briefly, a combination of thresholding and median filtering allows determining sharp (strut) peaks and surface peaks amongst artefacts from debris and reflection. This fully automated approach shows a way of providing clinically relevant information without the need of assessing separate B-scans.

5.7 Classification

We have already mentioned quantitative measures which seem fit for automated classification. An autocorrelation score can be suitable for expressing the likelihood of an expected shape to correlate with an OCT image. When a relation is complex, statistical tools can become essential for automated classification. There are many reduction and clustering algorithms, such as Least Median of Squares, PCA, Genetic Algorithms, or Linear Discriminant Analysis (LDA).

It is common practice to express the likelihood of a data point to belong to a group as a Mahalanobis distance. This is a score representing the distance between groups, taking

into account the distribution of data points of one group. These techniques can be combined, or applied in a decision tree. E.g., the structural features obtained by Lingley-Papadopoulos were differentiated in a decision tree, where in each step a 2-group classification model was designed [190].

Autocorrelation scores of an A-scan with itself have been shown to characterise spoof fingerprints [201]. Fingerprint dummies are created by using materials like wax or silicone rubber with a plasticine mould. These materials are homogenous and therefore have rapidly decreasing correlation scores. For fingerprints, de-correlation happens as well less strongly, and this difference is clear enough to distinguish artificial material from human tissue. There is no mention how exactly the area of an A-scan is selected for autocorrelation and hence we assume this was done manually. However the surface peaks of both spoof fingerprint and tissue can be expected to be pronounced and distinguishable, which should make it possible to automate layer recognition.

Weissman *et al.* used PCA to compare the agreement of their surface recognition algorithm with the line drawn by human operators [192]. Chen *et al.* used a combination of PCA and LDA to compare the performance of two systems with different axial resolution [189]. These studies, rather than assessing clinical parameters such as sensitivity and specificity, used classification as an objective performance measure. Qi *et al.* used PCA for reducing six statistical measures for texture features to two principal components PC1 and PC2 [188]. As mentioned before, this approach did not provide a benefit over using the six-dimensional data.

5.7.1 Glaucoma

Again we refer to ophthalmology for what is possible when feature parameters are readily available: In order to differentiate normal from glaucomatous eyes, Huang and Chen applied several classification methods [202]: LDA, Mahalanobis distance (MD), artificial neural networks (ANN), combined PCA/LDA, PCA/MD, and PCA/ANN. The parameters for classification were extracted from a set of 25 morphological parameters (retinal thickness and size measures). The study involved scans from 89 participants. Inclusion requirements for analysis were: 100 % of the A-scans had to be of good quality and above a specified SNR, and one of the features (cup/disc ratio) had to be big enough for automated detection.

This is noteworthy in respect to dense tissue assessment: Even although this classification approach was developed for images which usually allow extracting a wealth of morphological parameters, some images did not meet these criteria. In dense tissue OCT imaging, outside the ophthalmologic field, it is even less likely to encounter specimens which show a layered structure to this extent. As shown above, the characteristics of dense tissue layers pose a much greater challenge for reliable demarcation. This might be one reason why in dense tissue imaging textural features are candidates for classification.

5.7.2 Skin cancer

Jørgenson *et al.* used 14 image features which were fed into a support vector machine model to classify actinic keratosis from basal cell carcinomas [203]. Although the approach manages to classify 60 of 78 images correctly, there is no exact specification on how features were detected; e.g. “layer thickness” is “measured by ruler on the screen”, or a degree of “graininess” is categorised. The evaluators needed therefore training on feature recognition. However, the characteristics seem to have a degree of correlation with disease. It may be assumed that at least some characteristics can be described by methods mentioned above. This would have the potential for an objective and fully automated classification.

5.7.3 Breast tissue

Breast tissue contains human adipocytes that measure between 50 and 150 μm . These seem to render themselves to frequency based analysis: Ifimia *et al.* describe a system intended to guide fine needle aspirations for breast biopsies via A-scans (“reflectivity profiles”) [204]. Three groups of tissue, (adipose, fibroglandular, and cancerous) have different profiles and the A-scans therefore have a different slope and standard deviation. In this feasibility study three tissue specimens were measured. A study by Goldberg *et al.* was performed on 158 out of 260 scans from 58 patients [205]. In this study, the determination of the slope was automated: Noise reduction was performed by cutting off below the average noise level, determined by the signal from the first 200 μm . The surface was then defined as 100 μm below the first zero crossing of the 1st order derivative, in order to avoid contributions from specular reflection. Slope, standard deviation and spatial frequency content served as parameters to be separated by

a multivariate Gaussian classification model. The performance for a validation set consisting of 86 samples was 91.9 %.

Zysk and Boppart reported three methods to classify A scans of breast cancer biopsies into three relevant groups (adipose, cancerous, stroma); 1) Taking the Fourier domain component of an A-scan and compare with the mean of a training set 2) Periodicity analysis which looks at the distance between high intensity sub-surface peaks to the surface peak, and 3) a combination of both [206]. The periodicity analysis seems to benefit from the above mentioned bubbly structural pattern in OCT images of adipose breast tissue, resulting in A-scans with several distinct spikes. Cancerous tissue has a less characteristic structure and the A-scans are merely a loosely correlated decrease in intensity with increasing penetration depth. They calculated the frequency component for each A-scan and attempted to distinguish three types of tissue: This classification model was trained with data of three patients; data of a fourth patient was used for prediction. The number of available A-scans was 4015 in total (1666 adipose, 1408 cancerous, and 941 stroma).

Adjacent results were clustered to determine the presence of A-scans which had been classified as cancerous within a B-scan. An interesting approach mentioned in this paper is the classification of sliding windows on an A-scan, thereby allowing classification of different sub-surface areas.

5.8 Discussion

5.8.1 Common patterns and suggestions

There is a common pattern of five steps involved in extracting information from OCT data: Obtaining the measurements, pre-processing, feature registration, data reduction, and classification of data.

Pre-processing is the essential part of automation in feature description and effectively increases robustness for further analysis: A common pattern is de-noising, feature recognition, normalisation & refining. There are principally two approaches for evaluating OCT data: Evaluating features of the A-scan, or applying image analysis tools to B-scans. Knowledge about the specimen surface is a common requirement. This can be achieved by the measurement protocol, e.g. by pressing a surface with a glass

slide, or computationally. A frequently encountered method seems to be the crude cut-off threshold that is determined by the noise level above the specimen.

The ideal choice of pre-processing and analysis method depends on the sample type, the scan mode, and the actual research question asked. However it is possible to draw a variety of conclusions from automation algorithms considered in this review: All fully automated approaches rely on a combination of several steps. A-scan based analysis usually profits from averaging several A-scans in order to increase the SNR. Outlier A-scans can be corrected for by using a sliding median window or Olympic exclusion (removing a set of extreme values). Some publications disregard the first 10 pixel or $\sim 100 \mu\text{m}$ of a surface due to reflection artefacts [194],[205]. This seems appropriate when the sub-surface data, rather than the exact surface location, is relevant. Textural features seem to show a potential for distinguishing pathologies that are related to a vaguely progressing lack of structure, where sifting B-scans would become a straining task. Cross-correlation techniques need to employ several shapes to cover biological variation [192],[201]. Several publications exclude unsuitable data manually. For proof-of-concept this is acceptable, but this carries the potential of introducing bias and should be based on an objective parameter, e.g. overall intensity or dynamic range of an image. Statistical methods therefore have a large potential for helping automated classification of OCT data.

5.8.2 Outlook

Automation is a way of helping an operator with image interpretation. Unfortunately, algorithms cannot be expected to achieve 100 % accuracy. This should not be regarded as a major disadvantage: Even if data evaluation is only able to provide clear results from data with sufficient diagnostically relevant contrast, resources can be channelled to concentrate on the more difficult cases. Furthermore, automation has the potential to provide an objective measure. A classification performance lower than 100 % can still outperform human judgment, where for some conditions agreement can be considerably lower: Studies assessing the consensus of expert histopathologists on the same study material, which represent the gold standard in tissue grading, have found kappa values (which illustrate the chance-corrected agreement) as low as 0.36 between eight experts for cervical biopsies [207], or 0.49 between three experts when classifying oesophageal

specimens [153]. Such statistics become relevant when assessing how accurate a method would need to be before it can be used as a clinical tool. E.g., when trying to replace biopsy-dependent histopathology with a realtime diagnostic tool, the novel tool may have a similar or slightly worse performance, but may have logistic benefits.

Acquisition speed and resolution, which presented early challenges in OCT development, have experienced remarkable improvements. Limitations caused by data transfer rate, storage, and computing power are ceasing to be an issue: volumetric movies (4-D OCT) can become standard. This is a double edged sword; with the possibility of obtaining vast data sets in near real-time, it becomes inherently impractical for a human operator to evaluate this amount of data, and virtually impossible to do so at the rate of data acquisition. OCT allows real-time data acquisition, and it would be a pity not to exploit the speed benefit of computational approaches to provide a real-time diagnosis. In private conversations, the authors gained the impression that at least some OCT experts seem to have the opinion that the lack of real-time diagnosis is not a restriction. The authors want to emphasise that – even if this opinion is not commonplace – we strongly feel that information should be available as quickly as possible. For instance, topical diffusion of contrast agents takes time, therefore ruining the advantage of having a real-time imaging modality. There are pitfalls for automation, some papers find that assessment by a trained operator fairs better than the unsupervised approach [193],[194]. Algorithms do, however, have the potential to run in parallel to the scanning procedure and to provide an objective support for clinical decision making in real-time. This might even make displaying the OCT *image* redundant, much in the way spectroscopic techniques may show maps rather than raw spectral data. One field of interest today is therefore data evaluation in near real-time (Figure 5.2).



Figure 5.2: Scheme of landmark developments for OCT. TD: Time domain OCT, FD: Frequency domain OCT; 3 D: Volumetric OCT. The present challenge is automated data evaluation in near real-time.

After all, this manuscript would not have been possible without all the efforts towards automated classification of OCT data, which may be the next chapter in the success story of OCT.

5.8.3 Acknowledgements

This work is funded by the Pump Prime Fund of Cranfield University and Gloucestershire Hospitals NHS Foundation Trust. The authors would like to thank Professor Hugh Barr and Ms. Joanne Hutchings for kindly providing the oesophageal sample for Figure 5.1 (ethics approval by Gloucestershire LREC), and Dr. Catherine Kendall and Mr. Martin Isabelle for proofreading this manuscript.

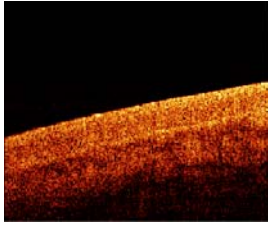

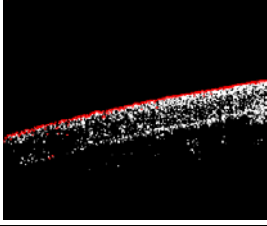
Dr. Nicholas Stone holds a Career Scientist Fellowship which is funded by the UK National Institute of Health Research.

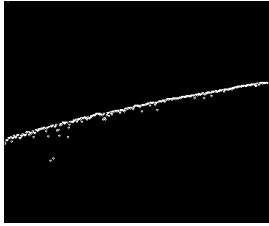
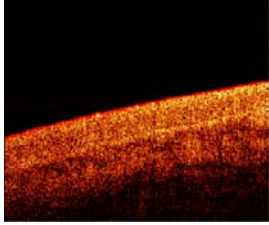
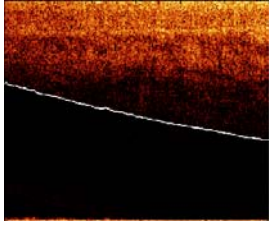

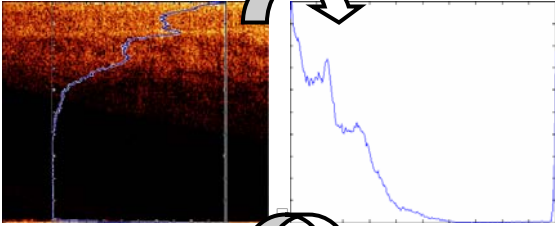
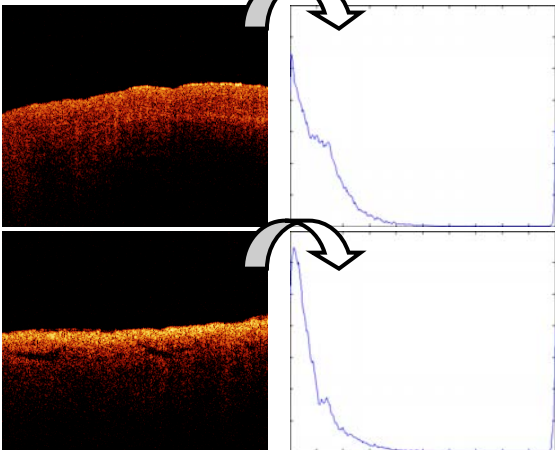
6 Designing a classification algorithm

6.1 Introduction

Classification of OCT data is an emerging field and as a consequence there are a lot of makeshift approaches (Chapter 5). The rationale behind the choice of pre-processing steps is often difficult to establish. This is fair at a proof-of-principle stage: certain assumptions have to be made in the view of limitations of computing power, programming language specifications, and quality and comparability of the data that is to be analysed. This is by no means a bad thing as often it's only worthwhile to put resources into improving an approach when the proof-of-principle has been displayed. For this reason, the following section describes how features are extracted from OCT data so that they can be used for Principal Components Analysis. For clarity, Table 6.1 shows the steps of the pre-processing algorithm and which Chapter describes them in more detail.

Table 6.1: Overview of approaches for classification on the basis of texture

Visual representation	Description
	Chapter 6.2.1 The image, denoised using a cut-off
	Chapter 6.2.2 Binary mask with a intensity threshold (here a 0.95 threshold, meaning that the 5 % most intense pixels are white)
	Chapter 6.2.2 The top values in each row are assumed as surface

	<p>Chapter 6.2.3</p> <p>The surface from the binary threshold. The effect of outliers may be mitigated using a sliding median window</p>
	<p>Chapter 6.2.3</p> <p>The surface as overlay on the B-scan</p>
	<p>Chapter 6.2.4</p> <p>The B-scan normalised to the surface using the surface vector, here with a white line indicating the original lower border</p>
	<p>Chapter 6.2.4</p> <p>The surface normalised B-scan as substrate for analysis. This is equally applicable to volumetric data.</p>
	<p>Chapter 6.2.5</p> <p>In this case, the mean A-scan is chosen as descriptor.</p>
	<p>Chapter 7</p> <p>The idea is now that different structures (which result from different pathology) can be represented by different descriptors. These descriptors serve as vectors to be fed into multivariate analysis.</p>

6.2 Feature extraction

Having learned from the previous analysis attempts, it seems worthwhile to investigate whether a classification algorithm based on A-scans would allow separation of different tissues. Similar to spectroscopic techniques which require shifting of a spectrum as normalisation, we expect that A-scans need to be normalised to the surface. Figure 6.1 displays 2 essential steps to achieve this: Firstly, the surface vector needs to be established. Secondly, this surface vector needs to be applied to the sample data in a process which we shall refer to as “wrapping”.

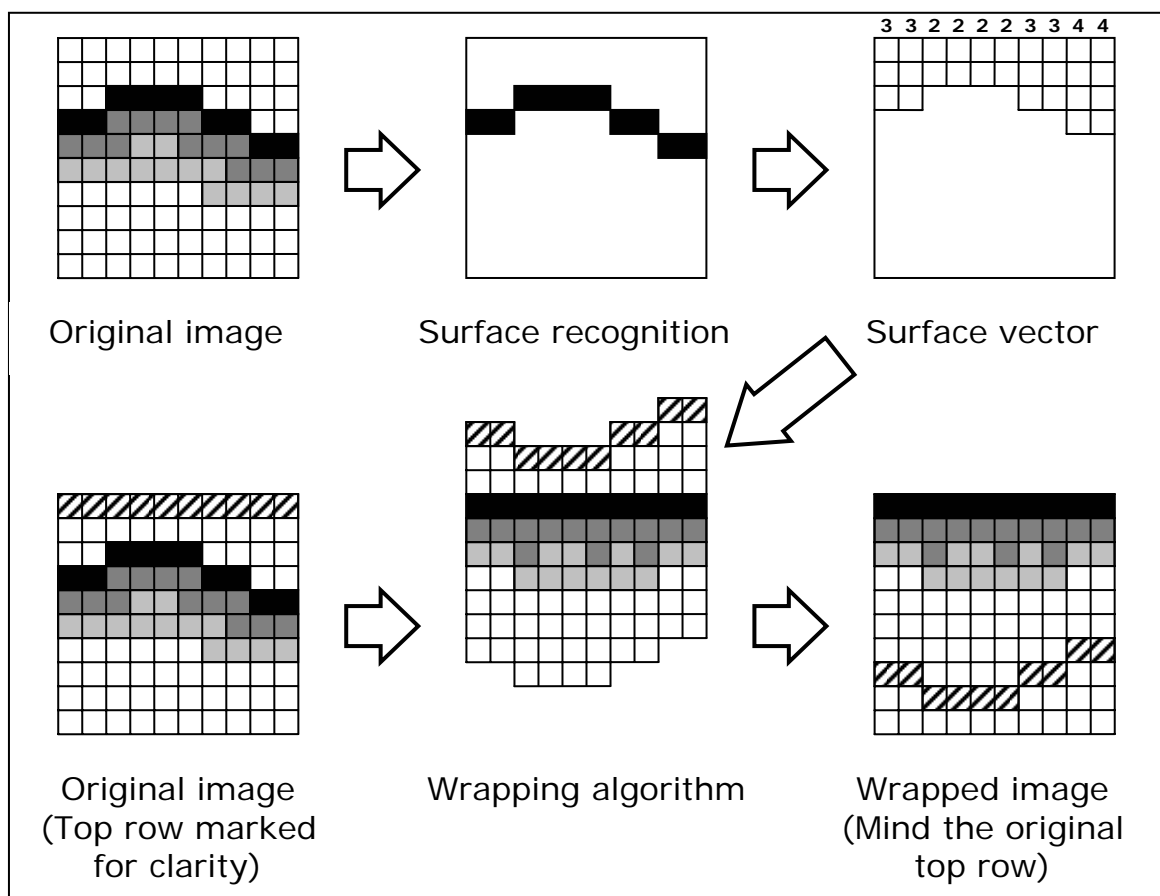


Figure 6.1: Scheme of surface normalisation, i.e., normalising an OCT image to the surface. The 2 steps are: a) surface recognition and b) the wrapping algorithm.

Although very precise algorithms do exist, they can be very time consuming and therefore delay what is only one of many pre-processing steps.

6.2.1 Noise reduction

All images used in this study measure 300×500 pixels, covering 3 mm width and approximately 5 mm depth⁶. The log of the intensity data (time domain signal) was stored as a matrix. Values were handled in double precision format. For the automated noise reduction, a ribbon of “air”, intensity information above the surface, was regarded as the background noise level [197], as indicated in Figure 6.2 (top row).

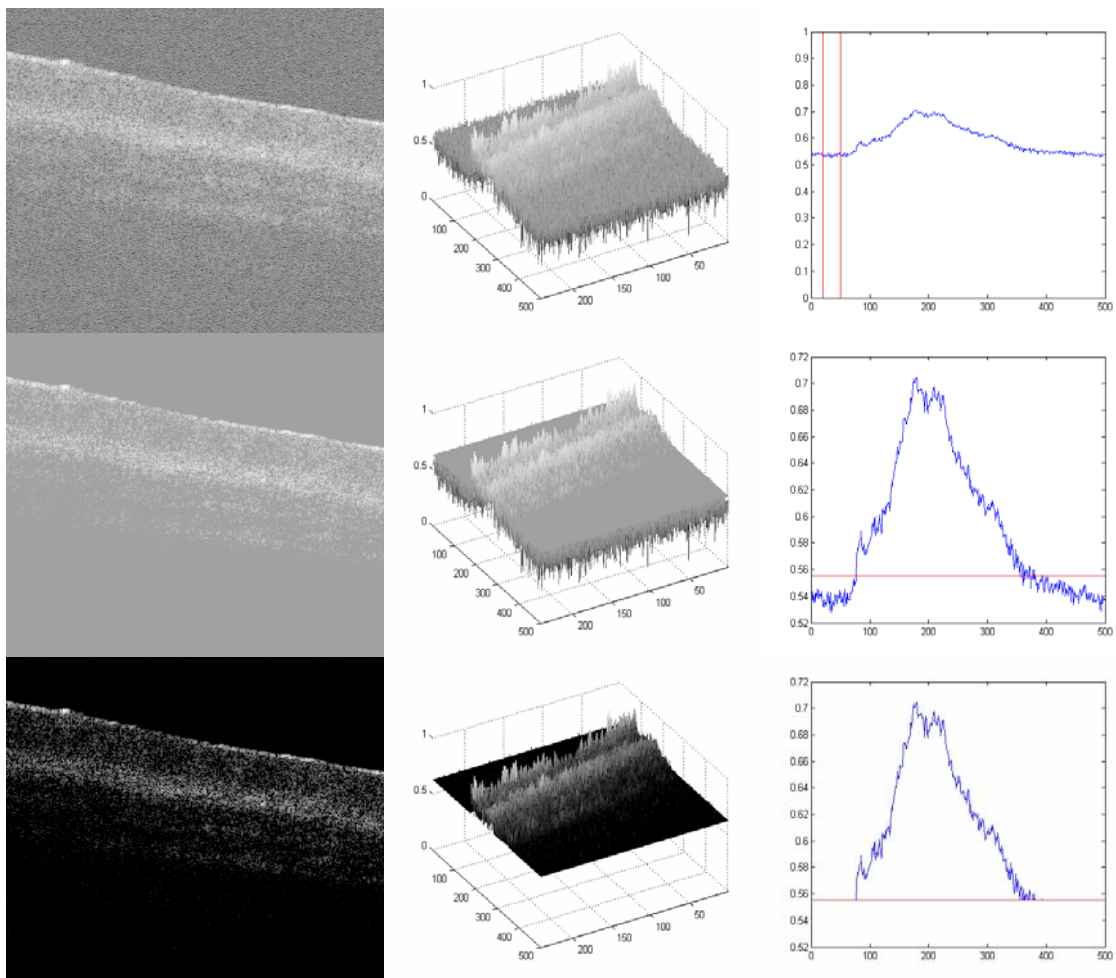


Figure 6.2: Noise reduction by cutting off the noise level Top row: original noisy image, 3 D intensity matrix, and mean reflection profile. The two red lines enclose an area above the surface which is considered as noise level. Middle row: The noise level determined from the top row plus $1.5 \times$ its standard deviation. Note that the profile shows the same graph, at a different scale. Bottom row: the cut-off has been applied and the dynamic range normalised.

⁶ This is not the penetration depth within the specimen, but the height of the scanned window as allowed by the distance change achieved by the reference mirror.

The mean of this plus $7 \times$ the standard deviation of this background signal was used as the intensity threshold. This relatively high threshold does crudely cut off low level signals from deeper areas (Figure 6.2, middle row, which shows the cut-off with the scale and colormap of the top row, and Figure 6.2, bottom row, which has a normalised scale and colormap). We found it, however, to be more important for our subsequent surface recognition to have a noise-free image, with the surface not being obscured by artefacts or layers.

6.2.2 Surface recognition

When using the structural time domain information as basis for classification, it is essential to ensure that the image rows equal the depth position within the tissue.

Surface recognition belongs to the family of pattern and boundary recognition within the field of digital image processing. As a rule of thumb, but not necessarily, exact results inherently require more processing time than an estimate. When classifying OCT data, which easily fills gigabytes of storage, processing speed is an issue.

For OCT there is no established method for unsupervised surface recognition. This is not straightforward as the point with the highest intensity within one A-scan is not necessarily correlated with the sample surface. This is also true for taking the first derivative of the intensity; the biggest change in intensity is not necessarily at the surface. Were this the case, surface recognition could be done quickly and elegantly using a weighing of the magnitude and the derivative of the signal intensity over the single A-scan. Specifically for biological samples this is prone to errors (cf. Figure 6.3).

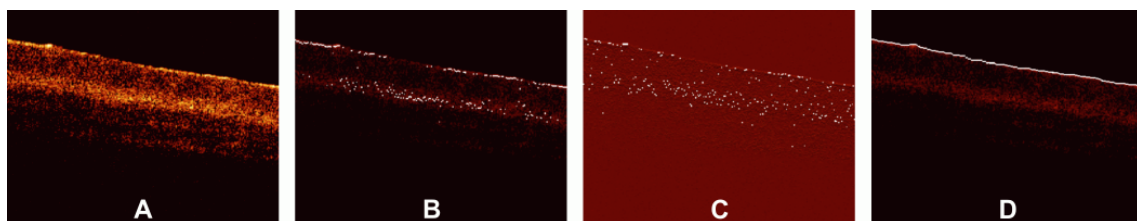


Figure 6.3: A) De-noised B-scan of a biological sample (porcine oesophagus). B) White dots mark the point of highest intensity within one A-scan (overlay over dimmed intensity image). C) White dots mark maximum rate of change in intensity within one A-scan. B & C show that the maximum intensity and derivative do not necessarily correlate with the surface. D) Surface pixels recognised from a binary threshold mask.

There is a small amount of literature about surface recognition in OCT and research groups seem to have methods which seem fit for the purpose of their publication, but are not always useful to implement for a different specimen type. In the literature of the previous Chapter there are methods such as shapelet based boundary recognition [192], Rotating Kernel Transformation (which emphasises straight lines within an image [191], erosion/dilation techniques on a binary threshold image [188], or polynomial fitting through median-blurred intensity peaks [198].

Binary thresholding is chosen throughout this work because it is accurate and fast. A visualisation of the full process is shown in Figure 6.4: The original image may be too noisy for surface recognition (Figure 6.4A). The binary filter creates a rule (stencil) which defines the pixels that shall be chosen for surface recognition (Figure 6.4B). After cutting off noise, it can be assumed that mask pixels are equivalent to tissue pixels, and the top pixels coincide with the surface. The top pixel of each A-scan in Figure 6.4B is therefore equivalent to the tissue surface (Figure 6.4C). It is important to note that this binary mask is only applied for surface recognition, not for the later analysis. The surface vector recognised from the binary image can then be used as correction factor for the original image (Figure 6.4D). The detailed process is described in Chapter 6.2.4.

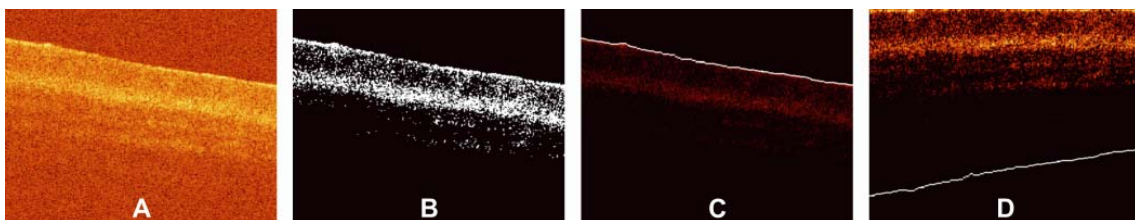


Figure 6.4: Surface normalising: A) Original noisy OCT image (false colour log of intensity envelope data), representative for many *ex vivo* scans: The sample surface, although somewhat straight, is not exactly parallel to the scanning direction. B) Binary image using a 0.90 quantile threshold. C) Surface recognised from B. D) Denoised image after surface normalisation using the vector recognised from the binary image for wrapping.

6.2.3 Refining the surface vector

Binary thresholding may fail due to noise, a superficial layer, or reflection artefacts (cf. Figure 6.5).

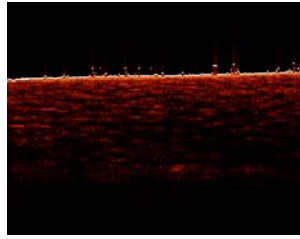


Figure 6.5: Reflection artefacts result in incorrectly detected surface pixels.

Such outliers can be corrected computationally or manually (semi-automated). Manual correction takes time and can introduce observer bias. Automated correction can be computationally expensive and therefore we want to justify and optimise its implementation.

Smoothing is visually pleasing, but that is not an argument for a classification algorithm. When the correct assignment of the surface is necessary, as e.g. for quantitative thickness measurements [196],[192], this is essential for maintaining the quality of results. However, for a classification algorithm, such deviations might not be significant. After all, correction takes time and poses the risk of unduly altering original data.

6.2.4 Applying surface vector correction (“wrapping”)

The application of surface data on the respective A-scans, e.g. transforming Figure 6.4C into Figure 6.4D), is straightforward in programming terms (Figure 6.1). However, the choice of many scientists for a proof-of-principle will probably be some kind of for-loop, in which the shift of data points is calculated for each column separately. This approach can easily become very time consuming. Specifically for large datasets, which are encountered in high-resolution 3 D volume imaging, such for-loops are a bottleneck which would make this approach for real-time imaging impossible. The solution for this is vectorisation, as shown in Figure 6.6.

Many software packages have built in abilities for vectorising code. However, if this is not the case, vectorisation of such an algorithm can put the inexperienced programmer in front of a steep learning curve and it might help having an approach laid out step by step. Hence we describe this in detail:

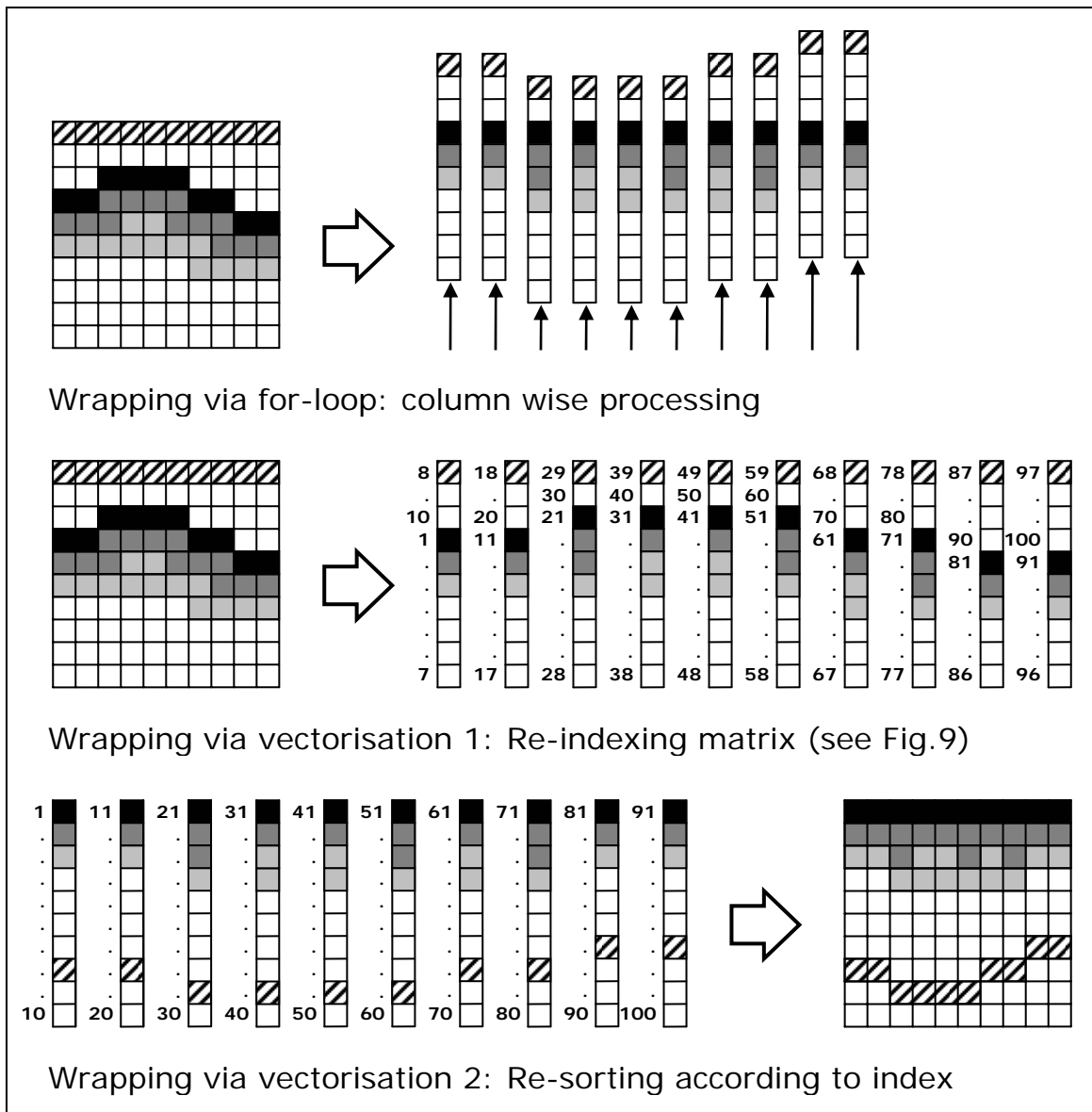


Figure 6.6: For-loops vs. vectorisation. The for-loop processes each column subsequently. Vectorisation reorders a matrix in one step.

We want to shift the z data of an A-scan by a known amount which is not necessarily correlated to any neighbour. The amount of shift is a function of the surface vector or surface matrix which has been established by the surface recognition. In a for-loop, each A-scan would be shifted with the corresponding amount located at the surface vector. For the vectorisation, we sort the data points per column in a way that the reordered sequence can serve for indexing the original data. For our example, let the B-scan be a 10×10 pixel array. A column vector for each A-scan is created. According to the

surface vector, this vector is wrapped around; e.g. if the surface is at pixel 3, the index for re-ordering this A-scan has to be [9 10 1 2 ... 7 8], in order to yield the sequence [3 4 5 6 ... 10 1 2]. This reordering sequence is created for each A-scan, and by using a gradient this can be done simultaneously. Taking the modulus ensures that every reordering vector ranges from 1...10. It is then necessary for the re-indexing matrix to consist of sequential numbers, e.g. a reordering matrix for a 10×10 pixel B-scan will have to range from 1 to 100. A-scan 1 of the 10×10 matrix will be covered by the numbers 1...10, A-scan 2 by 11...20 and so forth. This is realised by adding a gradient; 0 to the first A-scan, 10 to the second, etc.

The result of these operations can be imagined to look like an irregular staircase: Each column's lowest value is higher than the highest value of the preceding column, and no point within a column has got the same value. These sequential numbers are then used for re-sorting the original matrix. The calculation steps depicted in Figure 6.7 are the only ones necessary, and essentially are unaffected by the size of the matrix. The processing time is reduced with growing matrix sizes.

The disadvantage compared to traditional for-loops is that a computer system needs to have enough memory to hold an entire set of data. When the data exceeds the available memory, a combination is the suggested work-around: Splitting data in batches that fit into the memory to be processed in a vectorised way, and performing this vectorisation as loop over these batches.

6.2.5 Benefit

Although the emphasis of this work is on the shape of an A-scan, other features such as frequency content or integral of an A-scan, could be investigated. Further, statistic measures of a group of A-scans can be taken. This approach is therefore straightforward to be extended to volumetric data. Our system was acquiring B-scans and therefore we considered one B-scan as one set of A-scans. If the acquired data were volumetric, a mode of pre-grouping of sub-volumes, consisting of e.g. 25×25 A-scans, could be chosen.

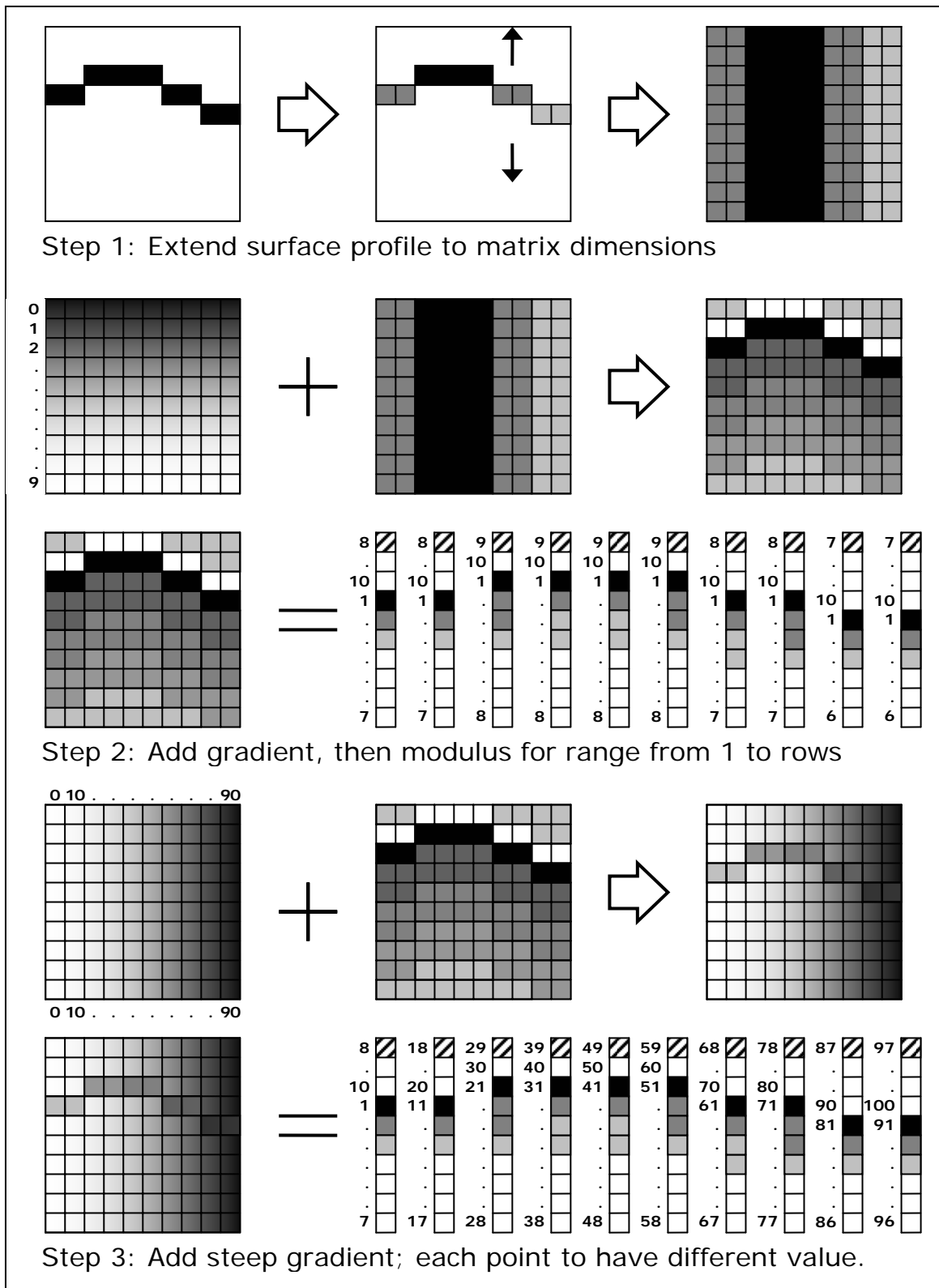


Figure 6.7: Scheme of a vectorisation.

6.3 Optimal parameters for classification

The parameters for noise reduction, thresholding, and further refining such as surface smoothing can affect the accuracy of the surface normalisation and need to be defined. The optimal pre-processing will depend on the actual processing step, a classification algorithm in our case. The optimal parameters for this may be an entirely different from other tasks such as segmentation.

6.3.1 Intensity normalising

A simple approach is to normalise data between the maximum and minimum intensity values. Histogram equalisation on the other hand is a well known, computationally inexpensive method to improve contrast by evenly distributing intensity values over the full data range. This usually optimises the utilisation of the dynamic range. However it can also result in increasing noise values. In medical imaging, contrast enhancing usually means a non-linear enhancement of a clinically relevant feature, which is different from scaling the dynamic range. However histogram equalisation has been reported as pre-processing step for analysing speckle patterns and it might have a benefit for subsequent classification [185], and has been extensively investigated [186]. The difference in contrast between a log of intensity image and a histogram equalised image is shown in Figure 6.8.

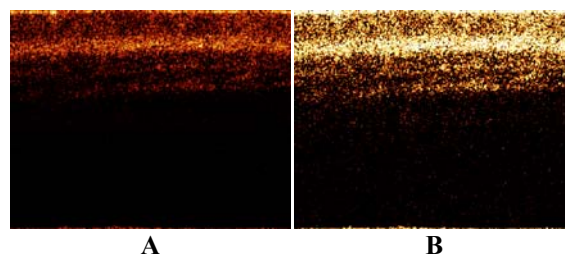


Figure 6.8: Visualisation of histogram equalisation, which is suggested for improving visual contrast. A) Log of intensity normalised sub-volume B) Histogram equalised data.

6.3.2 Cut off axial depth

Large amounts of data may make it necessary to reduce raw OCT data prior to reduction by PCA. Cropping rows of low SNR is such an approach.

Figure 6.9 illustrates how an OCT image can be divided into roughly three sections: 1) the area above the surface, containing pure noise, 2) sub-surface data with sufficient

SNR, 3) the area which exceeds penetration depth, with low SNR. Normalising to the surface results in two advantages: A) Statistical features are easy to extract: the penetration depth correlates with the row number of the image matrix. B) Areas containing a low SNR can be cut off.

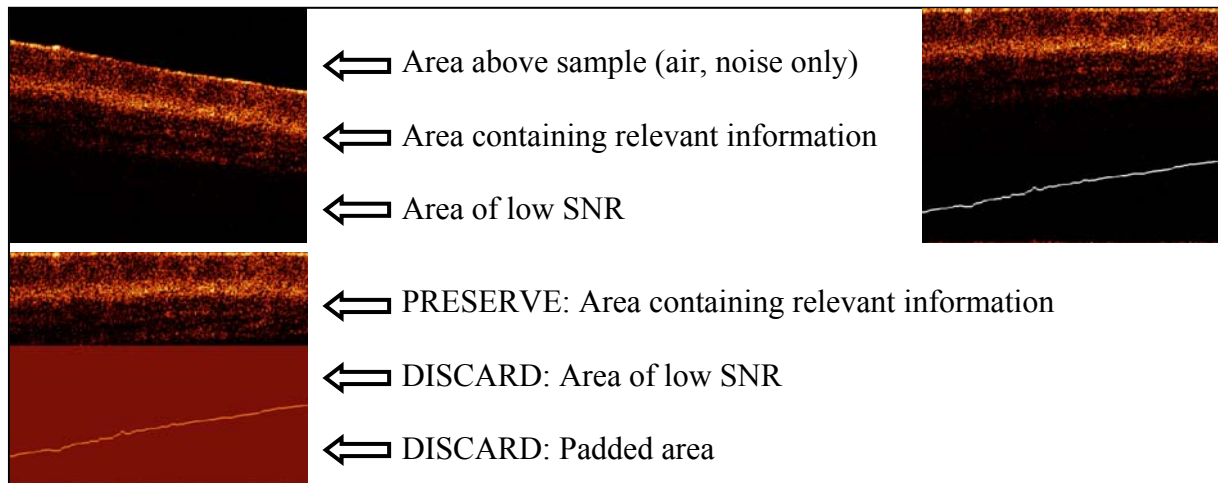


Figure 6.9: Cutting off parts of the image which contain low SNR. The line on the surface corrected images marks the border between original, shifted data and padded matrix points.

6.3.3 Amount of principal components

Principal components (PCs) do not have equal importance for distinguishing groups. Using ANOVA, the significance of a PC in respect to classifying a group can be expressed. Loadings of high significance will contribute more to a correct classification. Adding loadings of lower significance increases the classification performance slightly, and for a large number of data points even a small percentage improvement can be substantial. Adding too many loadings however can reach a point where a model is “overfitted”: Loadings of little significance contribute more noise than relevant information, and therefore decrease the performance.

6.3.4 Summary

Table 6.2: Pre-processing parameters which will be considered in subsequent classification.

Parameter	Description	Expectation
Axial cut-off	Delete values of low SNR Impact on classification performance	Computing time ↓ Memory consumption ↓ Negligible decrease in classification performance
Intensity normalisation	Intensity normalisation vs. Histogram equalisation	Histogram equalisation increases contrast ⇒ better performance
Surface recognition	Benefit of median smoothing vs. raw	Improved surface recognition ⇒ better performance
Amount of principal components	Too little PCs will not allow optimal separation, too many PCs feed noise into analysis	There should be a range of optimum numbers of PCs to be fed into LDA

Table 6.2 lists the pre-processing parameters for OCT images and their expected impact on subsequent classification using principal components-fed linear discriminant analysis. These were assessed using the vegetable data set described in Chapter 7. The findings, which are detailed in Appendix B, can be summarised in the following way:

Intensity normalising results in better classification than histogram equalisation. The value for binary thresholding for surface recognition can be chosen within a range of 90-97 % without affecting results. When a surface vector has been obtained using binary thresholding, there is a benefit from refining the surface. A sliding median window of 3-9 pixels can be generally applied without impairing classification. Cutting of rows of low SNR has little impact on results but has the benefit of reducing computational costs.

Hence the following points can be recommended: 1) using a median refined surface vector for surface normalisation; 2) normalising logarithmic data to the maximum intensity rather than using histogram equalisation; and 3) cutting off rows below the penetration depth for reducing computational effort. 4) Establishing the optimum of components to be used for a subsequent discrimination algorithm.

7 Testing the classification algorithm

Near real-time classification of Optical Coherence Tomography data using Principal Components fed Linear Discriminant Analysis

Florian Bazant-Hegemark^{1,2} and Nicholas Stone^{1,2}

1) Cranfield Health, Cranfield University at Silsoe, Bedfordshire MK45 4DT, UK

2) Biophotonics Research Group, Gloucestershire Royal Hospital, Great Western Road, Gloucester GL1 3NN, UK

Modified[†] from a manuscript that was published in
Journal of Biomedical Optics, 13(3), 034002, (2008)

7.1 Abstract

An optical coherence tomography (OCT) prediction algorithm has been designed and tested on a dataset of sample images (taken from vegetables and porcine tissues) to demonstrate proof-of-concept. Pre-processing and classification of data are fully automated, at a rate of 60,000 A-scans per minute on a 2.8 GHz Pentium 4 computer and can be considered to deliver in near real-time. A dataset consisting of 9 groups of different vegetables was classified correctly in 82 % of cases after cross-validation. Sets of fewer groups reach higher rates. The algorithm is able to distinguish groups with strong visual similarity amongst several groups of varying resemblance.

Surface recognition and normalising to the surface are essential for this approach. The mean divided by the standard deviation is a suitable descriptor for reducing a set of surface normalised A-scans. The method allows grouping of separate A-scans and is therefore straightforward to apply on 3 D data.

[†] Modifications are structural (Chapter numbering, British English spelling) and contextual (omitting redundant parts or higher detail that would not suit a journal publication). There are no modifications to results or interpretation.

OCT data can reliably be classified using Principal Components Analysis combined with Linear Discriminant Analysis. It remains to be shown whether this algorithm fails in the clinical setting, where inter-patient variation can be bigger than the deviations which are investigated as a disease marker.

Keywords: Classification model, Optical Coherence Tomography, Automated discrimination, Medical imaging, Pre-processing, Image processing

7.2 Introduction

7.2.1 Optical coherence tomography

Optical coherence tomography (OCT) is a relatively novel imaging modality first described by Fujimoto and co-workers [42], which is becoming established as a technique for non-invasive, real-time sub-surface imaging at resolutions of 2-10 μm . Non-destructive probing of this type is of interest specifically in medical imaging [89],[94] and also in other fields such as art conservation [208],[209], quality assurance, or homeland security (e.g. finger prints). Compared with other emerging imaging modalities, OCT provides high-resolution and imaging speed for detailed structural information - in vast datasets. We investigate automated classification which is quicker than visually assessing images and likely to be more reproducible.

PCA & LDA

Principal Components Analysis (PCA) and Linear Discriminant Analysis (LDA) are mathematical methods that, on a statistical basis, allow data reduction and clustering into groups [210]. There are papers on applying PCA for OCT-related analysis: Weissman *et al.* used PCA for comparing the agreement of their surface recognition algorithm with the line drawn by human operators [192]. Qi *et al.* used PCA for pre-processing: They were reducing six statistical measures for texture features to two principal components [188]. However, using this approach they reported no benefit over using the six-dimensional data.

Huang and Chen have also applied PCA and LDA for classification, but not on the OCT image itself. Rather, they were using a set of 25 morphological parameters to characterise ophthalmologic pathologies [202]. These parameters were thickness and size measures which could be extracted from the OCT image. In non-ophthalmic OCT imaging such defined layers are usually not observed.

Zysk and Boppart reported three methods for classifying three relevant groups (adipose, cancerous, stroma) in A-scans of breast cancer biopsies [206]: 1) Taking the Fourier transform of an A-scan and using its difference from a reference (the mean of a training set) for the analysis. 2) Periodicity analysis which looks at the distance between high intensity sub-surface peaks to the surface peak, and 3) a combination of both. The periodicity analysis seems to benefit from the “bubbly” structural pattern in OCT images of adipose breast tissue, which however is not present in every type of tissue. One element of this method (taking the FT of an A-scan) we will compare with our method. (We did not implement the periodicity technique). This enabled us to test the Fourier domain technique on sets containing more than 3 groups.

7.2.2 Research questions

In this paper we investigate whether OCT data can be classified in real-time, on the basis of A-scans, without any input of a human operator. We will investigate the classification performance on a small demonstrator group, and plan to assess the impact of surface normalisation. We will further investigate whether representative measures can be established for larger data sets which strain computer resources. Finally, we will investigate the performance on animal tissue.

7.3 Methods

7.3.1 Measurements

OCT device

The OCT device used in this study was a time domain bench top as described in Chapter 2. The light source is a super luminescent diode (SLD, Superlum) with a central wavelength of 1310 nm and provided an axial resolution of $\sim 15 \mu\text{m}$ in air and a $10 \mu\text{m}$ lateral resolution. The rapid delay line is realised using a double pass mirror and a reference mirror mounted on a galvanometer (Cambridge Instruments). The data acquisition was performed using Labview.

Samples

As a demonstrator of proof-of-concept, fruit and vegetables served as readily available samples with varying structure (Figure 7.1). Ten scans were taken of each vegetable; three vegetables of one kind form a group. In addition to this, porcine specimens were

obtained from a local butcher. The exact treatment of these samples was not recorded as the interest at this initial stage was whether different organs would classify at all, and whether a loosely defined different treatment of one specific organ (oesophagus) would be observable through an algorithm. Twenty different samples were prepared from each investigated organ, although all specimens originate from one animal. The size and shape of the samples allowed them to be placed in Petri dishes for the measurements. Where necessary, samples were kept stable with needles.

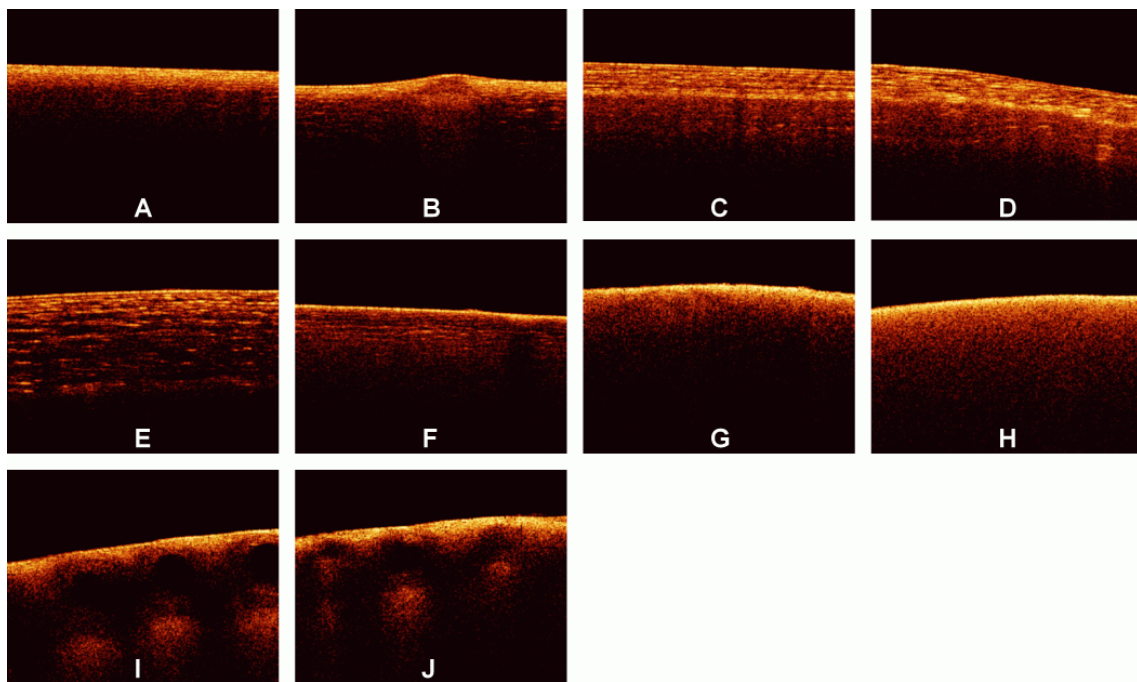


Figure 7.1: Representative OCT images of vegetables (after de-noising, to be fed into surface recognition). A) Celery-longitudinal B) Celery-axial C) Leek-longitudinal D) Leek-axial E) Onion F) Potato G) Mushroom stem H) Mushroom top I) Lemon J) Lime. All B-scans in this publication have a width of 3 mm. The authors would like to emphasise the somewhat similar appearance of C&D, G&H, or I&J.

7.3.2 Pre-processing

The image

The scanning protocol therefore required an area of air above the surface. We aimed to cover approximately $\frac{1}{5}$ - $\frac{1}{3}$ of the image height as this was convenient for manual handling. The surface of the sample was positioned to ensure that the full penetration depth could be depicted on the image. Specimens were placed so that the beam angle would not cause excessive reflections. Whilst these limitations for the distance of the

probe to the sample were chosen in respect to easy handling, no measurement was taken from the same site, as that would spoil the validity of the subsequent correlation.

After thresholding, intensity values of one B-scan were normalised to the maximum intensity peak, by scaling to the maximum peak being unity.

Surface recognition & normalising

Detecting the surface based on a binary threshold mask of the image, as shown in Chapter 6, is an approach we found to be robust for the purpose of our pre-processing. After finding no substantial influence on the result for threshold levels between 0.85 and 0.975, we arbitrarily defined a 0.95 quantile threshold for creating a binary image. The first pixel of this mask was taken as the surface. No smoothing or correction of eventual outliers (e.g. reflection artefacts) has been performed. Normalising to the surface was achieved by shifting each A-scan according to the surface vector.

Of these images we calculated several representative measures of the depth profile of the intensity variation within one image: the mean B-scan, median B-scan, standard deviation, or the mean divided by the standard deviation (M/SD) [211]. Furthermore, we looked into combinations of these, e.g. M/SD and mean as two representative measures for one B-scan. We also used the Fourier transform (FT) of an A-scan in order to be able to roughly compare our results with the approach of Zysk and Boppart [206].

7.3.3 Data analysis

Data handling, pre-processing, PCA and LDA were performed using Matlab V6.1.0.450, Release 12.1 (The MathWorks, Inc.), extended by the “PLS_Toolbox” (Eigenvector Research, Inc.) and our in house tools (“Classification_Toolbox”). Emphasis was put on speed and therefore execution times were assessed using the “tic”-“toc” functions and the “profile” function in MatLab. Bottlenecks, e.g. the routine to normalise A-scans to the surface, were vectorised (Appendix B).

We did not perform mean centring of the data as we found this to have no effect on the quality of results (data not shown). Using PCA, the data was reduced to the first 25 principal components (PCs). Assuming that the PC values follow a normal distribution, they were then sorted in order of their statistical significance, characterised by their respective F values determined by ANOVA ($p < 0.001$) [211]. (The number of PCs (25) was chosen arbitrarily and is not fundamental to the method. The amount of PCs does

affect the results (data not shown), but for the purpose of this publication, whereby different approaches are studied, a constant of 25 PCs are a fair representation.)

These 25 PCs were then used in their entirety for LDA, which is inherently able to suppress loadings of lower significance for the model. The training model was tested using leave-one-out cross-validation. For a small group (30 scans), one measure was left out: one A-scan when classifying separate A-scans, or: one representative measure when classifying the representative measures (leave-on-scan-out, LOSO). While for the LOSO cross-validation, each prediction data set is “unknown” to the algorithm, the training model can still include data from the very same sample. A better cross-validation approach is therefore to test a classification model with a prediction set whose samples are “unknown” in their entirety. This is achieved by splitting the data into a training group and a prediction group (leave-one-group-out, LOGO): For the large sample groups (330 scans) we created a training model from 220 randomly chosen scans, and evaluated the prediction performance for the remaining 110 scans which are unknown to the test model. These 110 scans were from different samples than the 220 training scans. This way, cross-validation could be performed on 3 different combinations. The mean of the results for the 3 cross-validations for one set of samples we accept as classification performance. For comparing pairs of similar vegetables and the animal samples, a prediction model was created using a randomly assigned half of the B-scans of each group and cross-validated using the other half of the B-scans. Also here, test and prediction groups were from different samples, and the mean of the 2 results was taken as overall performance.

7.4 Results

The images appear typical of common OCT image quality. Representatives of fruit and vegetable structures are shown in Figure 7.1. Distinguishing such images, when they are presented next to each other, is fairly easy. However, when working through a large quantity, some of these could start to look similar to an insufficiently trained or overworked eye.

The results section is divided into two groups: Part 1 shows the effects of pre-processing and shall demonstrate the rationale for choosing our parameters. Part 2 shows the actual classification results.

7.4.1 Part 1: Pre-processing

Surface recognition & normalising

Vectorisation reduced the processing time roughly by a factor of 4 to 5. The visual results of the unsupervised surface normalising have effectively been shown in Chapter 6. Figure 7.2 shows how normalising to the surface affects the mean and the M/SD of a B-scan.

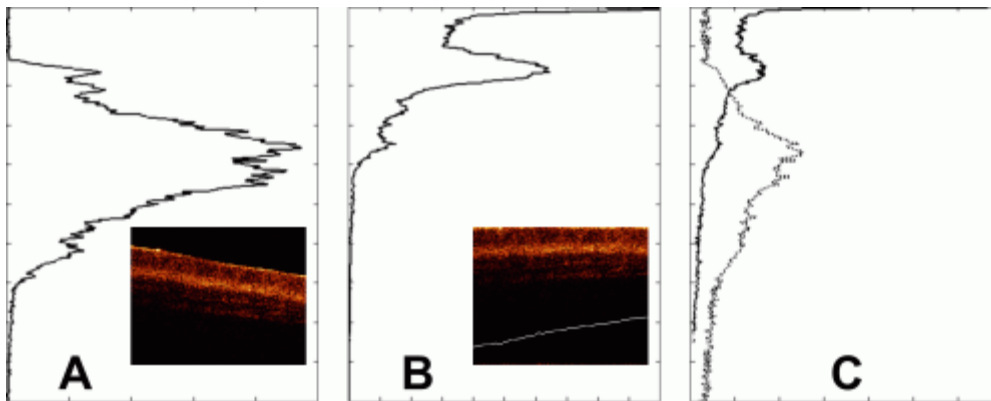


Figure 7.2: A) Mean A-scan of original data. B) Mean A-scan of surface normalised image. C) M/SD along row of A-scans; (—) surface normalised, (···) original.

Vegetables – separate A-scans

In a first approach, we used 10 images of 3 vegetables, resulting in 30 images or 9000 A-scans. Performing PCA on the surface normalised A-scans (Figure 7.3A) or not surface normalised FT curves (Figure 7.3B) shows a tendency to group but also considerable overlap. Please note that the plots in Figure 7.3 show only the 2 most significant (ANOVA) PCs for classification against each other. The subsequent LDA separates on the basis of a pool of several PCs, (25 in the examples of this study).

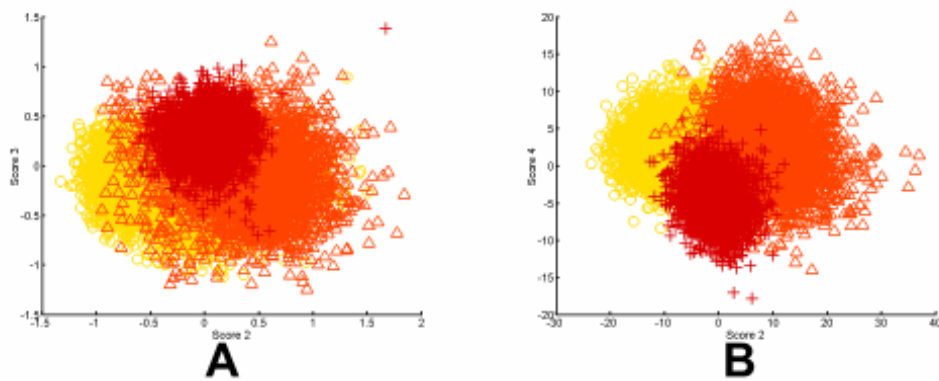


Figure 7.3: PCA of 3 vegetables, 10 images each. The plots show the PCs of 9000 A-scans with the 2 highest scores (determined by ANOVA) out of 25. A) Surface normalised. B) FT of A-scans (no surface normalisation). Legend: o...celery, Δ...onion, +...mushroom.

Table 7.1 shows the performance of classification models for this case. For the amplitude of the frequency content, surface normalisation does not improve the classification. For classifying time domain data (spatial information), surface normalisation yields a higher rate.

Table 7.1: Comparison for classification of 3 A-scan representations: The Fourier Transform, the Amplitude of the FT, and time domain data. The dataset consisted of 3 groups of 10 images each (9,000 A-scans). Scores are percentages correctly classified by the algorithm, after LOSO cross-validation.

Pre-processing	FT	Amplitude of FT	Depth profile (A-scan)
Without surface normalisation	65.4	96.1	65.4
Surface normalisation	92.8	95.8	92.8

Data reduction prior to PCA

The classification rates in Table 7.1 are not disappointing, but take a long time to calculate. Hence we reduced the B-scan to a representative measure, as shown above in Figure 7.4. From the mean of these images, the separation is clearer than in Figure 7.3.

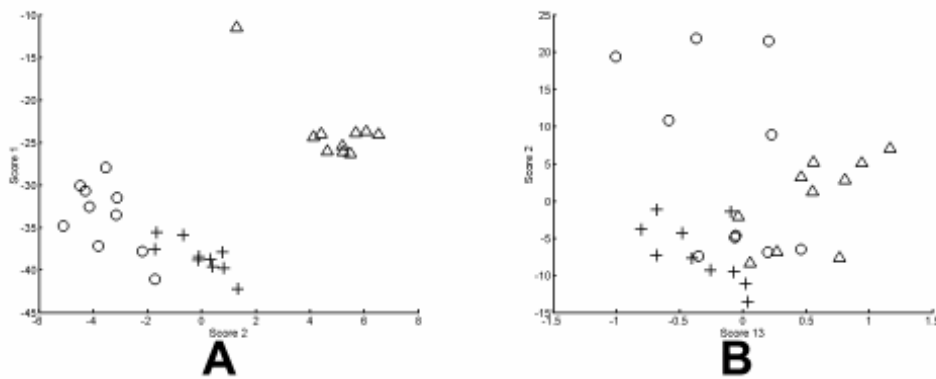


Figure 7.4: PCA of 3 vegetables, 10 images each. The data points show the 2 highest scoring PCs (ANOVA) of the 30 representative measures. A) Surface normalised. B) No surface normalisation. Legend: ○...celery, △...onion, +...mushroom.

For this approach, surface normalisation has a stronger impact than before: Classification was correct in 93.3 % for the surface normalised data (Figure 7.4A), but only 33.3 % for not surface normalised data (Figure 7.4B).

Data reduction – large dataset

A large dataset consisting of 330 images, categorised into 9 groups, was used for assessing various representative measures. These 99,000 A-scans could not be fed into analysis on a Pentium type processor with 1 gigabyte RAM, so it was essential to extract a representative measure. Table 7.2 shows the results for data reduction. For time domain data, models created using the mean normalised by the standard deviation (M/SD) had the best results, as shown in Table 7.2. For models from the FT of an A-scan, the standard deviation (SD) has the highest classification score. After LOGO cross-validation, the model made from the mean values has a higher score.

Table 7.2: Training performance (TP) and cross-validation (CV) of a dataset consisting of 9 groups; 7 groups of 30 images each and 2 groups of 60 images each (330 representative measures). Scores are percentages correctly classified by the algorithm, before and after cross-validation. The highest scores are bold.

Representative Measure	A-scan ¹			FT of A-scan ²		
	TP ³	LOSO ⁴	LOGO ⁵	TP ³	LOSO ⁴	LOGO ⁵
Median	86.7	80.0	81.6	88.8	83.3	82.7
Mean	88.2	84.2	81.6	90.0	86.4	86.7
Standard Deviation (SD)	86.4	77.6	72.9	91.2	85.5	85.1
M/SD	92.1	86.1	82.0	90.0	85.5	81.8

1) Surface normalised time domain intensity data, 2) Amplitude of FT (not surface normalised) A-scan, 3) Training Performance, 4) Leave-one-scan-out cross-validation, 5) Leave-one-group-out cross-validation

7.4.2 Part 2: Classification

From the results so far combinations with high scores were chosen for further comparison: surface normalisation with M/SD, and the mean and the standard deviation of the amplitude of the FT. We will abbreviate those as SN-MSD, AFT-M, and AFT-SD, respectively.

Vegetables

A different set consisting of 330 images, categorised into 7 groups, was classified correctly (after LOGO cross-validation) as follows: SN-MSD: 91.8 %; AFT-M: 93.9 %; AFT-SD: 79.1 %. For comparison, the not surface normalised approach reached 32.0 % in this case. For models consisting of more than 2 groups, a confusion matrix can be helpful for investigating how groups contribute to misclassification. It would take too much space here to provide a confusion matrix for each investigated model, but 2 examples shall be given in Figure 7.5: The left graph shows a very high classification (where a diagonal of 100 % in each group is the maximum achievable), and the right graph displays a higher misclassification rate.

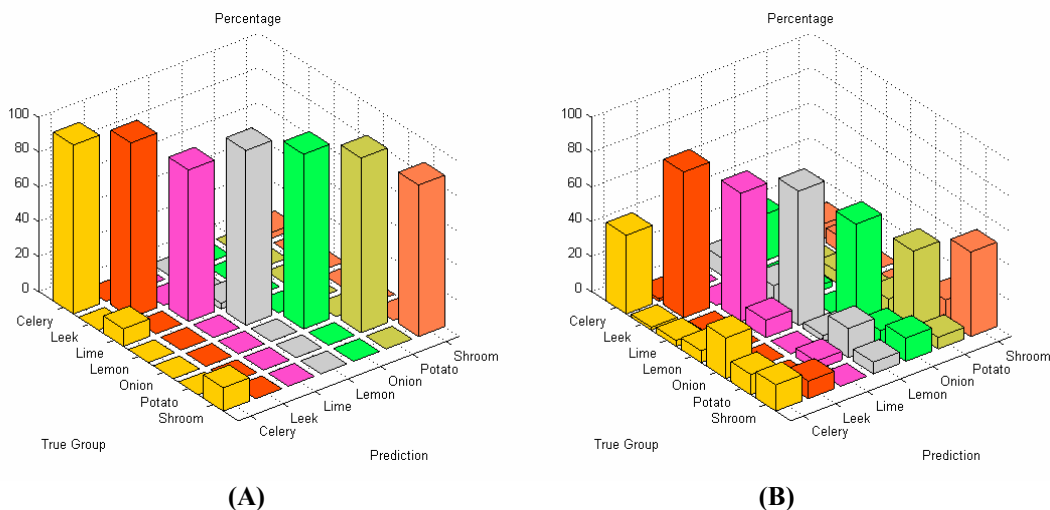


Figure 7.5: Graphical representation of exemplary confusion matrices with A) low, and B) high misclassification rates.

Similar image types

In the next step we assessed the classification of similar images, i.e. measurements of one type of vegetable with different orientation, or similar images (lemon vs. lime).

Representative B-scans are shown in Figure 7.1G-J. Classification results for 30 images each are shown in Table 7.3.

Table 7.3: Classification rates after LOGO cross-validation. Training models were created for pairs of similar images (2 × 30 images for a pair).

Type of similar images	Representative measure		
	SN-MSD	AFT-M	AFT-SD
Celery (2 orientations)	56.7 %	68.3 %	86.7 %
Leek (2 orientations)	61.7 %	63.3 %	80.0 %
Mushroom (2 orientations)	96.7 %	90.0 %	90.0 %
Lemon vs. Lime	83.3 %	95.0 %	95.0 %

Similar images in a large group

If groups are similar, further added groups do decrease the power of the separation. To show this, we grouped two vegetables to yield four groups. Separation is affected when these two groups have to be distinguished not only from each other, but also amongst other data, as shown in Table 7.4.

Table 7.4: Classification rates after LOGO cross-validation. Training models were created for sets of pairs of similar images (2 × 30 images for a pair).

Type of similar images	Representative measure		
	SN-MSD	AFT-M	AFT-SD
2 Mushroom orientations, Lemon vs. Lime	90.8 %	91.7 %	86.7 %
All 4 pairs (8 groups) from Table 7.3	67.1 %	74.2 %	82.1 %

Preliminary tissue pilot study

Extending on the vegetable findings, we used porcine samples for a preliminary tissue pilot study. We measured 20 sites each of four different tissue types (tongue, oesophagus, tracheae, and broncheoli). Representative images are shown in Figure 7.6. The results for PC-fed LDA classification are encouraging: After LOGO cross-validation, the four, albeit clearly distinct tissue types, are classified 90.0 % (SN-MSD), 82.5 % (AFT-M), and 70.0 % (AFT-SD) correctly.

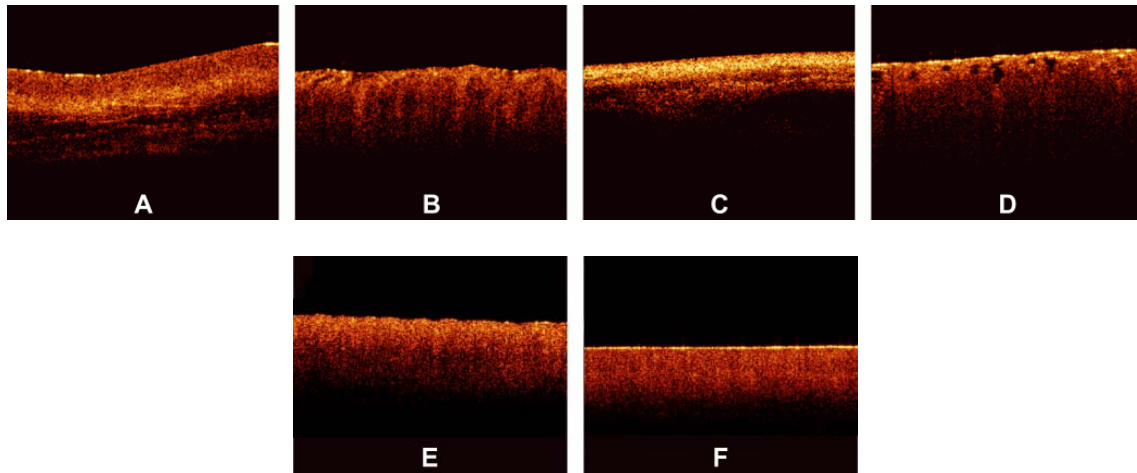


Figure 7.6: Representative OCT images of porcine samples. A) Squamous lining of the oesophagus B) tongue C) tracheae D) broncheoli E) oesophageal tissue after application of 5 % acetic acid as contrast agent F) 5 day old oesophageal tissue. Each image covers 3 mm in width.

In order to see whether more subtle changes can be picked up, we looked at samples from one tissue type (oesophageal) measuring them in a fresh state, after degradation for 5 days and after application of 5 % acetic acid, a contrast agent which is used in clinical settings. These 3 groups can be separated well between each other in 91.7 % (SN-M/SD), 93.3 % (AFT-M), and 76.7 % (AFT-SD) of the cases (LOGO cross-validated).

7.5 Discussion

OCT provides structural images which are intuitive to look at and this might even be an important aside for this technology to become accepted by many users. Evaluation of such structural data seems straightforward; however it remains a challenge in detailed areas. This is evident for an imaging technology which enables one to look closely at areas which previously have been impossible to observe.

Acquisition speed allows generation of high-resolution volumetric real-time data, this is difficult to handle and display meaningfully and will remain complex to interpret in the future: Searching through a vast voxel space can hardly be imagined to be an enjoyable task for a human operator.

7.5.1 Pre-processing

Time domain data and frequency content

Classification requires pre-processing. Both methods, the amplitude of the Fourier transform of an A-scan, and the surface normalised A-scan, give good results. The

Fourier transform which we used for creating a model differed in four points from Zysk and Boppart [206]: Firstly, our de-noising and normalising, as described above, was different. Secondly, we were using the FT of the A-scan for classification, and not the difference from a reference. Thirdly, Zysk and Boppart report that they did perform surface normalising (“truncating”) prior to FT, whereas we found that only to be beneficial when feeding the real and the imaginary part of the FT. However, we achieved better results when using the amplitude of the FT, directly of the non surface normalised A-scan. Finally, we took a representative measure for a set of A-scans and classified the set, rather than classifying single A-scans.

It then probably depends on the sample structure, e.g. presence of distinctive periodic patterns, whether the frequency information or time domain data allows a better classification. Time domain data needs to be surface normalised. This requires a robust and quick surface recognition and means an additional computational step.

We are aware that the data for this paper are suitable for binary thresholding as surface recognition, and that the converse is not necessarily always the case. Such scenarios could be imagined when the surface of interest is covered by a layer of mucus, a sheath, a balloon, or similar. Research into more robust and reliable surface detection algorithms would therefore be beneficial.

Representative measure

Reducing the data to a representative A-scan is necessary; our Pentium type processor with 1 gigabyte memory was not able to handle datasets of tens of thousands of A-scans. More powerful computers are of course available, but it is conceivable that OCT data will stretch these resources to the limit for the near future. We found that the M/SD of a surface normalised B-scan or the mean of the frequency content yield good results. For instance, at separating 2 similar celery orientations, the SN-MSD approach practically fails (56.7 %), whereas the AFT-SD reaches over 80 %. For classifying porcine tissue, the SN-MSD approach had the best result. In the case of variations within a B-scan, such as in Figure 7.7, standardising using the standard deviation does improve results for the subsequent classification.

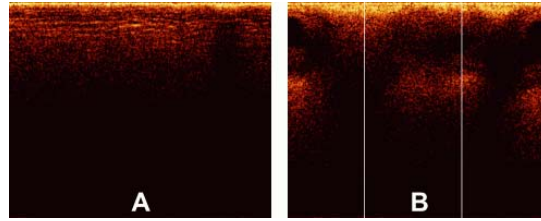


Figure 7.7: A) Surface normalised B-scans: The depth profile is homogeneous across the scan width and can be well described by a mean A-scan. B) Here structures result in different A-scans, e.g. A-scans at the two white lines will be different. In this situation, a standardised mean seems to better represent the B-scan.

In terms of pre-processing, we also investigated mean centring, and using more than one representative measure (data not shown). We have not performed any correction of the surface vector in this study, although that is beneficial (data not shown). Further, the performance of LDA can be improved by feeding an optimum number of principal components (PCs). However, this optimum depends on the number and kind of groups in the classification model. The different sample groups in this study do have different optimum numbers of PCs. For instance, for classifying 4 tissue types, the classification performance, using 25 PCs, was 90.0 % (SN-MSD) and 82.5 % (AFT-M). For these groups, the optimum for SN-MSD would be reached by feeding 11 PCs (97.5 % LOGO cross-validated), and for AFT-M 7 PCs (91.3 %). For the purpose of this demonstration, we decided to keep a fixed amount of PCs. It might well be worth investigating a combination of two or more measures, and refining automatically extracted measures. However, this is beyond the scope of this paper and is probably only worthwhile when refining the technique for a dedicated purpose. Summarising, it is possible to improve results by fine-tuning and combining parameters.

Speed

One criterion for processing is speed: OCT acquires data in real-time and hence the subsequent analysis should not impact on this characteristic. Pre-processing does slightly delay results: On a 2.8 GHz Pentium 4, classification of 330 B-scans takes approximately 85 seconds, of which 73 seconds are needed for surface normalisation and extracting a representative measure. That corresponds to roughly 60,000 A-scans per minute and makes us confident that the claim of providing a near real-time modality remains justified: It would be quite remarkable for a human operator to classify 330

images in 1½ minutes. However we were only using a run length encoded programming language for proof of principle; compiled programs run much faster.

7.5.2 Classification

Robust cross-validation methods are important for testing LDA. It is a powerful approach and for a small enough set of samples some separation into groups can always be expected. For instance, in the example of 30 images, the model created by LDA was able to classify 100 % correctly even without surface normalisation. After cross-validation, the surface normalised model was classified in 93.3 %, vs. 33.3 % for the not surface normalised model.

Relevance for medical data

The fact that the algorithm reaches over 90 % accuracy in certain cases is reassuring when looking at obviously different images, but for our data set the displayed approach is also able to differentiate images reasonably well which are difficult to tell apart by eye. Further, the algorithm is still able to classify similar images amongst a pool of other structural data.

This is important as we plan to apply this algorithm to medical data. There we expect a larger variation amongst healthy individuals, and only eventually the small, but significant alterations for early, non-symptomatic disease. As an example we want to mention epithelial cancers, carcinomas, where the epithelial thickness of the healthy population can vary considerably due to various factors. However, the subtle alterations of proliferative epithelial cells close to the basal layer are probably the feature which can be picked up by OCT scans. It still remains to be assessed whether the algorithm would fail in such scenarios.

Limitations and benefit

In this study the imaging protocol is strict in the sense that it requires the penetration depth to be fully exploited and to leave an area of air above the surface. However, most imaging will have to adhere to a protocol. As long as the operator is made aware of the requirements we believe this limitation not to be substantial. After all, we have acquired thousands of images in this way.

The advantage we see in the presented algorithm is that the few parameters utilised here are robust enough for the analysis to work automated and in near real-time. It does not

require assigning areas of interest. It does not rely on specific structures such as previously reported methods, but can still be used in combination with these algorithms. Hence this algorithm is straightforward to implement for existing OCT data.

It might be worthwhile to emphasise that, although we were taking representative measures of B-scans, this approach will work on volumetric data. For example, taking a representative measure of a volume consisting of 20×20 A-scans on a system with $10 \mu\text{m}$ lateral resolution will allow classification of 25 “tiles” per mm^2 , which is still superb for targeting clinical treatment.

Further work

Previous studies have shown that expert histopathologists do not always fully agree on classification of tissue specimens [153],[212], and an algorithm with lower than 100 % correct classification can still outperform subjective human assessment. In clinical work, OCT data would have to be compared to histopathologic results. Even though there are significant discrepancies found between expert pathologists on particular pathology groups [153],[212], histopathology is defined as the gold standard. Therefore we must base our analysis on the assumption that the sample pathology defined by histopathology is correct. In this study, we have not investigated the performance of the algorithm against a human operator, and hence assumed a 100 % correct classification as reference base line. In the future, we intend to compare classification results with those achieved by histopathologists on the same OCT images from biological specimens.

The next step is testing this algorithm on medical OCT images: We expect it to distinguish structurally different pathologies correctly. In ongoing work we are investigating how well it classifies disease groups with quite similar appearance, such as mild and moderate dysplasia in early cancer development.

7.5.3 Acknowledgements

This work is funded by the Pump Prime Fund of Cranfield University and Gloucestershire Hospitals NHS Foundation Trust.

The work presented in this article would not have been possible without the initial support of Prof. Ruikang K Wang, Dr. Shing “Sammy” Cheung, and Dr. Laurie Ritchie.

8 Applying the algorithm to clinical data

Optical Coherence Tomography (OCT) Imaging and Computer Aided Diagnosis of Human Cervical Tissue Specimens

Florian Bazant-Hegemark^{1,2}, Mike D Read³, Keith McCarthy⁴, Ruikang K Wang⁵, and Nicholas Stone^{1,2}

- 1) Cranfield Health, Cranfield University at Silsoe, Bedfordshire MK45 4DT, UK
- 2) Biophotonics Research Group, Gloucestershire Royal Hosp., Great Western Road, Gloucester GL1 3NN, UK
- 3) Women's Health Directorate, Gloucestershire Royal Hospital, Great Western Road, Gloucester GL1 3NN, UK
- 4) Department of Histopathology, Cheltenham General Hospital, Sandford Road, Cheltenham GL53 7AN, UK
- 5) Department of Biomedical Engineering, Oregon Health & Science University, Beaverton, OR 97006, US

Modified[†] after a manuscript that was published in
Proceedings SPIE **6627**, 66270F, (2007)

FBH designed, performed, and was principal investigator of the study, recruited patients, performed the OCT measurements, designed the algorithm, and wrote the manuscript. MDR recruited patients and obtained tissue samples. KMcC provided the histopathological diagnosis. RKW supplied instrumentation design and technical supervision. NS was clinical supervisor and contributed to algorithm design.

8.1 Abstract

The keyword for management of cervical cancer is prevention. The present program within the UK, the “National Health Service (NHS) cervical screening programme” (NHSCSP), is based on cytology. Although the program has reduced the incidence of cervical cancer, the treatment decision is based on diagnostic biopsying and requires follow-ups. There is a potential for reducing costs and workload within the NHS, and relieving anxiety of patients. In this study, a fibre optics time domain optical coherence tomography (OCT) system (1.3 μm centre wavelength, $\sim 15 \mu\text{m}$ resolution) was investigated for its capability to improve this situation. Tissue samples were obtained according to the ethics approval by Gloucestershire LREC, Nr. 05/Q2005/123. 1387

[†] Modifications include: British English spelling; cross-referencing; removing redundancies; and a greater level of detail.

images from 199 participants have been compared with histopathology results and categorised accordingly. Our OCT images do not reach the clarity and resolution of histopathology. Further, establishing and recognising features of diagnostic significance seems difficult. Automated classification would allow one to take decision-making from the subjective appraisal of a physician to an objective assessment. A model developed for this purpose can distinguish normal/CIN1 from CIN2/3 with 61.5 % accuracy ($\kappa = 0.52$). Confounding factors and options for improving results are analysed.

Keywords: OCT, epithelial malignancy, cancer prevention, computer aided diagnosis, precancer classification

8.2 Introduction

The management of cervical cancer in the western world may be regarded as a model for cancer prevention management. Screening and early diagnosis have indeed reduced the mortality (Chapter 1.4). However, decision making is still widely based on diagnosing tissue biopsies. Histopathology of these tissue samples is the gold standard, yet it is time consuming and invasive (Chapter 1.4).

Various emerging modalities that might play a role in improving this problem have been assessed, but are not yet introduced into the clinical setting. Amongst these technologies are non-destructive optical probes. OCT has been proposed as a method for investigating early epithelial cancers in the late 1990s [89],[94].

8.2.1 Cervical cancer

What we should recall from Chapter 1.2 is that the most prevalent form of cervical cancer is squamous cervical carcinoma, which is preceded by a long phase of preclinical disease. Dysplasia does not necessarily lead to cancer, but with increasing severity the likelihood of developing cancer rises. These early changes do not cause symptoms which would be noticeable by the patient. A large part in making the decision to surgically remove a precancerous lesion is based on the presence of CIN2 or worse; CIN1 does not usually justify treatment. However, CIN1 and CIN2 have similar markers in the colposcopic examination and therefore clinicians tend to rely on the opinion of a histopathologist.

8.2.2 Non-invasive probing

Emerging technologies which have been investigated for improving this situation include fluorescence spectroscopy, elastic light scattering, confocal microscopy (Chapter 4.3.1), infrared spectroscopy (Chapter 4.3.3), or Raman spectroscopy (Chapter 4.3.4). In OCT, there are a few studies which support its potential role in cervical precancer assessment (Chapter 4.3.2). However, none of the studies have resulted in a clinical tool that makes biopsies redundant (Chapter 4.4).

8.2.3 Investigation in this study

There is no commercial OCT unit available for cancer assessment – in contrast to ophthalmic OCT systems which have been widely introduced. The amount of studies applying OCT in precancer assessment, over the past 10 years, illustrates that the challenge of data evaluation cannot be addressed by improving image quality alone: In order to solve this, classification algorithms could be of benefit.

This study has established a cervical image collection to allow us to evaluate algorithms that are novel to OCT data classification: Principal Components Analysis (PCA) and Linear Discriminant Analysis (LDA). Both are statistical tools which have not yet been used for classifying A-scans according to their shape features (Chapter 5). There is potential for such an approach to distinguish OCT images in short time and with high accuracy, although this remains to be confirmed for a clinical database (Chapter 6). Here, we will investigate how this algorithm performs when trained to distinguish clinically relevant stages of cervical precancer.

8.3 Materials and methods

8.3.1 The study protocol

Human tissue samples were collected following the “favourable opinion”⁷ of the Gloucestershire Local Research Ethics Committee, Study Nr. 05/Q2005/123. Samples

⁷ The term “approval” was replaced by the term “favourable opinion” as this was deemed to be a more accurate expression; studies are therefore no longer “approved”. In practice, this should not make a difference.

were obtained from participants who needed a biopsy for diagnosis or treatment. These were acquired routinely for histopathologic examination, and therefore there was no need for additional biopsies to be taken for this study. The samples were categorised into three groups: hysterectomies, large loop excisions of the transformation zone (LLETZ) (both are treatments) and punch biopsies (diagnostic).

Samples from LLETZ, or “loop”, procedures are also referred to as “biopsies”. To clarify the difference, we refer to diagnostic samples, which resemble skin flakes of about 1-2 mm, as “punch biopsies”, to treatment samples, which are 1 or a few centimetres in size, as “LLETZ”, and we use the term “sample” rather than “biopsy” when addressing both groups. Figure 8.1 shows a specimen under the sampling arm.

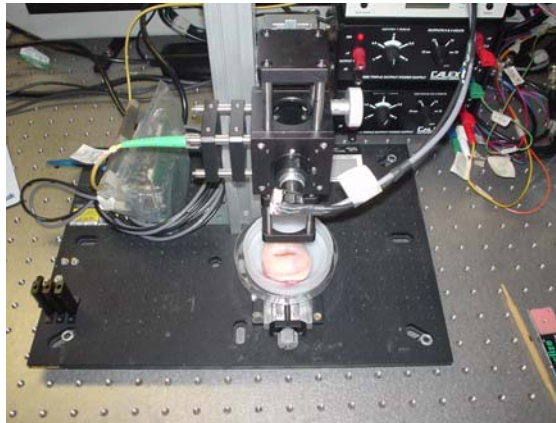


Figure 8.1: Photograph of an entire cervix sample, amputated from the uterus after a hysterectomy and placed in a specimen tray under the probe arm of the OCT system.

8.3.2 Measurements

For routine diagnosis, samples were put in formalin immediately after surgical removal, and processed by the histopathology department. However, in our study, samples were placed in physiological saline at room temperature before imaging was performed. The time between surgical removal and the measurement depended on the amount of participants per clinic and varied between 30 minutes and 3 hours. The measurements were performed as described in Chapter 7.3.1. After imaging, samples were placed in formalin and sent to histopathology. Thereby the routine diagnostic processing of the sample was not delayed.

The imaging sites were tagged using an orientation needle, thereby roughly marking the imaging positions. We had to assume that the pathology of the marked site and the imaged area are the same. It is noteworthy that due to the heterogeneity of cervical tissue and the micrometre thickness of both sections this approximate assumption may not hold true and the slice for histopathological assessment may be different from the OCT section, as demonstrated by Figure 8.2. From larger samples it was possible to obtain images from healthy areas (near the excision margins) and images of progressed CIN.

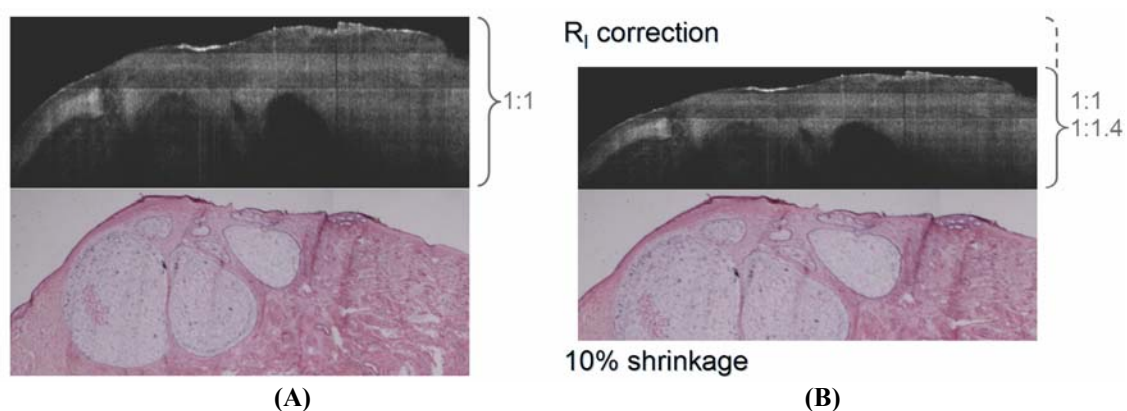


Figure 8.2: Difficulties with matching an OCT tomogram (Top) with the equivalent histopathology section (Bottom). These OCT images were created from data taken with OCT instrumentation that is capable of volumetric imaging. Of this volume, the best possible match and orientation from a series of histological slices was selected. (This comparison is credible through the particular spatial arrangement of the three cysts that is not present anywhere else in the sample.) N.B., this instrumentation was not available at the time of the study and matches from the study may be worse. (A) Top: Reconstruction from volumetric OCT dataset; Bottom: Equivalent site from histopathological sample. (B) Attempt of introducing correction factors, for refractive index mismatch and tissue shrinkage from fixation process.

8.3.3 Data evaluation

Using PC-fed LDA we were able to distinguish different types of biological samples and animal tissue with a correct classification rate of up to 93.3 % (Chapter 7.4.2).

Continuing from these findings, we evaluated the data here with different pre-processing approaches. We found that by far the biggest impact on classification is due to surface normalisation (Chapter 7.4.1). Further, the optimal amount of fed PCs can influence the performance (Chapter 7.5.1), and median filtering of the surface vector can be beneficial (Appendix C). We have used cross-validation as an estimator for their performance on new, unknown samples (Chapter 7.5.2). In this study we will

learn that certain biological specimens are not easy to measure (details below, Chapter 8.4.2), and that the best imaging conditions are assured for larger (LLETZ and hysterectomy) samples.

These findings on pre-processing we want to apply to this study. We have chosen 2 pre-processing methods and 2 specimen sources to be compared. The differences are shown in Table 8.1.

Table 8.1: Pre-processing and group composition for the 4 data evaluation arms

	Group 1	Group 2	Group 3	Group 4
Name	Entire set A	Entire set B	LLETZ A	LLETZ B
Symbol in graphs	◇	□	△	×
Specimen selection	Entire selection of samples, i.e. 1057 images from 187 patients		LLETZ and hysterectomy samples, i.e. 623 images from 86 patients	
Pre-processing:				
Surface recognition	Bimodal thresholding	Bimodal thresholding	Bimodal thresholding	Bimodal thresholding
Threshold level	0.95	0.90	0.95	0.90
Surface correction	none	15 pixel Median window	none	15 pixel Median window
Rows to be used	1-500 (full depth)	1-250 (half depth)	1-500 (full depth)	1-250 (half depth)

For each of the four models, the data was reduced to 35 PCs, which were sorted according to their significance (ANOVA) in respect to the disease staging provided by histopathology. Of these, between 10 and 35 PCs were then fed into LDA.

Ten different models were constructed in order to discriminate normal and the CIN1-3 groups. Using N, 1, 2, and 3 as group designators, these models were: N↔1, N↔2, N↔3, 1↔2, 1↔3, 2↔3, and clinically more significant decisions N1↔23, N↔123, N12↔3, and 1↔23. (Clinically worse groups, i.e. those positive for disease, are consistently assigned to the right side.) In total, there were 10 models, 26 PCs per model, and 4 pre-processing routes, resulting in 1040 models to be assessed. Of these 1040 models, we calculated the training performance and the cross-validated prediction performance. Multiple measurements of the same patient (which we did obtain) in a

training and prediction set would reduce the reliability of a cross-validation. Therefore, a leave-one-patient-out (LOPO) cross-validation was performed in this situation⁸.

A further estimator for the validity of results is the kappa (κ) value. κ statistics provide an accepted estimator to establish whether an agreement could have been achieved by chance, on a scale of -1 to 1. A $\kappa = 0$ means that the same performance would have been obtained by chance [213].

For a selected model, we will calculate sensitivity and specificity. There is a caveat as these two measures would not make sense clinically for every classification model, e.g. for distinguishing CIN2 from CIN3. For the N1 \leftrightarrow 23 model, the clinical relevance would be in predicting progressed disease. A “positive” would therefore be CIN2/3. A “negative” would be the lesion that does not require treatment or is healthy, i.e. CIN1. Table 8.2 shows how sensitivity and specificity are related to the group assignments in this study.

Table 8.2: Basis for sensitivity, specificity, and κ value calculation as used in the literature (left table), and as applied from group assignments in this study (right table).

Novel approach	Reference		Prediction by PC-fed LDA	Histopathology grading	
	Positive	Negative		Higher	Lower
Positive	True Pos.	False Pos.	Worse stage	True Pos.	False Pos.
Negative	False Neg.	True Neg.	Milder stage	False Neg.	True Neg.

8.4 Results

8.4.1 Recruitment

Over the course of 56 clinic sessions, 223 women consented to be included in the study. Not all of the recruited patients needed a biopsy and therefore the number of study

⁸ The LOPO CV was still undertaken with the full amount of images. E.g., for a full data set of 187 patients, 187 LOPO cross-validations were performed. This data set would contain 1057 images. If one patient provided 6 images, then the training set in this CV step was built from 1051 images and the prediction of the remaining, “unknown”, 6 images was determined. After 187 CVs, each of the 1057 images has a - either correct or wrong - prediction, and the sum of the correct predictions divided by the total 1057 images is the classification performance.

participants was lower. Samples were collected from 199 patients; in some patients multiple biopsies were taken. 245 tissue specimens were obtained, 149 punch biopsies, 92 LLETZ, and 4 specimens from hysterectomies.

The study was designed as a pilot study to recruit 200 patients and at least 5 tissue samples of each of 6 groups (ectocervical normal, CIN1, CIN2, CIN3, and endocervical normal and cancerous (adenocarcinoma)). Because the cancer progression markers within OCT images are not clearly defined throughout the literature, a power analysis was not appropriate at the time and the study was classified as a pilot study. From experience within the Women's Health Directorate, the Histopathology Department, and the Biophotonics Research group, it was estimated ("educated guess") that in order to recruit about 5 healthy tissue samples out of a population participating in a follow-up to positive screening results, larger numbers would be required. This estimate recognised that the number of CIN samples would be larger and it was also suggested to include tissue from hysterectomies that was not obtained as part of the screening programme. Because of these reasons and because the study did not alter the way patients receive care, this procedure was accepted by the ethics committee.

The breakdown of participants according to histopathology was: 12 normal, 49 CIN1, 53 CIN2, 63 CIN3, and 12 exclusions (a detailed breakdown with numbers of patients, samples, and images per group is given in Table 8.3). The mean age (\pm standard deviation) of participants was 31.8 ± 8.3 . We recorded, where possible, their ethnic background, parity, and hormonal status, in order to be able to investigate potential trends.

8.4.2 Images

1387 images were obtained from 245 samples; the breakdown of pathology identified was as followed: 20 healthy ("no sign of neoplasia"), 87 CIN1, 65 CIN2, and 73 CIN3. Figure 8.3 shows example images obtained of the individual pathologies as listed above. The cervical images are showing features which are consistent with cervical histology: Glands, blood vessels, cysts, mucus and epithelial layering can be made out. However, these features are not correlated with CIN progression, and can therefore not serve as markers. Not all of the images were suitable for further identification: Orientation of tiny punch biopsies was difficult, as they are prone to twisting and curling. Some

samples, although with sufficient penetration depth, did not cover the full lateral extent of the scan (Figure 8.4, Figure 8.5). 330 such images were not included in the classification set.

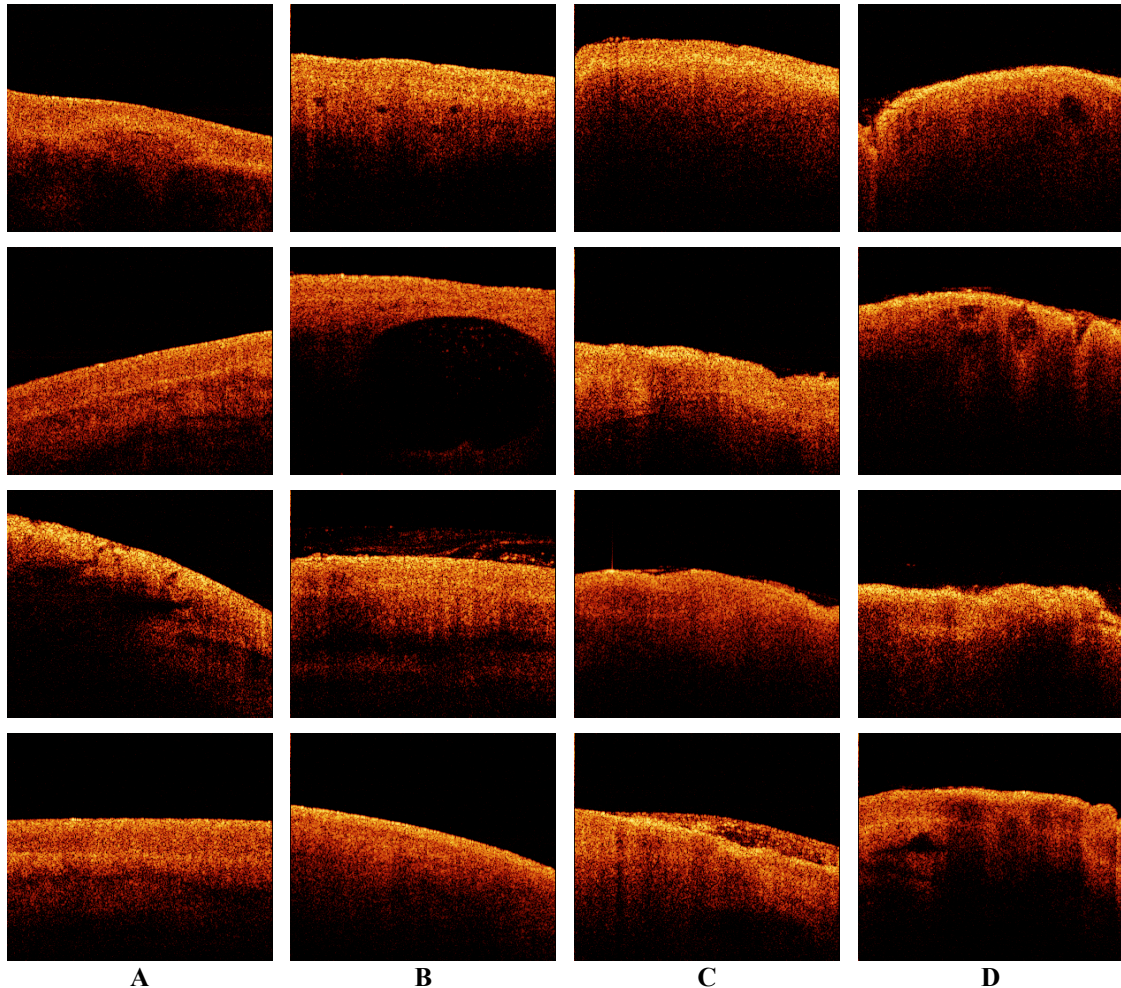


Figure 8.3: OCT images of good quality of the different CIN groups defined by histopathology. These images are representative for those images of good quality and full penetration depth. Each image has a width of 3 mm and a resolution of $\sim 10 \mu\text{m}$. Columns show: A) No sign of dysplasia, B) CIN1, C) CIN2, D) CIN3. From such images it is not straight forward to establish or recognise features which are relevant to disease progress.

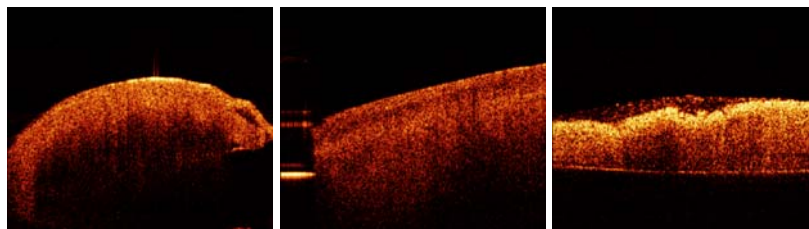


Figure 8.4: Examples for images whose useful areas had to be manually selected, or were entirely excluded from the classification due to lack of penetration depth.

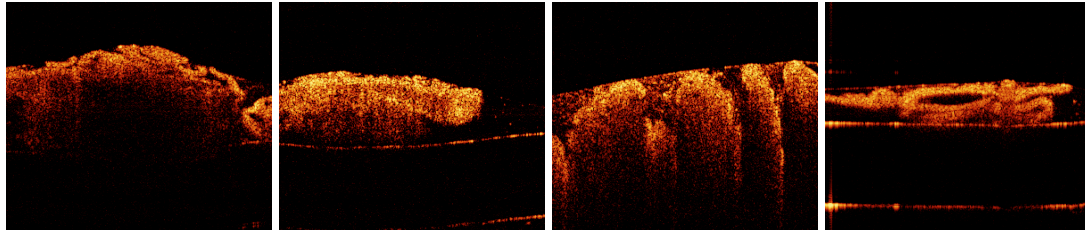


Figure 8.5: Examples for images which had to be excluded because of the difficulties with surface recognition due to orientation, curling, or wart structures.

Of the entire 1387 images from 245 samples from 199 patients, 1057 images from 212 samples from 187 patients were used for assessing the performance of training sets and cross-validation. Because there were different numbers of patients and measurements within each pathological group, we provide a table showing how many images, of how many patients, are used for the classification (Table 8.3).

Table 8.3: Composition of the 4 groups which are used to form the 10 models, and total numbers of selected/excluded samples and patients. The dashed box encircles the samples used for the “Entire” sets (in the following graphs designated by the symbols \diamond , \square), and the grey shaded box marks the samples used for the “LLETZ” sets (symbols \triangle , \times).

	selected into groups				total selected	excluded	total imaged
	“N”	“1”	“2”	“3”			
Staging	No sign of dysplasia	CIN1	CIN2	CIN3			
Images of:							
Punch biopsies	67	171	103	93	434	247	681
LLETZ	34	39	187	291	551	74	625
Hysterectomies	72	0	0	0	72	9	81
Total images	173	210	290	384	1057	330	1387
Participants providing:							
Punch biopsies	15	43	22	21	101	8	109
LLETZ	2	6	31	43	82	4	86
Hysterectomies	4	0	0	0	4	0	4
Total participants	21	49	53	64	187	12	199

8.4.3 Data evaluation

The results here show training performance, cross-validation, and κ values for 1040 ($26 \times 4 \times 10$) models. The graphs are designed to show only the optimal result for each model and may seem quite busy. In order to display the impact of the number of PCs,

comet tails plots are introduced. A guide on how to interpret the comet tail plots is given in Figure 8.6.

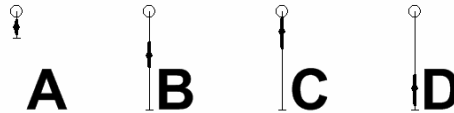


Figure 8.6: How to read the dependency of TPs on the number of PCs in the following figures: The comet tail plots are derived from whisker plots, stretching from maximum to minimum. A bold line marks the interquartile range, and the median value is marked separately. (A) shows little variation and hence little influence from the number of PCs. (B) shows a larger variation. (C) Here, the majority of PC numbers would yield good results. (D) shows a critical situation where the maximum TP can easily be confounded by a sub-ideal number of PCs.

Figure 8.7 is discussed here as guide for reading and interpreting the result figures: Values using the optimum number of PCs range between 58.0 % and 83.3 %, without cross-validation. The comet tail plots show the difference between worst and optimal PC choice: The longest tail, as it were, makes up for a difference of 10.3 percentage points.

We see that the refined pre-processing (symbols \square , \times) does not have a strong impact, in fact it performs worse in most cases. The scores for the LLETZ samples are higher than for the entire set. However, within the LLETZ samples there is a smaller proportion of healthy/CIN1 pathology.

Table 8.4 gives a detailed account of how big the sample numbers were, according to the data provided in Table 8.3. The strongest performance is achieved where CIN1 has to be distinguished from CIN3, and in models where those two pathology groups are separate. Again, these models have smaller sample numbers. The weakest separation is achieved for CIN2 vs. CIN3, and this affects models where these pathologies are not in the same group.

These models were not yet cross-validated and rise concern insofar as even the best result has a 16.7 % error in a task that effectively consists of matching a descriptor to a database that already includes this descriptor.

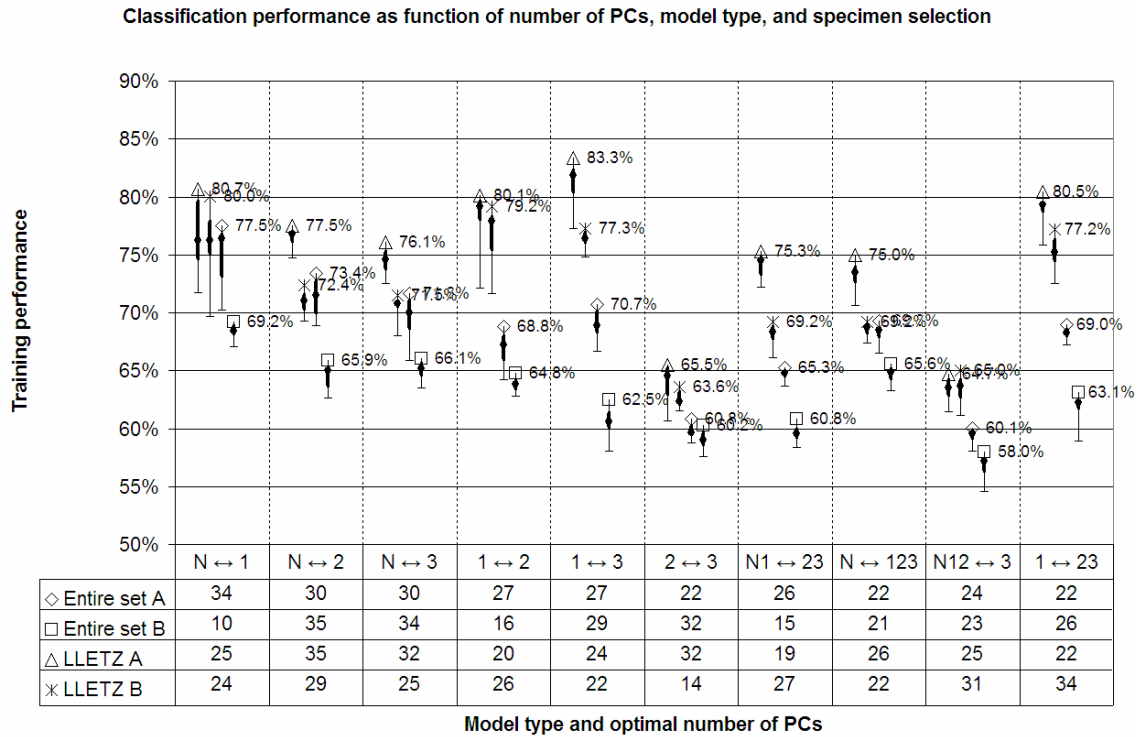


Figure 8.7: TPs of different classification models. Group designators: N...no sign of dysplasia, 1...CIN1, 2...CIN2, 3...CIN3. The percentages are the highest values achieved with the optimal number of PCs fed into LDA. These optimal numbers of PCs are shown in the data table below. E.g., N↔1 designates a model for separating healthy from CIN1. For this model, 34 PCs gave the highest score of 77.5 % to separate data from all specimen types (◇). The same model built from LLETZ and hysterectomy samples (△) achieved 80.7 % with 25 PCs. The comet tails of each data point show the range of TP achieved when feeding between 10 and 35 PCs.

Table 8.4: Number of images and group sizes of the 10 TP models.

	N↔1	N↔2	N↔3	1↔2	1↔3	2↔3	N1↔23	N↔123	N12↔3	1↔23
“Entire” set										
Left side	173	173	173	210	210	290	383	173	673	210
Right side	210	290	384	290	384	384	674	884	384	674
Total	383	463	557	500	594	674	1057	1057	1057	884
LLETZ set										
Left side	106	106	106	39	39	187	145	106	332	39
Right side	39	187	291	187	291	291	478	517	291	478
Total	145	293	397	226	330	478	623	623	623	517

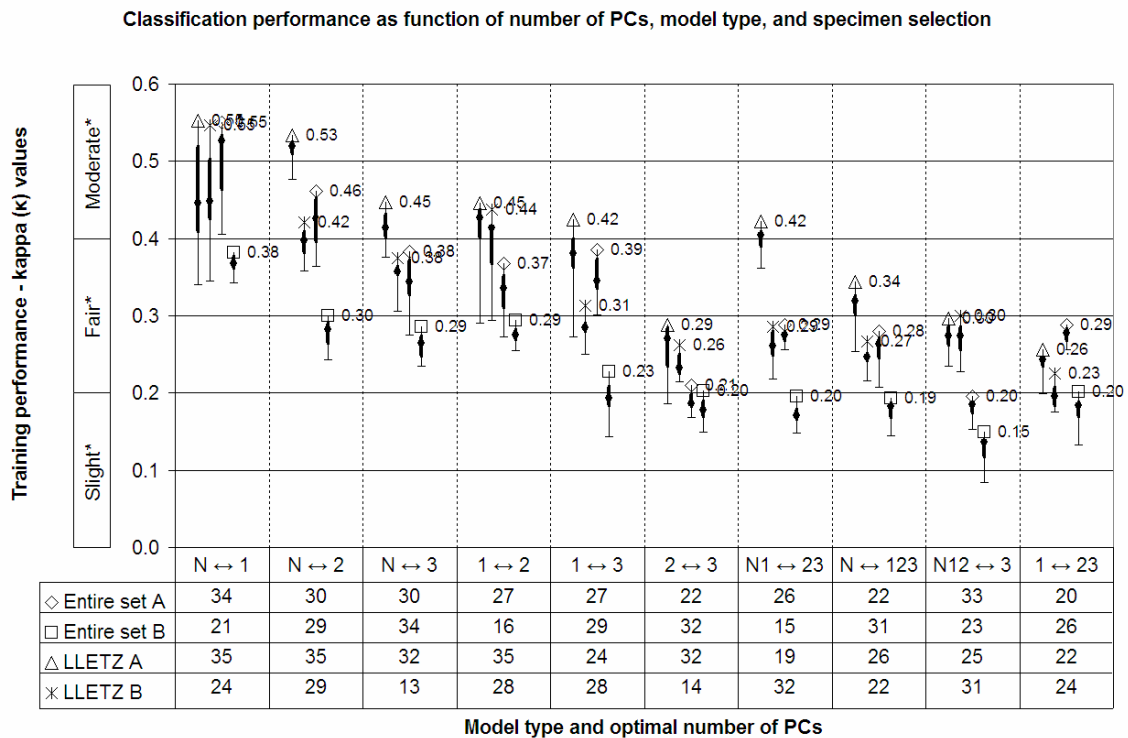


Figure 8.8: κ values for the models shown in Figure 8.7. Shown are the 40 values with highest performance in each model, and the range of 1040 models as function of the number of fed PCs, as explained in Figure 8.6. The data table shows the optimal number of PCs for these κ values, and they are mostly, but not entirely, identical to the TPs. (* Agreement interpretation after [214]).

In Figure 8.8, the κ values for the TP in Figure 8.7 show low values. Particularly for the models 1↔3 and 1↔23, which have high values according to Figure 8.7, we learn that much of this would have been obtained by guessing. However, as the κ values range from 0 to 1, we can claim that each model still achieves a slightly better performance than a mere guess. The model of direct clinical interest, N1↔23, has a TP of 75.3 %, and a κ of 0.42, when built from LLETZ samples.

We also see that the performance decreases with complexity of the models. When a group represents one clinical grade, κ is usually > 0.30 , groups including more clinical grades mostly have $\kappa < 0.30$. In these models, some scans were taken from the same sample or patient, respectively.

The following graphs show the results for LOPO CV, with the patient numbers, which equal the CV steps, detailed in Table 8.5. The prediction performance seems to be consistently lower than the TP, by about 10 % (Figure 8.9). In fact, the average of the entire 1040 performances drops from 68.7 % (TP) to 59.6 % (CV), and also the 40

optimal PC models fall from 70.1 % (TP) to 62.3 %. This was expected for the CV, as the classification for each patient was done on a model that did not know of this patient before. Although yielding lower figures, the CV performance is a more credible indicator for the diagnostic prediction performance.

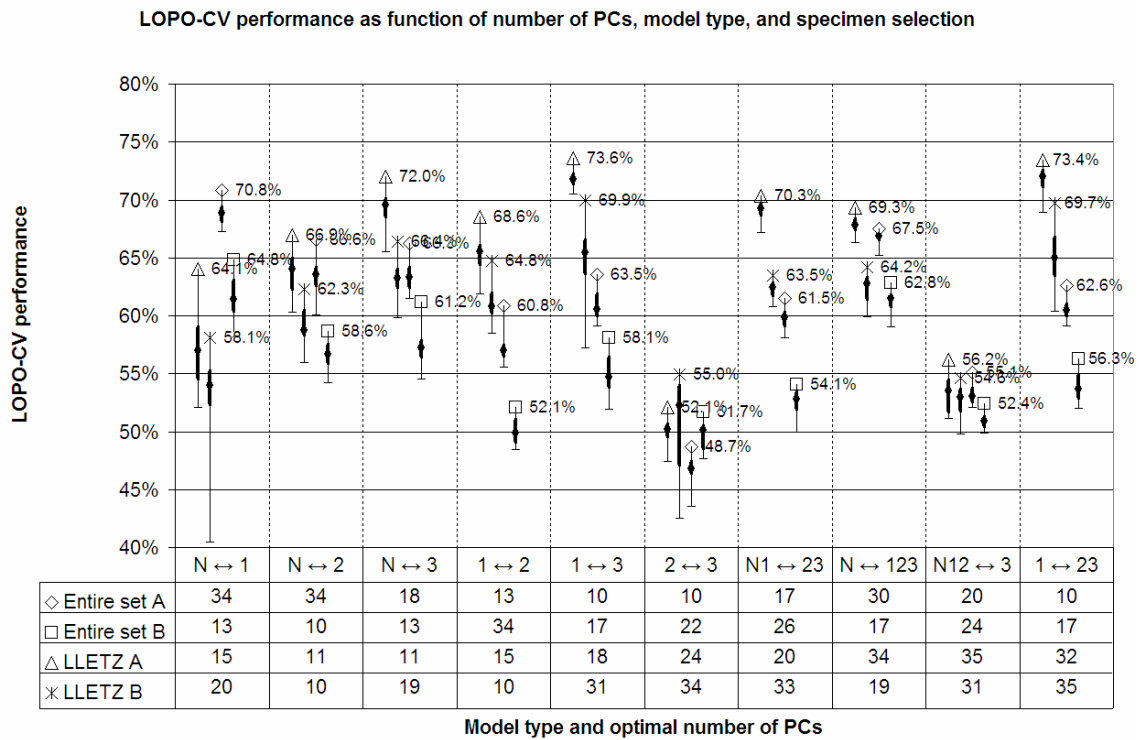


Figure 8.9: LOPO CV of different classification models. Interpretation and legend are the same as in Figure 8.7.

Table 8.5: Number of patients, and therefore cross-validations, of the 10 LOPO CV models. The amount of images used for cross-validation is shown in Table 8.4.

	N↔1	N↔2	N↔3	1↔2	1↔3	2↔3	N1↔23	N↔123	N12↔3	1↔23
“Entire” set										
Left side	21	21	21	49	49	53	70	21	123	49
Right side	49	53	64	53	64	64	117	166	64	117
Total	70	74	85	102	113	117	187	187	187	166
LLETZ set										
Left side	6	6	6	6	6	31	12	6	43	6
Right side	6	31	43	31	43	43	74	80	43	74
Total	12	37	49	37	49	74	86	86	86	80

Interestingly, κ values are higher for the prediction model (Figure 8.10). The amount of PCs does not seem to have an impact on most models in Figure 8.10, as illustrated by the short comet tails. However, there are high κ values for models separating CIN1 from CIN2, CIN3 or both. This suggests that better models are built in absence of normal tissue; which is a constraint that cannot easily be fulfilled *in vivo*. What is conspicuous in Figure 8.10 is that a majority of results seems to be situated near the 0.50 level, and that, although the CV of the data has similar results as the TP, the same cannot be said for the κ calculations. The CV κ values have higher values than the TP κ values. However, every calculation was done with the same program and the results shall therefore be accepted as correct.

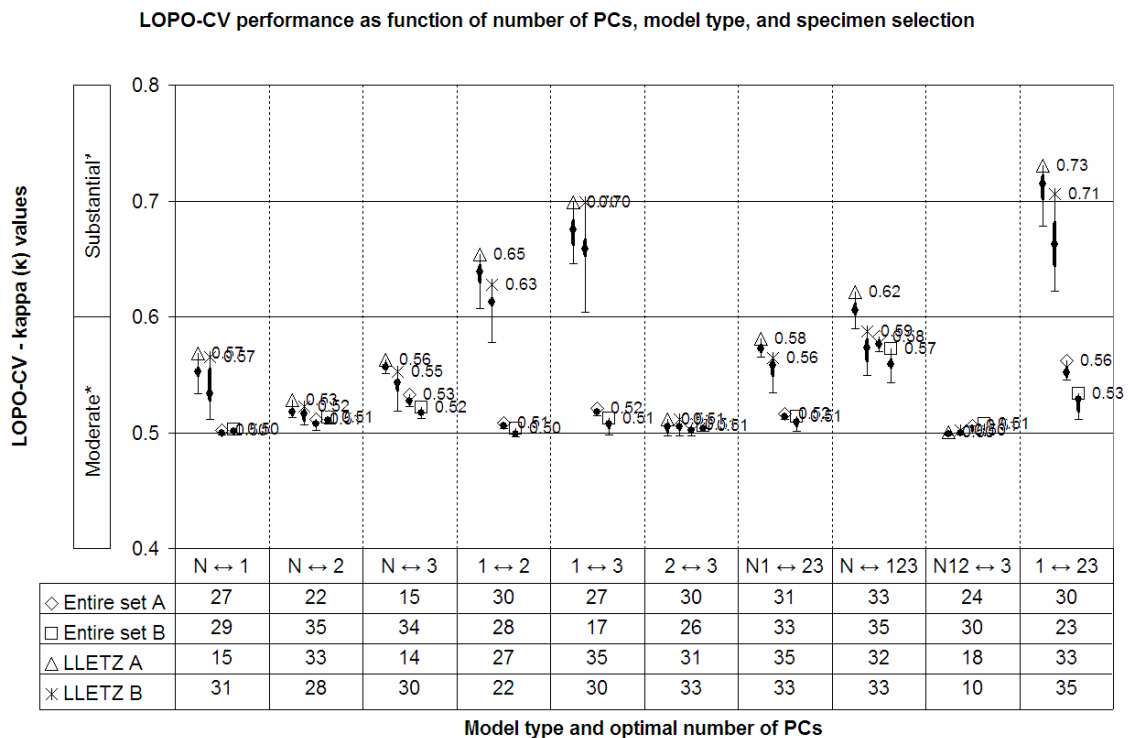


Figure 8.10: κ values for the models shown in Figure 8.9. Shown are the 40 values with highest performance in each model, and the range of 1040 models as function of the number of fed PCs, as explained in Figure 8.6. The data table shows the optimal number of PCs for these κ values. (* Agreement interpretation after [214]).

8.5 Discussion

8.5.1 Recruitment

Because this study did not interfere with the management a patient receives normally, patients were usually willing to participate. This was especially the case for patients with treatment. Although that option was emphasised to the participants prior to their consent, not a single one withdrew. Patients who attended a clinic for a diagnostic biopsy were in some cases more reluctant. We assume that although the provided information sheet clearly emphasised that no extra biopsies would be taken, this point might not have always come across. At the time of recruitment, we did not have a possibility to prove this to prospective participants. Now we can show that only 89.2 % of the recruits (199 of 223) became participants. Nevertheless our study size seems comparable with earlier studies in this field [99],[100].

It might be worthwhile to remind ourselves of Table 8.3, which showed that although a study was designed to recruit 200+ patients, further exclusion criteria mean that models consist of as little as 12 participants. This may serve as support that even small prospective studies, specifically when trying to cover several pathology groups, may require rather large participant numbers.

Table 8.3 shall be utilised to highlight further points:

- Although there were 1387 images from 199 participants, the smallest model (N↔1 for LLETZ & hysterectomy samples) consisted of only 145 (34 + 72 ↔ 39) images from 12 (2 + 4 ↔ 6) patients.
- The proportion of measurements from biopsies that was excluded is larger than the images from treatments (247 vs. 74 + 9).
- The proportion of progressed CIN is larger amongst LLETZ (34 + 39 vs. 187 + 291).
- Hysterectomies do not have CIN as they were performed for other reasons than cancer prevention.
- The proportion of early CIN is larger amongst punch biopsies.
- 67 punch biopsies from 15 patients presented no sign of dysplasia.
- 8 patients did have treatment for CIN although the subsequent examination showed that they had either no sign of dysplasia or CIN1.

These points support that in the current system diagnostic biopsies do act as a filter to provide treatment to only progressed CIN stages. There are, however, potentially redundant diagnostic biopsies, and treatments of patients which would have required surveillance only.

8.5.2 Sample quality

When measuring the samples, we had problems as the penetration depth of our system exceeded the size of the punch biopsies, as shown in Figure 8.4 and Figure 8.5. This does affect the data evaluation as it is based on the shape of a full A-scan. After all, OCT is a technology that may give insight on histological features, and does not render itself to cytological assessment. For small samples, manual handling was difficult and made proper imaging impossible. The images which are representative for an *in vivo* situation are therefore mainly from LLETZ procedures and hysterectomies. As a consequence, we had to exclude some of the images from our automated analysis based on the criterion whether an artefact would have occurred *in vivo*.

8.5.3 Data evaluation

The results from the non-clinical data set demonstrated that the automated classification can achieve good separation results (Chapter 6). However, this relies on the appropriate pre-processing. Pre-processing aims to emphasise features which can then be easily extracted and fed into the classification. Such features can be layer thickness, grade of layer disruption, or intensity patterns of A-scans.

These features need to correlate with the progression of precancerous lesions. The advantage of PC-fed LDA is that such features need not necessarily be known a priori. Establishing such features in precancerous OCT images of the cervix so far remains challenging. There are many structures such as glands, cysts, or characteristic wart surfaces, which do not correlate with CIN grades. It would be interesting to understand which “structural” features the algorithm has used for the classification of groups, even if this did not work perfectly.

This can usually be achieved by looking at the loadings of the strongest PCs, however this is a technique used in spectroscopic analysis of wavelength dependent photon interaction and does help elucidating an arbitrary A-scan shape. As PC-fed LDA was

applied to OCT data in this study, we try to illustrate representative structural features through what is presented in Figure 8.11:

Figure 8.11 starts with plotting the intersections of the 2 strongest⁹ PCs of representative measures. From this plot, groups of points are selected at north, south, east, west, and centre regions. They are not perfectly aligned as here we want to calculate the mean spectrum of these groups and therefore require areas with several points.

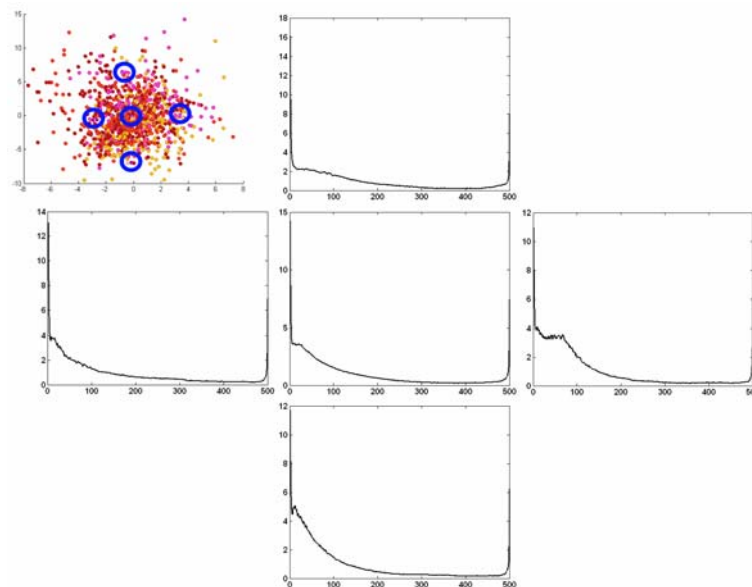


Figure 8.11: Upper left: scatter plot of the 2 strongest PCs of the entire data set, consisting of four groups (no sign of dysplasia, CIN1, CIN2, CIN3). Each B-scan descriptor is represented by a dot. By selecting mean descriptors of a group we can visualise the axes in an intuitive way. The blue circles roughly correspond to the PC dimensions (north, south, east, west, and centre). The five profiles are mean spectra from within the blue circles and can serve as an indicator that the two strongest PCs describe steepness and the presence and appearance of a plateau region. These profile features do not seem to relate with CIN stages.

The strongest feature for separation may be described as the steepness of the signal, or penetration depth. The second strongest feature seems to relate to the presence and appearance of a plateau within the scan. The plateau might be related to the epithelial thickness. However, the overlapping data clouds show that these features do not strongly correlate with precancer progression, and occur across all the CIN stages.

⁹ Strongest correlation with CIN progression, established by ANOVA

8.5.4 Results in context of literature

Table 8.6 offers the possibility of comparing our κ values with those available in the literature. There is no full meta-analysis available and performing one would have gone beyond the time frame of the project. It is however possible to identify some of the relevant work. These publications are slightly different as they do not compare standard histopathology with OCT, however they either assess inter-operator agreement between expert histopathologists, or between conventional histopathology and a novel diagnostic tool. Table 8.7 gives more information on the details of the reference studies.

Table 8.6: Comparison of highest κ values for “Entire” set and LLETZ set with literature.

Source	N↔1	N↔2	N↔3	1↔2	1↔3	2↔3	N1↔23	N↔123	N12↔3	1↔23
Entire set (◇)	◇: 0.50	◇: 0.51	◇: 0.53	◇: 0.51	◇: 0.52	◇: 0.51	◇: 0.52	◇: 0.58	◇: 0.51	◇: 0.56
LLETZ (△)	△: 0.57	△: 0.53	△: 0.56	△: 0.65	△: 0.70	△: 0.51	△: 0.58	△: 0.62	△: 0.50	△: 0.73
Source:										
Cai [215]								0.77-0.87		
Horn [216]							◇: 0.49 △: 0.63			
Kato [217]									0.58	
McCluggage [218]								0.30		
Parker [219]							0.81	0.55		
Stoler [220]							◇: 0.68 △: 0.69	◇: 0.47 △: 0.46		

Table 8.7: Details to references in Table 8.6.

First author, Year	Panel size	Sample number	Group name (equivalents in source)	Remarks (which values are used here)
Cai 2007, [215]	4	185	N1↔23: “non-CIN2/3+ vs. CIN2/3+” N↔123: “non-CIN+ vs. CIN+”	Weighted κ values
Horn 2008, [216]	6	247 punch, 249 cone biopsies	0.49 for biopsies 0.63 for treatments	Compare conventional vs. novel staining. Conventional κ values used here
Kato 1995, [217]	3	883	N12↔3: “CIN3 in histology”	Values for histology, CIN3
McCluggage 1996, [218]	6	125	6 tier staging	Unweighted κ values
Parker 2002, [219]	3	119	N1↔23: “HG vs. <HG” N↔123: “normal vs. dysplastic”	κ value used here is mean of 3 operator-pair wise comparisons
Stoler 2001, [220]	7	2237 biopsies, 535 LLETZ	N1↔23: “≤LSIL vs. ≥HSIL”, N↔123: “≤NEGATIVE vs. ≥ASCUS”	

8.5.5 Potential for improvements

In this work we have undertaken a classification of several models derived from clinical stages of precancerous disease. For the entire set, our LOPO CV classification result reached 61.5 % and a κ of 0.52 for distinguishing healthy/CIN1 from CIN2/3. This was however dependent on the number of PCs fed into LDA. Table 8.8 shows how sensitivity, specificity, classification and κ value change with the model. E.g., for the model that reaches 61.5 % correct classification, the κ value is 0.51. It does have the highest specificity (47.3 %), but the sensitivity is low (43.8 %). The highest sensitivity of 66.7 % would be obtained by feeding 34 PCs, where the classification performance reaches 58.1 %.

Table 8.8: Maximum performances and dependencies on PCs.

Amount of PCs for highest value	Classification	Sensitivity	Specificity	κ value
17	61.5 %	43.8 %	47.3 %	0.51
31	59.2 %	66.6 %	44.9 %	0.52
34	58.1 %	66.7 %	43.8 %	0.52

Why is the performance not higher and what could be investigated to improve these results? With *a posteriori* knowledge, different models could be designed from different sample sources and from different clinical stages. Such an approach may have a potential to reach higher classification performances, however that is irrelevant for a clinical decision making tool. When comparing our findings to the literature, there are clearly higher κ values. However, there are studies which have comparable or even lower values than provided by our approach. This seems a weak support for altering the current management. However, the results are promising enough to try and identify possible improvements.

SNR / Speed

For a fully automated algorithm, cancer progression markers need to be picked up properly and normalised in order to not disturb the analysis. Notably, the ability to extract features depends on the original quality of the image. Today, OCT systems provide higher resolution and higher acquisition speed. Both might be crucial for automated registration of image features and artefacts.

Correlation with gold standard

We also have to question the staging of our OCT images. Per study design, a truly accurate staging could only be expected from the tiny punch biopsies. As it turned out, these were the least useful samples for the subsequent evaluation due to handling problems. Although care was taken in measuring the correct areas of the larger specimens, it cannot be fully confirmed that these were 100 % corresponding to the slices examined by the histopathologists. A reasonable way to overcome this would be a different protocol that involves volumetric OCT, and therefore provides larger areas for matching biopsies and OCT scan. Within this PhD project, such a system was not available.

Data evaluation

PC-fed LDA is able to distinguish different data, and in hindsight the separation of non-clinical data (Chapter 6) was a worthwhile undertaking as it can serve as a proof of principle.

Normal epithelial tissue provides a wide range of structural features. The histological features of early precancerous changes are minute – after all they are symptomless. OCT is much more likely to detect a change in the reflection gradient within the epithelium rather than being able to resolve single cells.

A screening program is per definition aimed at the healthy population. It would therefore be possible to assess the intra-patient variation of measurements, rather than the inter-patient variation as it was done in this study. OCT would be technically advanced enough to image an entire cervix in real-time. Such an approach would avoid problems occurring from feeding different stages, age groups, and eliminate patient heterogeneity as a source of problem. In this study, we are not able to assure that.

8.5.6 Communication to clinician

Classification is a key instrument for providing a clinician with clear answers, rather than images. From our findings we believe that inter-patient variation in epithelial appearance is rather large and dominant over less pronounced markers for early malignancy (Figure 8.11). However, imaging of the entire cervix *in vivo* should avoid these problems: Since OCT would be applied in early stages of precancerous development, the majority of the cervix surface is a healthy area which can be used as

an intra-patient reference. Classification models would therefore not have to be as robust to inter-patient variation. This study was not designed for this evaluation. However, in a few cases images from the same patient are available (Figure 8.12). Such a model is too small for conclusions to be drawn, but it shows that assessing the intra-patient variation might be a possible way forward.

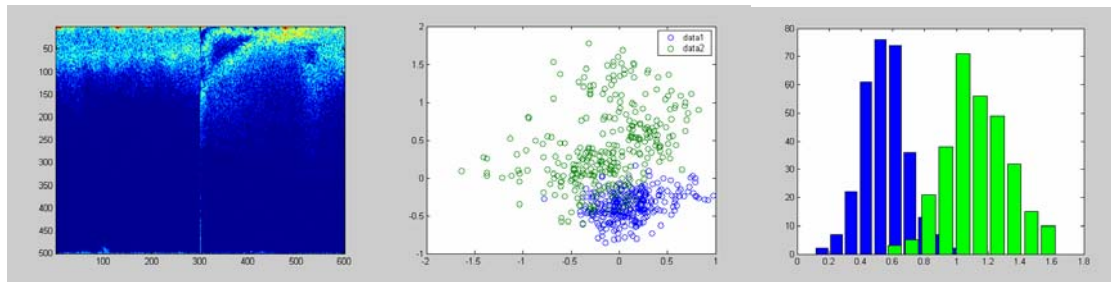


Figure 8.12: Two dissimilar, already surface normalised scans from the same patient, allowing discrimination. Here, no representative measure has been taken, and each dot represents an actual A-scan. Using a representative measure, these graphs would show 2 dots.

8.5.7 Acknowledgements

This work is funded by the Pump Prime Fund of Cranfield University and Gloucestershire Hospitals NHS Foundation Trust.

Dr. Nicholas Stone holds a Career Scientist Fellowship which is funded by the UK National Institute of Health Research.

The work presented in this article would not have been possible without the initial support of Dr. Shing “Sammy” Cheung.

9 The meaning of it all: Discussion

This PhD project combined several current issues in medical diagnostics: Epithelial pre-cancer detection, non-destructive diagnostics, and clinical decision making.

Prior to discussing the specific findings of each chapter, the work shall be put into context:

9.1 A framework for understanding this thesis

9.1.1 General approach

In the general understanding, a doctoral thesis should extend the current knowledge by a substantial finding.

- This could be the correlation of data points after a series of observational experiments (Figure 9.1). This correlation can then serve as rule, e.g. in form of an equation which is valid within a certain set of conditions.

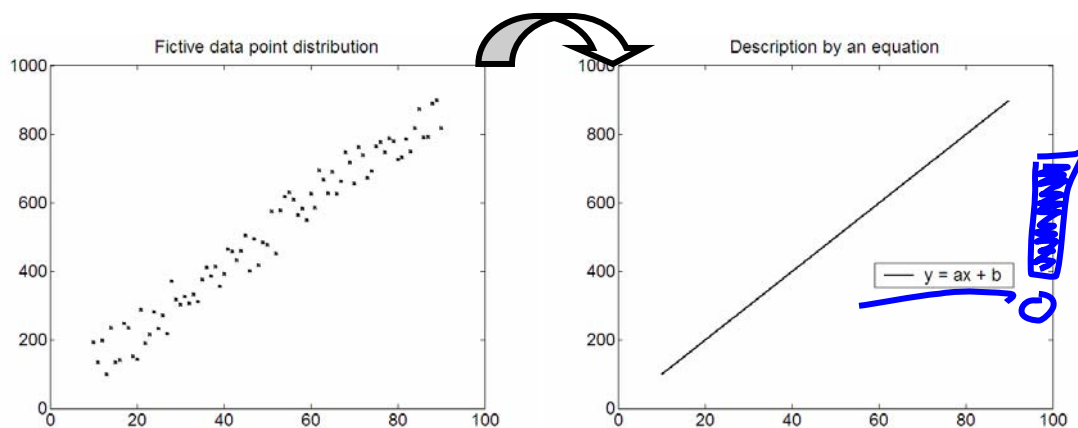


Figure 9.1: Example for observational finding: A series of experimental data points serves as basis to create a rule.

- Another finding a thesis could provide is a method to interpret existing data points. This would help obtaining or understanding future findings. Figure 9.2 shall illustrate the example of using a logarithmic representation, a frequently encountered approach for yielding a linear dependency of data points.

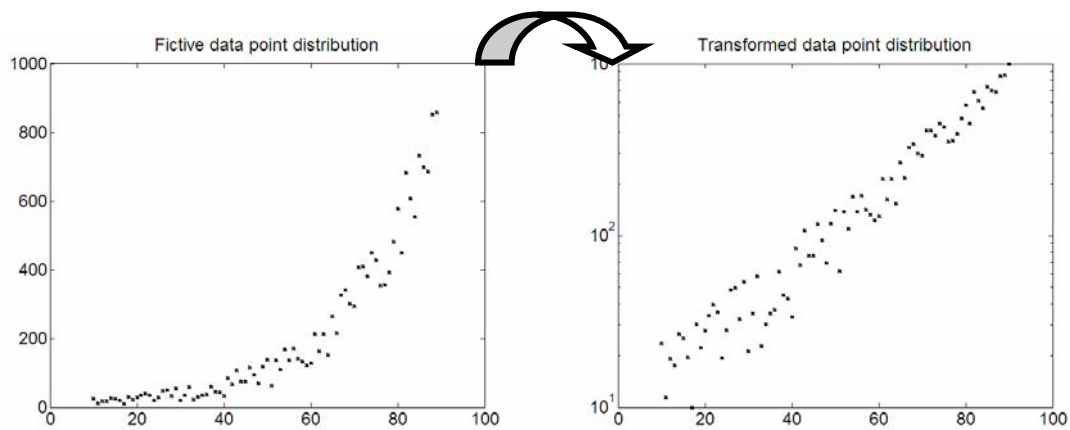


Figure 9.2: Example for methodological finding: Introducing a tool (here the logarithm) with the aim to facilitate interpretation.

9.1.2 Specificities in this thesis

On the basis of Figure 9.1 and Figure 9.2, the work in this thesis shall be appraised:

- Prior to any experimentation, two topics were reviewed: the existing literature on optical non-destructive imaging of cervical precancer, and on automating evaluation of OCT imaging data. This provided a justification for performing this research and showed the potential of OCT in precancer assessment and for a novel algorithm.
- In the experimental part, a collection of OCT images of cervical tissue samples has been established (Figure 9.3). This in itself is not entirely novel, as there are a few preceding publications [96]-[98]. It is however the first study of this size (200+ participants) undertaken in the UK. Results of this kind might help substantialise the (still few) findings in the literature.



Figure 9.3: Data collection in this thesis.

- An algorithm has been developed to help converting OCT image data into a measure that can be easily assessed, following the interpretation of Figure 9.2. It has been explored on non-clinical data to provide proof of principle. A novel approach introduced in this thesis is to reduce A-scans by PCA and to classify this data using LDA (Figure 9.4).

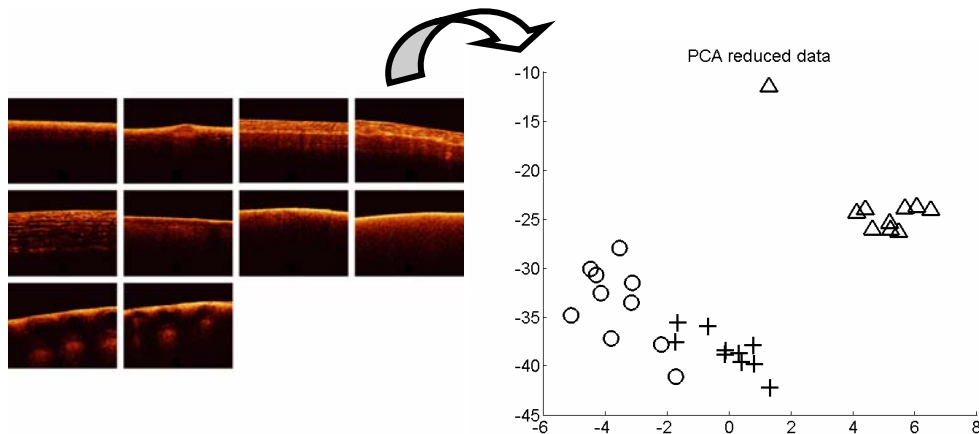


Figure 9.4: The process of converting OCT data by reduction to a hyper-dimensional data point, as developed in this thesis.

- The following difference has to be appreciated: The fictitious examples of Figure 9.1 and Figure 9.2 were 1 D, experimental data that is correlated with one variable. The PC-fed LDA accepts multi-dimensional data: Data points which are described by several dimensions form data clouds, which serve for creating hyper-dimensional maps. These maps are used to calculate a distance of unknown points, which are inversely proportional to the classification likelihood (Figure 9.5). This likelihood score is then a 1 D (linear) correlation.

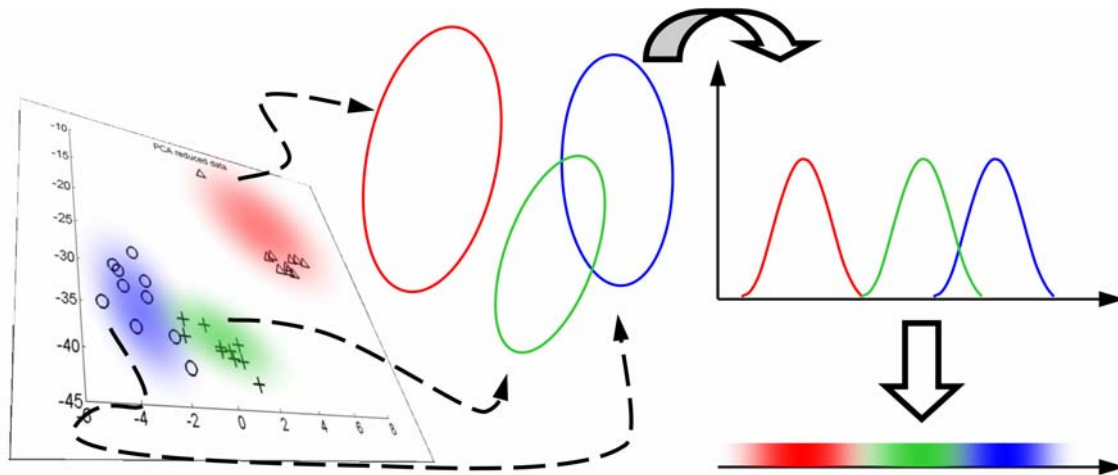


Figure 9.5: Establishing a linear correlation from multi-dimensional data (here illustrated by 2 dimensions).

It is now worth investigating whether there is a possibility to use OCT in cervical cancer management, e.g. by classifying the cancer risk as depicted in Figure 9.6?

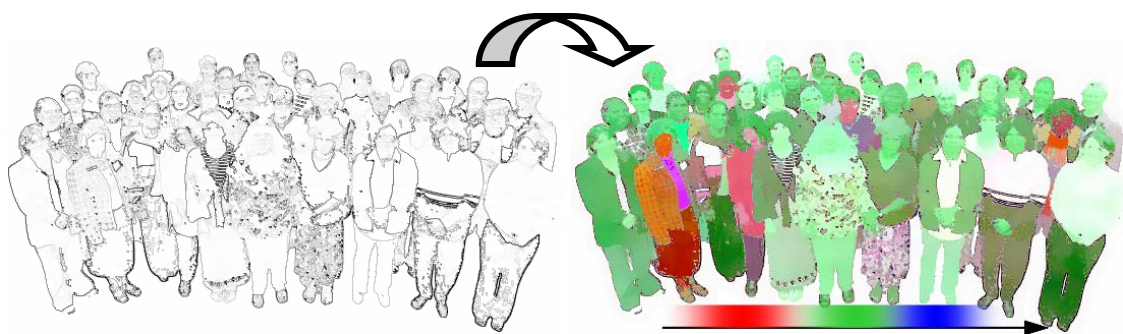


Figure 9.6: The aim of this thesis: using OCT to predict the risk of precancer progression.

- The algorithm, PC-fed LDA, has been applied in an attempt to classify clinical data of cervical tissue samples. For this situation, the algorithm achieved a 61.5 % correlation and does not fare better than reference values achieved by standard histopathology (Figure 9.7). For this model, sensitivity and specificity do not exceed 66.7 % and 47.3 %.
- In this study, OCT has been used to image *ex vivo* samples, where it would be positioned to compete with histopathology. If such a score can be achieved *in vivo*, this would provide an advantage. An *in vivo* tool of this kind would more likely be a screening tool to guide biopsies.

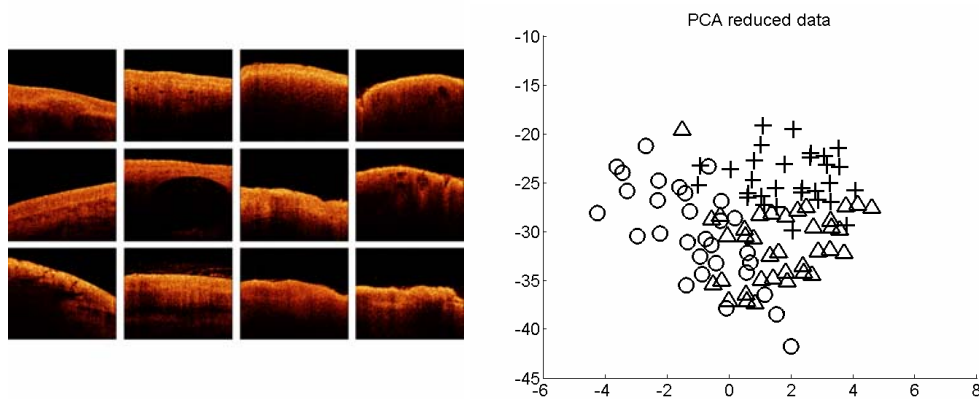


Figure 9.7: A correlation for the cervical OCT image collection has been performed. This was made difficult by overlapping data clouds and subsequent weak separation.

As answer to what has been achieved during the work of this thesis, the following list shall help as guideline:

- A review on optical high-resolution mapping of the cervix
- A review of analysis methods for OCT data
- A generic approach for converting dense tissue OCT data into an objective measure
- A specific data collection that has been described by the novel method

9.2 The findings in detail

9.2.1 Modalities for assessing cervical precancer

There are many diagnostic tools under investigation for the cervix, and an exhaustive list would probably have to include every measurement principle possible. In Chapter 4, a very narrow field was reviewed: non-invasive optical high-resolution modalities. OCT differs from confocal microscopy, infrared spectroscopy and Raman spectroscopy: The newest OCT systems are quick enough to provide wide-field, high-resolution (lateral and depth) imaging. The low powers employed mean that samples do not suffer photo-bleaching. On the other hand, the contrast is rather weak.

Chapter 4.3.2 on OCT showed that although there are about 10 years worth of research into OCT for epithelial (dense tissue) cancers and some publications on OCT of the cervix, there is actually no clinical tool available. OCT so far falls short of reliably showing the crucial difference between CIN1 and CIN2/3.

It was also pointed out that there is not yet a publication on volumetric OCT imaging of the cervix. In this case, one will expect much more data, which would have to be scanned for faint markers for lesions suspicious of early cancer. This is provided that such markers will be clearly identified, agreed on by experts and supported by trials.

Creating our own collection of OCT images seemed not enough for this purpose. We deemed it necessary to be prepared for high throughput classification. Therefore, a review on automated analysis of OCT data was performed, which constitutes Chapter 5.

9.2.2 OCT data evaluation

With the introduction of OCT, there had been attempts of automating data evaluation. This was driven by ophthalmology, with the possibility to segment multiple layers and derive features mostly related to layer thickness and integrity. For dense tissue, where such layering is more an exception than the general rule, these algorithms cannot unfold to their full potential, and subsequently there are only a few publications.

This was established after reviewing the literature on automation of OCT data (Chapter 5). This further established that many studies include manual selection as inclusion/exclusion criteria. Prior to this work, the classification of OCT data on the basis of time domain A-scans has not been thoroughly investigated for this purpose. The design of this approach is the content of Chapter 6.

9.2.3 Principal components-fed linear discriminant analysis

Some requirements need to be met in order to yield successful results with the techniques PCA and LDA: Each pixel of an A-scan vector needs to be correlated with the measured imaging depth relative to the surface.

As a proof of principle, we investigated whether OCT data from different samples can be classified. The advantage of the approach would be that a structural clinical feature, as stated in the end of Chapter 2.5, need not be identified. Rather, the degree of correlation within a model would serve as marker.

If an algorithm applied to classify clinical images fails, it would not be clear whether this is due to algorithm design, or due to image properties. In this thesis, the initial approach was therefore to attempt a classification of images with obvious differences (Chapter 7.4.1). We confirmed the assumption that for classification of time domain

A-scans, the surface needed to be normalised. Once this is achieved in a reliable manner, the algorithm is quite powerful and able to distinguish several groups.

After classifying vegetables, which poses a rather artificial task, we moved closer to the aim of classifying tissue by feeding scans of animal (porcine) organ samples. Finally, we classified different states of the same organ. In all of the scenarios, the PC-fed LDA classified up to 93.3 % correct, an estimate backed up by thorough cross-validation that has potential for higher scores (Chapter 7.4.2).

9.2.4 PC-fed LDA for analysis of epithelial precancer

All the research and design from the previous chapters should culminate in Chapter 8, the application of the algorithm to clinical data. Chapter 8 describes the tissue collection and measurement protocols for the different sources for advanced stages and milder or even non-cancerous tissue. The immediate impact from source heterogeneity was sample handling: diagnostic tissue samples are significantly smaller than interventional resections. Too small samples proved difficult for imaging, and did not allow feeding such images into analysis.

The classification performs weak: Using the entire set of useful images to separate normal/CIN1 from CIN2/3, the algorithm achieves a cross-validated score of 61.5 % and a κ value of 0.52 (Chapter 8.4.3). Although this is better than guessing, it is not encouraging for a clinical tool.

In comparison to other values in the literature, this is not convincing enough to suggest an immediate change of diagnostic procedures.

The model was built with images from *ex vivo* samples obtained using a fibre optics TDOCT system. Future measurements could provide volumetric data with added information content, e.g. from functional imaging or contrast agents. It can be expected that *in vivo* images do not suffer from problems with orientation and penetration depth.

Cervical tissue undergoes many changes which affect epithelial thickness and mucus consistency. There is therefore a large variety of tissue composition within a healthy population. In theory a large enough sampling number, i.e. patients, should be sufficient to cover these variations. In practice, we cannot ascertain that our population has covered the entire range of possible variation and it is not possible to estimate the required size for such discrimination. It would be large, if anything.

9.3 Outlook

For assessing the value of the findings, we could pose the question whether this will have an impact on cervical precancer management in the near future. Not that a direct impact would suffice or be a requirement to judge the value of this work, but it could certainly serve as an indicator.

Even if findings are promising, it is not granted that such results indeed make it into the clinical setting. Health and safety issues, ease of handling and interpretation, and comfort for the patient are important issues. (Appendix A treats the topic of “non-invasive” measurements). New protocols need to fit into existing structures and have a high chance of success only if they fit well into existing logistics, as achieved with the liquid based cytology (LBC).

It is unlikely that any modality will entirely replace a screening programme such as the NHSCSP in the near future.

- This program is well established.
- Even slight changes (reducing false negatives by LBC) take years, are still not fully implemented although more efficient and cost reducing.
- However, modalities can serve as adjunct and guidance for biopsying. Such a modality is therefore positioned as the second line of defence after an abnormal screening result.

Why then perform research on a novel tool such as OCT for the cervix, and not another organ? The answer is that the cervix is a good model for assessing clinical tools.

- Research can be integrated into existing recruitment schemes and target participants without costing them any extra time, physical harm, or worry.
- The routinely obtained results can serve as reference. A study can be designed to incur no extra cost for the reference result.
- Results are transferable to other sites of the body and countries, where such trials are not easy to perform, and where a prevention program has not yet been established. Examples are: mouth, larynx, oesophagus, stomach, upper and lower intestines, prostate, bladder, lymph nodes, or the uterus.

Can the existing literature be trusted as reference for results? A meta-analysis on 59 studies on Pap-smear staging reports a range of 11 % to 99 % for sensitivity and from

14 % to 97 % for specificity, with a highly negative correlation [221]. Such a meta-analysis has not been found for histopathological assessment, but already the few citations in this thesis show some degree of variation [215]-[220].

9.4 Suggestions to continue from this work

OCT data has long been synonymous with structure. When structure is intuitive to assess, it might seem not worth the effort of developing an algorithm.

However, in large datasets (as encountered in OCT); automation would give a time advantage. We assume that structural changes are not clearly correlated with epithelial malignancy, and changes due to the menstrual cycle as well as inter-patient variability will require a lot of expertise for evaluating the images.

How could this be addressed? From volumetric images, the intra-patient variability might be a more reliable measure for precancer. OCT is a modality quick enough to acquire a volume of the entire cervix. Developing a 3 D-tool and performing a new, second prospective study was outside the scope of this project. It is however a suggestion drawn from this study: To undertake a study on volumetric data from tissue samples.

There are other interesting topics which could not be addressed within this work:

Contrast agents

In this study, contrast agents were not investigated. Contrast agents are routinely applied in the clinical setting, and whatever their mechanism, might also enhance the precancer progression marker in OCT images. Acetic acid has been shown to improve contrast for nuclei in confocal microscopy, and the impact of acetic acid has been proven to show up as distinguishable by PC-fed LDA of animal tissue in Chapter 7.4.2. This might merit an investigation into contrast agents or refractive index matching, and their effect on CIN discrimination in OCT images.

Functional OCT

In cartilage and scar imaging, the polarisation content is more informative than a purely structural imaging. Functional OCT, a field which includes polarisation sensitive OCT, might show contrast mechanisms in precancerous tissues that are not available from structural OCT.

Quantitative refractive index measurements

The most generalising assumption in OCT is the constant refractive index correction, usually a factor of 1.4 for dense tissue. Using approaches such as dual wavelength OCT or angular measurements, a volumetric refractive index map of tissue would allow a more thorough description.

Multi-modality

So far, suggestions have been limited to OCT. Multi-modal systems could combine OCT, Raman spectroscopy, or fluorescence imaging into one system. There are practical difficulties to overcome, however combined systems have been engineered and might show a way forward for a complementing, non-destructive biochemical/structural analysis on a micrometre scale.

Application scope

In this work, the scope was disease prevention. It is well perceivable that OCT has a benefit in tumour margin assessment or in situ treatment monitoring.

Certain approaches within this work were undertaken in due course, and may have not been the optimal choice possible. These include fields which might merit a more thorough investigation that was outside the scope of this thesis:

- Research into algorithms for robust and fast surface recognition, and a scoring method for the validity of findings.
- In order to compare findings of future studies, a thorough meta-analysis of the literature on κ values for histopathological staging of cervical precancer would be of interest.
- Although the slowest step of the algorithm, the surface normalisation, was vectorised (details in Appendix B), it has not been fully optimised. A true real-time algorithm that allows overlaying classification likelihoods in the form of false colour maps would certainly be a more useful guiding tool.

9.5 Summary and final words

Referring back to Table 3.1, we can now attempt to summarise how our research questions are answered (Table 9.1):

Table 9.1: Summary of the hypotheses investigated in this thesis, findings and conclusions.

Research question	Aim	Findings	Conclusion
Can OCT be used as staging tool for pre-cancerous lesions in the cervix?	Define a role of OCT for the clinical setting in the near future.	See below.	There is potential. Possibilities for improving performance are identified.
Does OCT perform different than other high-resolution imaging tools?	Define advantages of OCT over other technologies.	Acquisition speed, non-destructive real-time imaging, and high-resolution over a wide FOV and depth were identified as key benefits.	Worthwhile to investigate OCT (speed, wide-field high-resolution).
Can OCT data evaluation in dense tissue be automated?	Define conditions under which these algorithms work.	Approaches exist but have requirements: identified marker, and manual exclusion criteria. Further, the speed is never specified.	Provided review of automated dense tissue evaluation. There might be a benefit of a quick classification tool.
Is linear discriminant analysis a suitable tool for OCT data classification?	Investigate whether a classification has a correct rate (arbitrary) of >70 %.	PC-fed LDA performed > 70 %.	Demonstrated proof of principle. PC-fed LDA can classify dense tissue OCT images.
Does automated analysis of cervical OCT images have a clinical benefit?	Investigate whether the classification is worse than standard clinical histopathology scores in literature.	For the cervical image collection, the κ values are lower than in the literature. The model of clinical significance can reach values listed in Table 8.8.	Cannot suggest OCT to replace clinical tool.

OCT is an imaging modality that has quickly found a way into clinical applications for its various advantages. An algorithm like the one presented in this work might enhance the way OCT data can be interpreted. It might sound apparent that the clinical feature needs to stand out. Even if this remains a challenge for epithelial cancers, there is potential in other fields (Appendix D).

References

- [1] Healthsquare.com. *Women's Health – Cervical Cancer*. <http://www.healthsquare.com/fgwh/wh1ch38.htm> (accessed October 2007).
- [2] Healthsquare.com. *Women's Health – Pap Test*. <http://www.healthsquare.com/fgwh/wh1ch39.htm> (accessed October 2007).
- [3] ArabicOBGYN.net. *Cervix*. <http://www.arabicobgyn.net/doc/cervix.htm> (accessed October 2007).
- [4] Burd EM. Human Papillomavirus and Cervical Cancer. *Clinical Microbiology Reviews*. **16**(1), 1-17 (2003).
- [5] Weinberg RA. How Cancer Arises. *Scientific American Magazine*. **275**(3), 62-70 (1996).
- [6] Follen M, Levenback CF, Iyer RB, Grigsby PW, Boss EA, Delpassand ES, Fornage BD, Fishman EK. Imaging in Cervical Cancer. *Cancer*, **98**(9 Suppl), 2028-2038 (2003).
- [7] Department of Health. Cervical Screening – The Facts. *NHS Cancer Screening Programmes*. Bulletin: 272855/Cervical Screening, (2006).
- [8] Shafi MI, Welton K. Colposcopy and cervical intraepithelial neoplasia. *Obstetrics, Gynaecology & Reproductive Medicine*, **17**(6), 173-180 (2007).
- [9] Gaiotto MA, Focchi J, Ribalta JL, Stávale JN, Baracat EC, Lima GR, Guerreiro da Silva ID. Comparative study of MMP-2 (matrix metalloproteinase 2) immune expression in normal uterine cervix, intraepithelial neoplasias, and squamous cells cervical carcinoma. *Am J Obstet Gynecol*. **190**(5), 1278-1282 (2004).
- [10] Sahai E. Mechanisms of cancer cell invasion. *Current Opinions in Genetics and Development*. **15**(1), 87-96 (2005).
- [11] BSCCP – The British Society for Colposcopy and Cervical Pathology. *Guidelines*. <http://www.bsccp.org.uk/index.asp?PageID=20> (accessed October 2007).
- [12] ASCCP – American Society for Colposcopy and Cervical Pathology. *Practice Recommendations – Practice Management Materials*. <http://www.asccp.org/edu/practice/cervix.shtml#colposcopy>
- [13] Addis IB, Hatch KD, Berek JS. Intraepithelial Disease of the Cervix, Vagina, and Vulva. Chapter 17 in: *Berek JS, Novak E. Berek & Novak's gynecology*, Edition 14. Lippincott Williams & Wilkins, ISBN: 0-781-76805-5, pp.561ff (2006).
- [14] Feng W, Xiao J, Zhang Z, Rosen DG, Brown RE, Liu J, Duan X. Senescence and apoptosis in carcinogenesis of cervical squamous carcinoma. *Modern Pathology*. **20**, 961-966 (2007).
- [15] Behtash N, Mehrdad N. Cervical cancer: screening and prevention. *Asian Pac J Cancer Prev*. **7**(4), 683-686 (2006).
- [16] Lindeque BG. Management of cervical premalignant lesions. *Best Pract Res Clin Obstet Gynaecol*. **19**(4), 545-561 (2005).
- [17] Kyrgiou M, Koliopoulos G, Martin-Hirsch P, Kehoe S, Flannely G, Mitrou S, Arbyn M, Prendiville W, Paraskeva E. Management of minor cytological abnormalities: a systematic review and meta-analysis of the literature. *Cancer Treat Rev*. **33**(6), 514-520 (2007).
- [18] Cancer Research UK. Cancer Stats – Cervical Cancer – Cervical Screening Programme Results <http://info.cancerresearchuk.org/cancerstats/types/cervix/screening/cervicalscreeningresults/?a=5441> (accessed October 2007)
- [19] Elit LM. Pitfalls in the diagnosis of cervical intraepithelial neoplasia. *J Low Genit Tract Dis*. **8**(3), 181-187 (2004).
- [20] Carreon JD, Sherman ME, Guillén D, Solomon D, Herrero R, Jerónimo J, Wacholder S, Rodríguez AC, Morales J, Hutchinson M, Burk RD, Schiffman M. CIN2 is a much less reproducible and less valid diagnosis than CIN3: results from a histological review of population-based cervical samples. *Int J Gynecol Pathol*. **26**(4), 441-446 (2007).

- [21] Hammes LS, Naud P, Passos EP, Matos J, Brouwers K, Rivoire W, Syrjänen KJ. Value of the International Federation for Cervical Pathology and Colposcopy (IFCPC) Terminology in predicting cervical disease. *J Low Genit Tract Dis.* **11**(3), 158-165 (2007).
- [22] Owens CL, Moats DR, Burroughs FH, Gustafson KS. "Low-grade squamous intraepithelial lesion, cannot exclude high-grade squamous intraepithelial lesion" is a distinct cytologic category: histologic outcomes and HPV prevalence. *Am J Clin Pathol.* **128**(3), 398-403 (2007).
- [23] Iatrakis G, Kourounis G, Georgopoulos N, Karachotzitis J. Treatment delay and pathology results in women with low-grade squamous intraepithelial lesions: A preliminary study. *Eur J Gynaecol Oncol.* **25**(3), 376-378 (2004).
- [24] Diakomanolis E, Haidopoulos D, Chatzipapas I, Rodolakis A, Stefanidis K, Markaki S. Negative cone biopsies: A reappraisal. *J Reprod Med.* **48**(8), 617-621 (2003). Erratum in: *J Reprod Med.* **48**(10), 833 (2003).
- [25] The Information Centre, Part of the Government Statistical Service. *Cervical screening statistics 2006/7.* ISBN: 978-1-84636-166-1. Bulletin: 2007/14/HSCIC (2007).
- [26] Martin-Hirsch P, Rash B, Martin A, Standaert B. Management of women with abnormal cervical cytology: treatment patterns and associated costs in England and Wales. *BJOG.* **114**(4), 408-415 (2007).
- [27] Basen-Engquist K, Shinn EH, Warneke C, de Moor C, Le T, Richards-Kortum R, Follen M. Patient distress and satisfaction with optical spectroscopy in cervical dysplasia detection. *Am J Obstet Gynecol.* **189**, 1136-1142 (2003).
- [28] Orbell S, Hagger M, Brown V, Tidy J. Appraisal theory and emotional sequelae of first visit to colposcopy following an abnormal cervical screening result. *Br J Health Psychol.* **9**(4), 533-555 (2004).
- [29] Balasubramani L, Orbell S, Hagger M, Brown V, Tidy J. Do women with high-grade cervical intraepithelial neoplasia prefer a see and treat option in colposcopy? *BJOG.* **114**(1), 39-45 (2007).
- [30] Livasy CA, Maygarden SJ, Rajaratnam CT, Novotny DB. Predictors of recurrent dysplasia after a cervical loop electrocautery excision procedure for CIN-3: a study of margin, endocervical gland, and quadrant involvement. *Mod Pathol.* **12**(3), 233-238 (1999).
- [31] Rojat-Habib MC, Cravello L, Bretelle F, Roger V, Liprandi A, de Burtel I, d'Ercole C, Pellissier JF, Blanc B. Value of endocervical margin examination of conization specimens. Prospective study conducted on 150 patients [Article in French]. *Gynecol Obstet Fertil.* **28**(7-8), 518-525 (2000).
- [32] Tyler LN, Andrews N, Parrish RS, Hazlett LJ, Korourian S. Significance of margin and extent of dysplasia in loop electrosurgery excision procedure biopsies performed for high-grade squamous intraepithelial lesion in predicting persistent disease. *Arch Pathol Lab Med.* **131**(4), 622-624 (2007).
- [33] Cardoza-Favarato G, Fadare O. High-grade squamous intraepithelial lesion (CIN 2 and 3) excised with negative margins by loop electrosurgical excision procedure: the significance of CIN 1 at the margins of excision. *Hum Pathol.* **38**(5), 781-786 (2007).
- [34] Maluf PJ, Adad SJ, Murta EF. Outcome after conization for cervical intraepithelial neoplasia grade III: relation with surgical margins, extension to the crypts and mitoses. *Tumori.* **90**(5), 473-477 (2004).
- [35] Johnson N, Khalili M, Hirschowitz L, Ralli F, Porter R. Predicting residual disease after excision of cervical dysplasia. *BJOG.* **110**(10), 952-955 (2003).
- [36] Fadare O, Rodriguez R. Squamous dysplasia of the uterine cervix: tissue sampling-related diagnostic considerations in 600 consecutive biopsies. *Int J Gynecol Pathol.* **26**(4), 469-474 (2007).
- [37] Kyrgiou M, Tsoumpou I, Vrekoussis T, Martin-Hirsch P, Arbyn M, Prendiville W, Mitrou S, Koliopoulos G, Dalkalitsis N, Stamatopoulos P, Paraskevaidis E. The up-to-date evidence on colposcopy practice and treatment of cervical intraepithelial neoplasia: the Cochrane colposcopy & cervical cytopathology collaborative group (C5 group) approach. *Cancer Treat Rev.* **32**(7), 516-523 (2006).

References

- [38] Tabor A, Berget A. Cold-knife and laser conization for cervical intraepithelial neoplasia. *Obstet Gynecol.* **76**(4) 633-635 (1990).
- [39] Bodner K, Bodner-Adler B, Wierrani F, Kubin A, Szölts-Szölts J, Spängler B, Grünberger W. Cold-knife conization versus photodynamic therapy with topical 5-aminolevulinic acid (5-ALA) in cervical intraepithelial neoplasia (CIN) II with associated human papillomavirus infection: a comparison of preliminary results. *Anticancer Res.* **23**(2C), 1785-1788 (2003).
- [40] Yamaguchi S, Tsuda H, Takemori M, Nakata S, Nishimura S, Kawamura N, Hanioka K, Inoue T, Nishimura R. Photodynamic therapy for cervical intraepithelial neoplasia. *Oncology.* **69**(2), 110-116, (2005).
- [41] Andikyan V, Kronschnabl M, Hillemanns M, Wang X, Stepp H, Hillemanns P. Fluorescence diagnosis with 5-ALA thermogel of cervical intraepithelial neoplasia [Article in German]. *Gynakol Geburtshilfliche Rundsch.* **44**(1), 31-37 (2004).
- [42] Huang D, Swanson EA, Lin CP, Schuman JS, Stinson WG, Chang W, Hee MR, Flotte T, Gregory K, Puliafito CA, Fujimoto JG. Optical coherence tomography. *Science* **254**, 1178-1181 (1991).
- [43] Liu B, Brezinski ME, Theoretical and practical considerations on detection performance of time domain, Fourier domain, and swept source optical coherence tomography. *Journal of Biomedical Optics.* **12**(4), 044007 (2007).
- [44] Fujimoto JG, Drexler W, Morgner U, Kärntner F, Ippen E, Optical Coherence Tomography. *Optics & Photonics News*, **2000**:24-31 (2000).
- [45] Fujimoto JG, Pitris C, Boppart SA, Brezinski ME. Optical Coherence Tomography: An Emerging Technology for Biomedical Imaging and Optical Biopsy. *Neoplasia.* **2**(1-2), 9-25 (2000).
- [46] Welzel J. Optical coherence tomography in dermatology: a review. *Skin Research and Technology.* **7**, 1-9 (2001).
- [47] Andersen PE, *Introduction – Optical coherence tomography.*
http://risoe-staged.risoe.dk/About_risoe/research_departments/OPL/LSO/OCT/~media/risoe_dk/research/oct/documents/oct_introduction_andersen.ashx (accessed December 2005).
- [48] Yang Y, Bagnaninchi PO, Whiteman SC, van Pittius DG, El Haj AJ, Spiteri MA, Wang RK. A naturally occurring contrast agent for OCT imaging of smokers' lung. *J. Phys. D, Appl. Phys.* **38**, 2590-2596 (2005).
- [49] Wang RK, Elder JB. Propylene Glycol As a Contrasting Agent for Optical Coherence Tomography to Image Gastrointestinal Tissues. *Lasers in Surgery and Medicine.* **30**, 201-208 (2002).
- [50] He Y, Wang RK. Dynamic optical clearing effect of tissue impregnated with hyperosmotic agents and studied with optical coherence tomography. *J. Biomed. Opt.* **9**(1), 200-206 (2004).
- [51] Agrawal A, Utzinger U, Brookner C, Pitris C, Follen Mitchell M, Richards-Kortum R. Fluorescence Spectroscopy of the Cervix. Influence of Acetic Acid, Cervical Mucus, and Vaginal Medications. *Lasers in Surgery and Medicine.* **25**, 237-249 (1999).
- [52] Collier T, Shen P, de Pradier B, Sung K-B, Richards-Kortum R. Near Real Time Confocal Microscopy of Amelanotic Tissue: Dynamics of Aceto-Whitening Enable Nuclear Segmentation. *Optics Express.* **6**(2), 40-48 (2000).
- [53] Lee TM, Oldenburg AL, Sitafalwalla S, Marks DL, Luo W, Toublan FJ-J, Suslick KS, Boppart SA. Engineered microsphere contrast agents for optical coherence tomography. *Optics Letters.* **28**(17), 1546-1548 (2003).
- [54] Chen J, Saeki F, Wiley BJ, Cang H, Cobb MJ, Li ZY, Au L, Zhang H, Kimmey MB, Li X, Xia Y. Gold Nanocages- Bioconjugation and Their Potential Use as Optical Imaging Contrast Agents. *Nano Letters.* **5**(3), 473-477 (2005).
- [55] Sokolov K, Follen M, Aaron J, Pavlova I, Malpica A, Lotan R, Richards-Kortum R. Real-Time Vital Optical Imaging of Precancer Using Anti-Epidermal Growth Factor Receptor Antibodies Conjugated to Gold Nanoparticles. *Cancer Research.* **63**, 1999-2004 (2003).

- [56] Xu C, Ye J, Marks DL, Boppart SA. Near-infrared dyes as contrast-enhancing agents for spectroscopic optical coherence tomography. *Opt. Lett.* **29**, 1647-1649 (2004).
- [57] Oldenburg AL, Toublan FJJ, Suslick KS, Wei A, Boppart SA. Magnetomotive contrast for in vivo optical coherence tomography. *Optics Express*. **13**(17), 6597-6614 (2005).
- [58] U.S. National Library of Medicine and the National Institutes of Health. PubMed Service. <http://www.ncbi.nlm.nih.gov/sites/entrez> (accessed between November 2004 and October 2007)
- [59] Thompson Reuters. ISI Web of Knowledge. <http://isiwebofknowledge.com/> (accessed between November 2006 and October 2007)
- [60] Drezek RA, Richards-Kortum R, Brewer MA, Feld MS, Pittis C, Ferenczy A, Faupel ML, Follen M. Optical imaging of the cervix. *Cancer*, **98**(9 Suppl), 2015-2027 (2003).
- [61] Sokolov K, Follen M, Richards-Kortum R. Optical spectroscopy for detection of neoplasia. *Current Opinion in Chemical Biology*. **6**, 651-658 (2002).
- [62] Sokolov K, Aaron J, Hsu B, Nida D, Gillenwater A, Follen M, MacAulay C, Adler-Storthz K, Korgel B, Descour M, Pasqualini R, Arap W, Lam W, Richards-Kortum R. Optical Systems for In Vivo Molecular Imaging of Cancer. *Technology in Cancer Research & Treatment*. **2**(6), 491-504 (2003).
- [63] Baak JP, Kruse AJ, Robboy SJ, Janssen EA, van Diermen B, Skaland I. Dynamic behavioural interpretation of cervical intraepithelial neoplasia with molecular biomarkers. *J Clin Pathol*. **59**(10), 1017-1028 (2006).
- [64] Mathur SP, Mathur RS, Creasman WT, Underwood PB, Kohler M. Early non-invasive diagnosis of cervical cancer: beyond Pap smears and human papilloma virus (HPV) testing. *Cancer Biomark*. **1**(2-3), 183-191 (2005).
- [65] Brink AA, Snijders PJ, Meijer CJ. HPV detection methods. *Dis Markers*. **23**(4), 273-281 (2007).
- [66] Syrjänen KJ. Immunohistochemistry in assessment of molecular pathogenesis of cervical carcinogenesis. *Eur J Gynaecol Oncol*. **26**(1), 5-19 (2005).
- [67] Shack R, Baker R, Buchroeder R, Hillman D, Shoemaker R, Bartels PH. Ultrafast laser scanner microscope. *J Histochem Cytochem*. **27**(1), 153-159 (1979).
- [68] Takamatsu T, Fujita S. Microscopic tomography by laser scanning microscopy and its three-dimensional reconstruction. *J Microsc*. **149**(3), 167-174 (1988).
- [69] Carlsson K. Three-dimensional specimen reconstruction by confocal microscopy and digital image processing. *Bull Assoc Anat (Nancy)*. **75**(229), 105-108 (1991).
- [70] Boon ME, Kok LP, Sutedia G, Dutrieux RP. Confocal sectioning of thick, otherwise undiagnosable cell groupings in cervical smears. *Acta Cytol*. **37**(1), 40-48 (1993).
- [71] Kahn E, Lizard G, Frouin F, Roignot P, Chardonnet Y, Di Paola R. Factor analysis of confocal image sequences of human papillomavirus DNA revealed with fast red in cervical tissue sections stained with TOTO-iodide. *Anal Quant Cytol Histol*. **22**(2), 168-174 (2000).
- [72] Crist, K. A. Kim, K. Goldblatt, P. J. Boone, C. W. Kelloff, G. J. You, M. DNA Quantification in Cervical Intraepithelial Neoplasia Thick Tissue Sections by Confocal Laser Scanning Microscopy. *Journal Of Cellular Biochemistry*. **25**(Suppl.), 49-56 (1996).
- [73] Pavlova I, Sokolov K, Drezek R, Malpica A, Follen M, Richards-Kortum R. Microanatomical and Biochemical Origins of Normal and Precancerous Cervical Autofluorescence Using Laser-scanning Fluorescence Confocal Microscopy. *Photochemistry and Photobiology*. **77**(5), 550-555, (2003).
- [74] Drezek RA, Collier T, Brookner CK, Malpica A, Lotan R, Richards-Kortum R, Follen M. Laser scanning confocal microscopy of cervical tissue before and after application of acetic acid. *Am J Obstet Gynecol*. **182**(5), 1135-1139 (2000).
- [75] Zuluaga AF, Drezek R, Collier T, Lotan R, Follen M, Richards-Kortum R. Contrast agents for confocal microscopy: how simple chemicals affect confocal images of normal and cancer cells in suspension. *Journal of Biomedical Optics*. **7**(3), 398-403 (2002).

References

- [76] Carlson KD, Pavlova I, Collier T, Descour M, Follen M, Richards-Kortum R. Confocal microscopy: Imaging cervical precancerous lesions. *Gynecol. Oncol.* **99**(3), 84-88 (2005).
- [77] Collier T, Follen M, Malpica A, Richards-Kortum R. Sources of scattering in cervical tissue - determination of the scattering coefficient by confocal microscopy. *Applied Optics.* **44**(10), 2072-2081 (2005).
- [78] Collier T, Guillaud M, Follen M, Malpica A, Richards-Kortum R. Real Time Reflectance Confocal Microscopy: Comparison of Two Dimensional Images and Three Dimensional Image Stacks for Detection of Cervical Precancer. *Journal of Biomedical Optics.* **12**(2), 024021 (2007).
- [79] Sung K-B, Liang C, Descour M, Collier T, Follen M, Richards-Kortum R. Fiber-Optic Confocal Reflectance Microscope With Miniature Objective for In Vivo Imaging of Human Tissues. *IEEE Transactions On Biomedical Engineering.* **49**(10) 1168-1172 (2002).
- [80] Jean F, Bourg-Heckly G, Viellerobe B, Fibered confocal spectroscopy and multicolor imaging system for in vivo fluorescence analysis. *Opt. Express.* **15**, 4008-4017 (2007).
- [81] Zhuo S, Chen J, Luo T, Jiang X, Xie S, Chen R. Two-layered multiphoton microscopic imaging of cervical tissue. *Lasers Med Sci.* [Epub ahead of print], doi 10.1007/s10103-008-0570-2, (2008).
- [82] Skala MC, Squirrell JM, Vrotsos KM, Eickhoff JC, Gendron-Fitzpatrick A, Eliceiri KW, Ramanujam N. Multiphoton Microscopy of Endogenous Fluorescence Differentiates Normal, Precancerous, and Cancerous Squamous Epithelial Tissues. *Cancer Res.* **65**(4), 1180-1186, (2005).
- [83] Carlson K, Chidley M, Sung KB, Descour M, Gillenwater A, Follen M, Richards-Kortum R. In vivo fiber-optic confocal reflectance microscope with an injection-molded plastic miniature objective lens. *Appl Opt.* **44**(10), 1792-1797 (2005).
- [84] Luck B, Carlson K, Collier T, Sung K-B. Confocal microscopy [detecting and diagnosing cancers]. *IEEE Potentials.* **23**(1), 14-17, (2004).
- [85] Choi HJ, Choi IH, Kim TY, Cho NH, Choi HK. Three-dimensional visualization and quantitative analysis of cervical cell nuclei with confocal laser scanning microscopy. *Anal Quant Cytol Histol.* **27**(3), 174-180 (2005).
- [86] Luck B, Bovic A, Richards-Kortum R. Segmenting cervical epithelial nuclei from confocal images Gaussian Markov random fields. In: *Image Processing 2003, ICIP 2003, Vol. 2*, 14-17 (2003).
- [87] Rahman M, Abd-El-Barr M, Mack V, Tkaczyk T, Sokolov K, Richards-Kortum R, Descour M. Optical imaging of cervical pre-cancers with structured illumination: An integrated approach. *Gynecologic Oncology.* **99**(3, Suppl.), 112-115 (2005).
- [88] Collier T, Lacy A, Richards-Kortum R, Malpica A, Follen M. Near real-time confocal microscopy of amelanotic tissue: detection of dysplasia in ex vivo cervical tissue. *Acad Radiol.* **9**(5), 504-512 (2002).
- [89] Tomlins PH, Wang RK. Theory, developments and applications of optical coherence tomography. *J. Phys. D, Appl. Phys.* **38**, 2519-2535 (2005).
- [90] Yang C. Molecular Contrast Optical Coherence Tomography: A Review. *Photochem Photobiol.* **81**(2), 215-237 (2005).
- [91] Podoleanu AG. Optical coherence tomography. *British Journal of Radiology.* **78**, 976-988 (2005).
- [92] Jenkins MW, Adler DC, Gargasha M, Huber R, Rothenberg F, Belding J, Watanabe M, Wilson DL, Fujimoto JG, Rollins AM. Ultrahigh-speed optical coherence tomography imaging and visualization of the embryonic avian heart using a buffered Fourier Domain Mode Locked laser. *Opt. Express.* **15**, 6251-6267 (2007).
- [93] Sergeev AM, Gelikonov V, Gelikonov G, Feldchtein F, Kuranov R, Gladkova N, Shakhova N, Snopova L, Shakhov A, Kuznetsova I, Denisenko A, Pochinko V, Chumakov Y, Streltsova O. In vivo endoscopic OCT imaging of precancer and cancer states of human mucosa. *Opt. Express.* **1**(13), 432-440 (1997).

- [94] Feldchtein FI, Gelikonov GV, Gelikonov VM, Kuranov RV, Sergeev AM, Gladkova ND, Shakhov AV, Shakhova NM, Snopova LB, Terent'eva AB, Zagainova EV, Chumakov Y, Kuznetzova IA. Endoscopic applications of optical coherence tomography. *Optics Express*. **3**(6), 257-270 (1998).
- [95] Pitris C, Goodman A, Boppart SA, Libus JJ, Fujimoto JG, Brezinski ME. High-resolution imaging of gynecologic neoplasms using optical coherence tomography. *Obstet Gynecol*. **93**(1), 135-139 (1999).
- [96] Escobar PF, Belinson JL, White A, Shakhova NM, Feldchtein FI, Kareta MV, Gladkova ND. Diagnostic efficacy of optical coherence tomography in the management of preinvasive and invasive cancer of uterine cervix and vulva. *Int J Gynecol Cancer*. **14**(3), 470-474 (2004).
- [97] Sapozhnikova VV, Shakhova NM, Kamensky VA, Petrova SA, Snopova LB, Kuranov RV. Capabilities of fluorescence spectroscopy using 5-ALA and optical coherence tomography for diagnosis of neoplastic processes in the uterine cervix and vulva. *Laser Physics*. **15**(12), 1664-1673 (2005).
- [98] Escobar PF, Rojas-Espaillet L, Belinson JL. Optical diagnosis of cervical dysplasia. *Int J Gynaecol Obstet*. **89**(1), 63-64 (2005).
- [99] Zuluaga AF, Follen M, Boiko I, Malpica A, Richards-Kortum R. Optical coherence tomography: a pilot study of a new imaging technique for noninvasive examination of cervical tissue. *Am J Obstet Gynecol*. **193**(1), 83-88 (2005).
- [100] Escobar PF, Rojas-Espaillet L, Tisci S, Enerson C, Brainard J, Smith J, Tresser NJ, Feldchtein FI, Rojas LB, Belinson JL. Optical coherence tomography as a diagnostic aid to visual inspection and colposcopy for preinvasive and invasive cancer of the uterine cervix. *Int J Gynecol Cancer*. **16**(5), 1815-1822 (2006).
- [101] Hsiung PL, Pantanowitz L, Aguirre AD, Chen Y, Phatak D, Ko TH, Bourquin S, Schnitt SJ, Raza S, Connolly JL, Mashimo H, Fujimoto JG. Ultrahigh-resolution and 3-dimensional optical coherence tomography ex vivo imaging of the large and small intestines. *Gastrointest Endosc*. **62**(4), 561-574 (2005).
- [102] Turchin IV, Sergeeva EA, Dolin LS, Kamensky VA, Shakhova NM, Richards-Kortum R. Novel algorithm of processing optical coherence tomography images for differentiation of biological tissue pathologies. *J Biomed Opt*. **10**(6), 064024 (2005).
- [103] Yun SH, Tearney GJ, Vakoc BJ, Shishkov M, Oh WY, Desjardins AE, Suter MJ, Chan RC, Evans JA, Jang I, Nishioka NS, de Boer JF, Bouma BE. Comprehensive volumetric optical microscopy in vivo. *Nature Medicine*. **12**, 1429-1433 (2006).
- [104] Barton JK, Guzman F, Tumlinson A. Dual modality instrument for simultaneous optical coherence tomography imaging and fluorescence spectroscopy. *J. Biomed. Opt.* **9**(3), 618-623 (2004).
- [105] Wong PTT, Wong RK, Caputo TA, Godwin TA, Rigas B. Infrared spectroscopy of exfoliated human cervical cells: evidence of extensive structural changes during carcinogenesis. *Proc. Natl. Acad. Sci. USA*. **88**(24), 10988-10992 (1991).
- [106] Walsh M, German M, Singh M, Pollock H, Hammiche A, Kyrgiou M, Stringfellow H, Paraskevaidis E, Martin-Hirsch P, Martin F. IR microspectroscopy: potential applications in cervical cancer screening. *Cancer Letters*, **246**(1-2), 1-11 (2007).
- [107] Cohenford MA, Godwin TA, Cahn F, Bhandare P, Caputo TA, Rigas B. Infrared spectroscopy of normal and abnormal cervical smears: evaluation by principal component analysis. *Gynecol Oncol*. **66**(1), 59-65 (1997).
- [108] Fung Kee Fung M, Senterman M, Eid P, Faught W, Mikhael NZ, Wong PT. Comparison of Fourier-transform infrared spectroscopic screening of exfoliated cervical cells with standard Papanicolaou screening. *Gynecol Oncol*. **66**(1), 10-15 (1997).
- [109] Chiriboga L, Xie P, Yee H, Vigorita V, Zarou D, Zakim D, Diem M. Infrared spectroscopy of human tissue. I. Differentiation and maturation of epithelial cells in human cervix. *Biospectroscopy*. **4**, 47-53 (1998).

- [110] Wood BR, Quinn MA, Tait B, Ashdown M, Hislop T, Romeo M, McNaughton D. FTIR microspectroscopic study of cell types and potential confounding variables in screening for cervical malignancies. *Biospectroscopy*. **4**(2), 75-91 (1998).
- [111] Neviliappan S, Fang Kan L, Tiang Lee Walter T, Arulkumaran S, Wong PT. Infrared spectral features of exfoliated cervical cells, cervical adenocarcinoma tissue, and an adenocarcinoma cell line (SiSo). *Gynecol Oncol*. **85**(1), 170-174 (2002).
- [112] Cohenford MA, Rigas B. Cytologically normal cells from neoplastic cervical samples display extensive structural abnormalities on IR spectroscopy: implications for tumor biology. *Proc Natl Acad Sci USA*. **95**(26), 15327-15332 (1998).
- [113] Chang JI, Huang YB, Wu PC, Chen CC, Huang SC, Tsai YH. Characterization of human cervical precancerous tissue through the fourier transform infrared microscopy with mapping method. *Gynecol Oncol*. **91**(3), 577-583 (2003).
- [114] Podshyvalov A, Sahu RK, Mark S, Kantarovich K, Guterman H, Goldstein J, Jagannathan R, Argov S, Mordechai S. Distinction of cervical cancer biopsies by use of infrared microspectroscopy and probabilistic neural networks. *Appl Opt*. **44**(18), 3725-3734 (2005).
- [115] Mordechai S, Sahu RK, Hammody Z, Mark S, Kantarovich K, Guterman H, Podshyvalov A, Goldstein J, Argov S. Possible common biomarkers from FTIR microspectroscopy of cervical cancer and melanoma. *J Microsc*. **215**(1), 86-91 (2004).
- [116] Lowry SR. The analysis of exfoliated cervical cells by infrared microscopy. *Cell Mol Biol*. **44**(1), 169-177 (1998).
- [117] Romeo M, Burden F, Quinn M, Wood B, McNaughton D. Infrared microspectroscopy and artificial neural networks in the diagnosis of cervical cancer. *Cell Mol Biol*. **44**(1), 179-187 (1998).
- [118] Romeo MJ, Quinn MA, Burden FR, McNaughton D. Influence of benign cellular changes in diagnosis of cervical cancer using IR microspectroscopy. *Biopolymers*. **67**(4-5), 362-366 (2002).
- [119] Wong PTT, Senterman MK, Jackli P, Wong RK, Salib S, Campbell CE, Feigel R, Faught W, Fung MFK. Detailed account of confounding factors in interpretation of FTIR spectra of exfoliated cervical cells. *Biopolymers*. **67**(6), 376-386 (2002).
- [120] Chiriboga L, Yee H, Diem M, Wood B. Comment on: Neviliappan S. *et al*. Infrared spectral features of exfoliated cervical cells, cervical adenocarcinoma tissue, and an adenocarcinoma cell line (SiSo). *Gynecol Oncol*. **85**(1), 170-174 (2002). *Gynecol Oncol*. **91**(1), 275-276 (2003). [Author reply *Gynecol Oncol*. **91**(1), 276-277 (2003)].
- [121] Chiriboga L, Xie P, Yee H, Zarou D, Zakim D, Diem M. Infrared spectroscopy of human tissue. IV. Detection of human cervical tissue via infrared microscopy. *Cell. Mol. Biol. Lett*. **44**, 219-229 (1998).
- [122] Romeo MJ, Wood BR, Quinn MA, McNaughton D. Removal of blood components from cervical smears: implications for cancer diagnosis using FTIR spectroscopy. *Biopolymers*. **72**(1), 69-76 (2003).
- [123] Morris BJ, Lee C, Nightingale BN, Molodsky E, Morris LJ, Appio R, Sternhell S, Cardona M, Mackerras D, Irwig LM. Fourier transform infrared spectroscopy of dysplastic, papillomavirus-positive cervicovaginal lavage specimens. *Gynecol Oncol*. **56**(2), 245-249 (1995).
- [124] Walsh MJ, Singh MN, Pollock HM, Cooper LJ, German MJ, Stringfellow HF, Fullwood NJ, Paraskevaidis E, Martin-Hirsch PL, Martin FL. ATR microspectroscopy with multivariate analysis segregates grades of exfoliative cervical cytology. *Biochem Biophys Res Commun*. **352**(1), 213-219 (2007).
- [125] Wood BR, Chiriboga L, Yee H, Quinn MA, McNaughton D, Diem M. Fourier transform infrared (FTIR) spectral mapping of the cervical transformation zone, and dysplastic squamous epithelium. *Gynecol Oncol*. **93**(1), 59-68 (2004).
- [126] Steller W, Einkenkel J, Horn LC, Braumann UD, Binder H, Salzer R, Krafft C. Delimitation of squamous cell cervical carcinoma using infrared microspectroscopic imaging. *Anal Bioanal Chem*. **384**(1), 145-154 (2006).

- [127] Mark S, Sahu RK, Kantarovich K, Podshyvalov A, Guterman H, Goldstein J, Jagannathan R, Argov S, Mordechai S. Fourier transform infrared microspectroscopy as a quantitative diagnostic tool for assignment of premalignancy grading in cervical neoplasia. *J Biomed Opt.* **9**(3), 558-567 (2004).
- [128] Wood BR, Bambery KR, Evans CJ, Quinn MA, McNaughton D. A three-dimensional multivariate image processing technique for the analysis of FTIR spectroscopic images of multiple tissue sections. *BMC Medical Imaging.* **6**(12), (2006).
- [129] Stone N, Kendall C, Shepherd N, Crow P, Barr H. Near-Infrared Raman Spectroscopy for the Classification of Epithelial Pre-Cancers and Cancers. *Journal of Raman Spectroscopy.* **33**(7), 564-573 (2002).
- [130] Utzinger U, Richards-Kortum RR. Fiber optic probes for biomedical optical spectroscopy. *J Biomed Opt.* **8**(1), 121-147 (2003).
- [131] Mahadevan-Jansen A, Follen-Mitchell M, Ramanujam N, Utzinger U, Richards-Kortum R. Development of a Fiber Optic Probe to Measure NIR Raman Spectra of Cervical Tissue In Vivo. *Photochemistry and Photobiology.* **68**(3), 427-431 (1998).
- [132] Yazdi Y, Ramanujam N, Lotan R, Follen-Mitchell M, Hittelman W, Richards-Kortum R. Resonance Raman spectroscopy at 257nm excitation of normal and malignant cultured breast and cervical cells. *Appl. Spectrosc.* **53**, 82-85 (1999).
- [133] Lieber CA, Mahadevan-Jansen A. Automated method for subtraction of fluorescence from biological Raman spectra. *Appl Spectrosc.* **57**(11), 1363-1367 (2003).
- [134] Hutchings J, Kendall CA, Stone N. Rapid Raman microscopic biochemical imaging for potential histological screening. In: *Proc. SPIE*; 6853A-33 (2008).
- [135] Hart-Prieto MC, Matousek P, Towrie M, Parker AW, Wright M, Ritchie AW, Stone N. The use of picosecond Kerrgated Raman spectroscopy to suppress signals from both surface and deep layers in bladder and prostate tissue. *Journal of Biomedical Optics*, **10**, 044006 (2005).
- [136] Krishna CM, Prathima NB, Malini R, Vadhiraja BM, Bhatt RA, Fernandes DJ, Kushtagi P, Vidyasagar MS, Kartha VB. Raman spectroscopy studies for diagnosis of cancers in human uterine cervix. *Vibrational Spectroscopy.* **41**(1), 136-141 (2006).
- [137] Stone N, Kendall C, Smith J, Crow P, Barr H. Raman spectroscopy for identification of epithelial cancers. *Faraday Discussions: Applications of Spectroscopy to Biomedical Problems.* **126**, 141-157 (2004).
- [138] Liu CH, Das BB, Sha Glassman WL, Tang GC, Yoo KM, Zhu HR, Akins DL, Lubicz SS, Cleary J, Prudente R, *et al.* Raman, fluorescence, and time-resolved light scattering as optical diagnostic techniques to separate diseased and normal biomedical media. *J Photochem Photobiol B.* **16**(2), 187-209 (1992).
- [139] Mahadevan-Jansen A, Follen-Mitchell M, Ramanujam N, Malpica A, Thomsen S, Utzinger U, Richards-Kortum R. Near-Infrared Raman Spectroscopy for In Vitro Detection of Cervical Precancers. *Photochem Photobiol.* **68**(1), 123-132 (1998).
- [140] Mourant JR, Short KW, Carpenter S, Kunapareddy N, Coburn L, Powers TM, Freyer JP. Biochemical differences in tumorigenic and nontumorigenic cells measured by Raman and infrared spectroscopy. *J Biomed Opt.* **10**(3), 031106 (2005).
- [141] Utzinger U, Heintzelman DL, Mahadevan-Jansen A, Malpica A, Follen M, Richards-Kortum R. Near-Infrared Raman Spectroscopy for in Vivo Detection of Cervical Precancers. *Applied Spectroscopy*, **55**(8), 955-959 (2001).
- [142] Lyng FM, Faoláin EÓ, Conroy J, Meade AD, Knief P, Duffy B, Hunter MB, Byrne JM, Kelehan P, Byrne HJ. Vibrational spectroscopy for cervical cancer pathology. From biochemical analysis to diagnostic tool. *Experimental and Molecular Pathology*, **82**, 121-129 (2007).
- [143] Robichaux-Viehoever A, Kanter E, Shappell H, Billheimer D, Jones III H, Mahadevan-Jansen A. Characterization of Raman Spectra Measured in Vivo for the Detection of Cervical Dysplasia. *Applied Spectroscopy*, **61**(9), 986-993 (2007).

References

- [144] Jess PRT, Smith DDW, Mazilu M, Dholakia K, Riches AC, Herrington CS. Early detection of cervical neoplasia by Raman spectroscopy. *Int. J. Cancer*. **121**, 2723-2728 (2007).
- [145] Krafft C, Sergo V. Biomedical applications of Raman and infrared spectroscopy to diagnose tissues. *Spectroscopy*. **20**, 195-218 (2006).
- [146] Krishna CM, Sockalingum GD, Vadhiraja BM, Maheedhar K, Rao AC, Rao L, Venteo L, Pluot M, Fernandes DJ, Vidyasagar MS, Kartha VB, Manfait M. Vibrational spectroscopy studies of formalin-fixed cervix tissues. *Biopolymers*. **85**(3), 214-221 (2007).
- [147] Faoláin EÓ, Hunter MB, Byrne JM, Kelehan P, Lambkin HA, Byrne HJ, Lyng FM. Raman spectroscopic evaluation of efficacy of current paraffin wax section dewaxing agents. *J Histochem Cytochem*. **53**(1), 121-129 (2005).
- [148] Huang Z, McWilliams A, Lam S, English J, McLean DI, Lui H, Zeng H. Effect of formalin fixation on the near-infrared Raman spectroscopy of normal and cancerous human bronchial tissues. *Int J Oncol*. **23**(3), 649-655 (2003).
- [149] Motz JT, Gandhi SJ, Scepanovic OR, Haka AS, Kramer JR, Dasari RR, Feld MS. Real-time Raman system for in vivo disease diagnosis. *J Biomed Opt*. **10**(3), 031113 (2005).
- [150] Wu TT, Qu JY. Optical imaging for medical diagnosis based on active stereo vision and motion tracking. *Opt. Express*. **15**, 10421-10426 (2007).
- [151] Majumder SK, Kanter E, Robichaux-Viehoever A, Jones III H, Mahadevan-Jansen A. Near-infrared Raman spectroscopy for in-vivo diagnosis of cervical dysplasia: a probability-based multi-class diagnostic algorithm. In: *Proc. SPIE* 6430, 64300Q (2007).
- [152] Baker R, Matousek P, Ronayne KL, Parker AW, Rogers K, Stone N. Depth profiling of calcifications in breast tissue using picosecond Kerr-gated Raman spectroscopy. *Analyst*. **132**(1), 48-53 (2007).
- [153] Kendall C, Stone N, Shepherd N, Geboes K, Warren B, Bennett R, Barr H. Raman spectroscopy, a potential tool for the objective identification and classification of neoplasia in Barrett's oesophagus. *J. Pathol*. **200**, 602-609 (2003).
- [154] Orfanoudaki IM, Themelis GC, Sifakis SK, Fragouli DH, Panayiotides JG, Vazgiouraki EM, Koumantakis EE. A clinical study of optical biopsy of the uterine cervix using a multispectral imaging system. *Gynecol Oncol*. **96**(1), 119-131 (2005).
- [155] Alvarez RD, Wright TC. Effective cervical neoplasia detection with a novel optical detection system: A randomized trial. *Gynecologic Oncology*, **104**(2), 281-289 (2007).
- [156] Lurie S, Eliaz M, Boaz M, Levy T, Golan A, Sadan O. Distribution of cervical intraepithelial neoplasia across the cervix is random. *Am J Obstet Gynecol*. **196**(2), 125 (2007).
- [157] Franco EL, Duarte-Franco E, Ferenczy A. Prospects for controlling cervical cancer at the turn of the century. *Salud pública de méxico*, **45**(3 Suppl), 367-375 (2003).
- [158] Monsonego J. HPV infections and cervical cancer prevention. Priorities and new directions. In: *Highlights of EUROGIN 2004 International Expert Meeting*, Nice, France, October 21-23, 2004. *Gynecologic Oncology* (2004).
- [159] Vergote I, van der Zee AGJ, Kesic V, Sert B, Robova H, Rob L, Reed N, Luesley DM, Leblanc E, Hagen B, Gitsch G, Du Bois A, Di Vagno G, Colombo N, Beller U, Ayhan A, Jacobs I. ESGO statement on cervical cancer vaccination. *Int J Gynecol Cancer*. **17**, 1183-1185 (2007).
- [160] Ellis DI, Goodacre R. Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and Raman spectroscopy. *Analyst* **131**, 875-885 (2006).
- [161] Stelzer EHK. Contrast, resolution, pixelation, dynamic range and signal-to-noise ratio: fundamental limits to resolution in fluorescence light microscopy. *Journal of Microscopy*. **189**(1), 15-24 (1998).
- [162] Patil CA, Bosschaart N, Keller MD, van Leeuwen TG, Mahadevan-Jansen A, (2007) Combined Raman spectroscopy and optical coherence tomography device for tissue characterization. *Optics Letters* **33**(10):1135-1137

- [163] Kanter E, Walker R, Marion S, Hoyer P, Barton JK, (2005) Optical coherence tomography imaging and fluorescence spectroscopy of a novel rat model of ovarian cancer. *Progress in biomedical optics and imaging* 6(30):58610P.1-58610P.8
- [164] Xi C, Marks DL, Parikh DS, Raskin L, Boppart SA. Structural and functional imaging of 3D microfluidic mixers using optical coherence tomography. *Proc Natl Acad Sci.* **101**(20), 7516-7521 (2004).
- [165] Bonesi M, Churmakov DY, Ritchie LJ, Meglinski IV. Turbulence monitoring with Doppler Optical Coherence Tomography. *Laser Physics Letters.* **4**(4), 304-307 (2006).
- [166] Park B, Pierce M, Cense M, de Boer J. Real-time multi-functional optical coherence tomography. *Opt. Express.* **11**, 782-793 (2003).
- [167] Su J, Tomov IV, Jiang Y, Chen Z. High-resolution frequency-domain second-harmonic optical coherence tomography. *Appl. Opt.* **46**, 1770-1775 (2007).
- [168] Lazebnik M, Marks DL, Potgieter K, Gillette R, Boppart SA. Functional optical coherence tomography for detecting neural activity through scattering changes. *Opt. Lett.* **28**, 1218-1220 (2003).
- [169] Maheswari RU, Takaoka H, Kadono H, Homma R, Tanifuji M. Novel functional imaging technique from brain surface with optical coherence tomography enabling visualization of depth resolved functional structure in vivo. *J Neurosci Methods.* **124**(1), 83-92 (2003).
- [170] Morgner U, Drexler W, Kärtner FX, Li XD, Pitris C, Ippen EP, Fujimoto JG. Spectroscopic optical coherence tomography. *Opt. Lett.* **25**, 111-113 (2000).
- [171] Faber DJ, Mik EG, Aalders MC, van Leeuwen TG. Toward assessment of blood oxygen saturation by spectroscopic optical coherence tomography. *Opt Lett.* **30**(9), 1015-1017 (2005).
- [172] Knüttel A, Boehlau-Godau M. Spatially confined and temporally resolved refractive index and scattering evaluation in human skin performed with optical coherence tomography. *J. Biomed. Opt.* **5**(1), 83-92 (2000).
- [173] Kholodnykh AI, Petrova IY, Larin KV, Motamedi M, Esenaliev RO. Precision of Measurement of Tissue Optical Properties with Optical Coherence Tomography. *Appl. Opt.* **42**, 3027-3037 (2003).
- [174] Esenaliev RO, Larin KV, Larina IV, Motamedi M. Noninvasive monitoring of glucose concentration with optical coherence tomography. *Opt. Lett.* **26**, 992-994 (2001).
- [175] van der Meer FJ, Faber DJ, Sassoon DMB, Aalders MC, Pasterkamp G, van Leeuwen TG. Localized measurement of optical attenuation coefficients of atherosclerotic plaque constituents by quantitative optical coherence tomography. *IEEE Transactions on Medical Imaging*, **24**(10), 1369-1376 (2005).
- [176] Levitz D, Thrane L, Frosz MH, Andersen PE, Andersen CB, Valanciunaite J, Swartling J, Andersson-Engels S, Hansen PR. Determination of optical scattering properties of highly-scattering media in optical coherence tomography images. *Optics Express.* **12**(2) 249-259, (2004).
- [177] Jeon SW, Shure MA, Baker KB, Huang D, Rollins AM, Chahlavi A, Rezaia AR. A feasibility study of optical coherence tomography for guiding deep brain probes. *J Neurosci Methods* **154**(1-2), 96-101 (2006).
- [178] Ramrath L, Hofmann UG, Huettmann G, Moser A, Schweikard A (2007) Towards Automated OCT-Based Identification of White Brain Matter. Pp. 414-418 In: *Bildverarbeitung für die Medizin - Algorithmen – Systeme – Anwendungen.* Pub.: Springer Berlin Heidelberg, ISBN 978-3-540-71091-2.
- [179] Kuranov RV, Sapozhnikova VV, Prough DS, Cicenaitis I, Esenaliev RO. In vivo study of glucose-induced changes in skin properties assessed with optical coherence tomography. *Phys. Med. Biol.* **51** 3885-3900 (2006).
- [180] Kaiser JF. On a simple algorithm to calculate the 'energy' of a signal. *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '90).* **1**: 381-384 (1990).

- [181] Gossage KW, Tkaczyk TS, Rodriguez JJ, Barton JK. Texture analysis of optical coherence tomography images: feasibility for tissue classification. *J. Biomed. Opt.* **8**(3), 570-575 (2003).
- [182] Hillman TR, Adie SG, Seemann V, Armstrong JJ, Jacques SL, Sampson DD. Correlation of static speckle with sample properties in optical coherence tomography. *Opt. Lett.* **31**, 190-192 (2006).
- [183] Brodatz. *Textures*. Dover Publications Inc., ISBN: 0-486-40699-7, (2000).
- [184] Randen T. *Brodatz Textures*. <http://www.uu.uio.no/~tranden/brodatz.html> (accessed August 2007).
- [185] Gossage KW, Smith CM, Kanter EM, Hariri LP, Stone AL, Rodriguez JJ, Williams SK, Barton JK. Texture analysis of speckle in optical coherence tomography images of tissue phantoms. *Phys. Med. Biol.* **51**, 1563-1575 (2006).
- [186] Yu K, Ji L, Wang L, Xue P. How to optimize OCT image. *Optics Express*. **9**(1), 24-35 (2001).
- [187] MacNeill BD, Jang I-K, Bouma BE, Iftimia N, Takano M, Yabushita H, Shishkov M, Kauffman CR, Houser SL, Aretz HT, DeJoseph D, Halpern EF, Tearney GJ. Focal and Multi-Focal Plaque Macrophage Distributions in Patients With Acute and Stable Presentations of Coronary Artery Disease. *Journal of the American College of Cardiology*. **44**(5), 972-979 (2004).
- [188] Qi X, Sivak Jr. MV, Isenberg G, Willis JE, Rollins AM. Computer-aided diagnosis of dysplasia in Barrett's esophagus using endoscopic optical coherence tomography. *Journal of Biomedical Optics*. **11**(4) 044010 (2006).
- [189] Chen Y, Aguirre AD, Hsiung P-L, Huang S-W, Mashimo H, Schmitt JM, Fujimoto JG. Effects of axial resolution improvement on optical coherence tomography (OCT) imaging of gastrointestinal tissues. *Optics Express* **16**(4), 2469-2485, (2008).
- [190] Lingley-Papadopoulos CA, Loew MH, Manyak MJ, Zara JM. Computer recognition of cancer in the urinary bladder using optical coherence tomography and texture analysis. *J Biomed Opt.* **13**(2), 024003 (2008).
- [191] Rogowska J, Bryant CM, Brezinski ME. Cartilage thickness measurements from optical coherence tomography. *J. Opt. Soc. Am. A.* **20**, 357-367 (2003).
- [192] Weissman J, Hancewicz T, Kaplan P. Optical coherence tomography of skin for measurement of epidermal thickness by shapelet-based image analysis. *Opt. Express*. **12**, 5760-5769 (2004).
- [193] Gambichler T, Moussa G, Regeniter P, Kasseck C, Hofmann MR, Bechara FG, Sand M, Altmeyer P, Hoffmann K. Validation of optical coherence tomography in vivo using cryostat histology. *Phys. Med. Biol.* **52**, 75-85 (2007).
- [194] Korde VR, Bonnema GT, Xu W, Krishnamurthy C, Ranger-Moore J, Saboda K, Slayton LD, Salasche SJ, Warneke JA, Alberts DS, Barton JK. Using Optical Coherence Tomography to Evaluate Skin Sun Damage and Precancer. *Lasers Surg Med.* **39**(9), 687-695 (2007).
- [195] Lee Y-K, Rhodes WT. Nonlinear image processing by a rotating kernel transformation. *Optics Letters*. **15**(23), 1383-1385 (1990).
- [196] Rogowska J, Brezinski ME. Image processing techniques for noise removal, enhancement and segmentation of cartilage OCT images. *Phys. Med. Biol.* **47**, 641-655 (2002).
- [197] Fernández DC, Salinas HM, Puliafito CA. Automated detection of retinal layer structures on optical coherence tomography images. *Opt. Express*. **13**, 10200-10216 (2005).
- [198] Hori Y, Yasuno Y, Sakai S, Matsumoto M, Sugawara T, Madjarova V, Yamanari M, Makita S, Yasui T, Araki T, Itoh M, Yatagai T. Automatic characterization and segmentation of human skin using three-dimensional optical coherence tomography. *Opt. Express*. **14**, 1862-1877 (2006).
- [199] Danielsson PE. Euclidean distance mapping. *Computer Graphics and Image Processing*. **14**(3), 227-248 (1980).
- [200] Bonnema G, Cardinal K, Williams S, Barton J. An automatic algorithm for detecting stent endothelialization from volumetric optical coherence tomography datasets. *Phys Med Biol.* **53**(12), 3083-3098 (2008).

- [201] Cheng Y, Larin KV. Artificial fingerprint recognition by using optical coherence tomography with autocorrelation analysis. *Appl. Opt.* **45**, 9238-9245 (2006).
- [202] Huang ML, Chen HY. Development and comparison of automated classifiers for glaucoma diagnosis using Stratus optical coherence tomography. *Invest Ophthalmol Vis Sci.* **46**(11) 4121-4129, (2005).
- [203] Jørgenson TM, Tycho A, Mogensen M, Bjerring P, Jemec GBE. Machine-learning classification of non-melanoma skin cancers from image features obtained by optical coherence tomography. *Skin Research and Technology* **14**(3), 364-369 (2008).
- [204] Iftimia NV, Bouma BE, Pitman MB, Goldberg B, Bressner J, Tearney GJ. A portable, low coherence interferometry based instrument for fine needle aspiration biopsy guidance. *Rev. Sci. Instrum.* **76**, 064301 (2005).
- [205] Goldberg BD, Iftimia NV, Bressner JE, Pitman MB, Halpern E, Bouma BE, Tearney GJ. Automated algorithm for differentiation of human breast tissue using low coherence interferometry for fine needle aspiration biopsy guidance. *J Biomed Opt.* **13**(1), 014014 (2008).
- [206] Zysk AM, Boppart SA. Computational methods for analysis of human breast tumor tissue in optical coherence tomography images. *J. Biomed. Opt.* **11**(5), 054015 (2006).
- [207] Ismail SM, Colclough AB, Dinnen JS, Eakins D, Evans DMD, Gradwell E, O'Sullivan JP, Summerell JM, Newcombe RG. Observer variation in histopathological diagnosis and grading of cervical intraepithelial neoplasia. *BMJ.* **298**, 707-710 (1989).
- [208] Targowski P, Rouba B, Wojtkowski M, Kowalczyk A. The application of optical coherence tomography to non-destructive examination of museum objects. *Studies in Conservation.* **49**, 107-114 (2004).
- [209] Liang H, Cid M, Cucu R, Dobre G, Podoleanu A, Pedro J, Saunders D. En-face optical coherence tomography – a novel application of non-invasive imaging to art conservation. *Opt. Express.* **13**, 6133-6144 (2005).
- [210] Wold S, Esbensen K, Geladi P. Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems.* **2**, 37-52 (1987).
- [211] Brereton RG. Experimental Design. And: Pattern Recognition. Chapters 1 & 4 in: *Chemometrics – Data Analysis for the Laboratory and Chemical Plant.* John Wiley & Sons Ltd, ISBN: 0-471-84977-6, pp.15-53 and pp.191-224 (2003).
- [212] Shepherd NA. Barrett's esophagus: its pathology and neoplastic complications. *Esophagus.* **1**(1), 17-29 (2003).
- [213] Fisher L, van Belle G. Categorical Data: Contingency Tables. Chapter 7 in: *Biostatistics – A methodology for the health sciences.* John Wiley & Sons Ltd, ISBN: 0-471-16609-X, pp.256 (1993).
- [214] Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* **33**, 159-174 (1977).
- [215] Cai B, Ronnett BM, Stoler M, Ferenczy A, Kurman RJ, Sadow D, Alvarez F, Pearson J, Sings HL, Barr E, Liaw K-L. Longitudinal Evaluation of Interobserver and Intraobserver Agreement of Cervical Intraepithelial Neoplasia Diagnosis Among an Experienced Panel of Gynecologic Pathologists. *American Journal of Surgical Pathology.* **31**(12), 1854-1860 (2007).
- [216] Horn LC, Reichert A, Oster A, Arndal SF, Trunk MJ, Ridder R, Rassmussen OF, Bjelkenkrantz K, Christiansen P, Eck M, Lorey T, Skovlund VR, Ruediger T, Schneider V, Schmidt D. Immunostaining for p16INK4a Used as a Conjunctive Tool Improves Interobserver Agreement of the Histologic Diagnosis of Cervical Intraepithelial Neoplasia. *American Journal of Surgical Pathology.* **32**(4), 502-512 (2008).
- [217] Kato I, Santamaria M, De Ruize PA, Aristizabal N, Bosche FX, De Sanjosé S, Muñoz N. Inter-observer variation in cytological and histological diagnoses of cervical neoplasia and its epidemiologic implication. *Journal of Clinical Epidemiology.* **48**(9), 1167-1174 (1995).

References

- [218] McCluggage WG, Bharucha H, Caughley LM, Date A, Hamilton PW, Thornton CM, Walsh MY. Interobserver variation in the reporting of cervical colposcopic biopsy specimens: comparison of grading systems. *Journal of Clinical Pathology*. **49**, 833-835 (1996).
- [219] Parker MF, Zahn CM, Vogel KM, Olsen CH, Miyazawa K, O'Connor DM. Discrepancy in the Interpretation of Cervical Histology by Gynecologic Pathologists. *Obstetrics & Gynecology*. **100**, 277-280 (2002).
- [220] Stoler MH, Schiffman M. Interobserver Reproducibility of Cervical Cytologic and Histologic Interpretations - Realistic Estimates From the ASCUS-LSIL Triage Study (ALTS) Group. *Journal of the American Medical Association*. **285**, 1500-1505 (2001).
- [221] Walter SD, Irwig L, Glasziou PP. Meta-Analysis of Diagnostic Tests With Imperfect Reference Standards. *Journal of Clinical Epidemiology*. **52**(10), 943-951 (1999).

Appendix A - Defining “non-invasive” and “non-destructive”

Within literature, there is a tendency to portray OCT as non-invasive. This is a justifiable claim for ophthalmic OCT, where the power of the incident laser is below safety requirements, and the probe is positioned at a certain distance from the organ. However, when intending to apply OCT for endoscopic applications, this claim can be argued. It seems necessary to define “non-invasive”.

A more accurate description of OCT may have to consider discomfort (patient) and logistics (care provider):

Discomfort

The common assumption and rationale for research is that patients would prefer an optical approach to tissue sampling. For colposcopy, studies have shown that this is the case for the majority of participants. This inherently means that there is also a proportion of patients who, given the choice, would opt for a swift biopsy over a scanning procedure. The degree of discomfort a procedure is causing is therefore relevant.

Logistics

Implementing a novel adjunct technology is hugely facilitated if it fits into an existing scheme. This includes the time frame of a procedure. For instance, the target of 15-20 minutes for a colposcopy session – that includes counselling, examination, and administrative work – is difficult to meet already today.

Appendix A.1 - Non-invasive

In terms of the working principle, OCT is non-invasive and non-destructive. The energy which is absorbed by tissue per unit time is low and complying with the safety standards of the American National Standards Institute. After all, OCT interrogates tissue by probing for elastically scattered photons, with no intent of transferring energy into tissue.

Appendix A.2 - Minimally invasive

Minimally invasive is a term coined for surgical procedures which cause minimal trauma (in contrast to open surgery). Instrumentation for such scenarios exists. OCT

fitted into endoscopic devices still requires access to the organ of interest, either via natural orifices or “keyhole” surgery. Those probes are in contact with the tissue surface. This requires a choice of harmless and non-irritating materials.

Appendix A.3 - Invasive

OCT is investigated for replacing biopsies, which are invasive and destructive sampling. A fair approach will have to consider when OCT imaging relies on a lengthy examination or contrast agents, or has the potential to irritate. Contrast agents may significantly enhance image quality, to an extent that their application could become a key step of a scanning protocol. It can be necessary to move organs or body components: For cardiovascular OCT, where blood strongly attenuates the OCT beam, flushing with saline opens a time window of a few seconds for scanning.

Appendix A.4 - Non-destructive, uncomfortable

For endoscopic imaging, it is fair to say that OCT is non-destructive. This allows repeated scans of a whole surface, which seems beneficial for diagnosis. A diagnostic biopsy might not be the main cause of pain (the removed tissue is usually a small skin flake). The respective inconvenience is caused by getting access to the organ site, i.e. the experience of a colposcopic session in its entirety. This, however, may well become a longer procedure if it involves data acquisition by an optical device such as OCT.

Appendix A.5 - Invasiveness in examinations of the cervix

For topical assessment of the human cervix, OCT would be implemented into a colposcopic setting. This procedure already includes application of saline solution and chemicals (acetic acid, iodine) as standard contrast agents, which creates an opportunity to implement such contrast mechanisms for optical measurements as well. However, it might be worthwhile to remind oneself of the potential disadvantages: An OCT examination may require more time and has the potential to cause skin irritation.

Appendix B - Optimal parameters for PC-fed LDA of dense tissue OCT data

Comparison: Pre-processing parameters for OCT Classification

Appendix B.1 - Pre-processing

Appendix B.1.1 - Surface recognition & surface normalising

Several quantiles as cut-off values for the threshold for the binary mask were examined. It was necessary for the algorithm to work on the basis that histogram equalisation did not change results for the thresholding.

Results were calculated for a variety of threshold values (0.85, 0.90, 0.925, 0.95, and 0.975) in order to establish their robustness.

Appendix B.1.2 - Creating surface misalignment

We corrected for single misaligned pixels with a sliding median window of several pixel size (3-17). However, the data was suitable for surface recognition by binary thresholding and delivered good results without smoothing. Hence one could not expect a major impact by refining. Therefore, surface misalignment was introduced in a proportion of the surface vector; e.g. 5 % of a 300 A-scan sub-volume meant corrupting 15 A-scan positions.

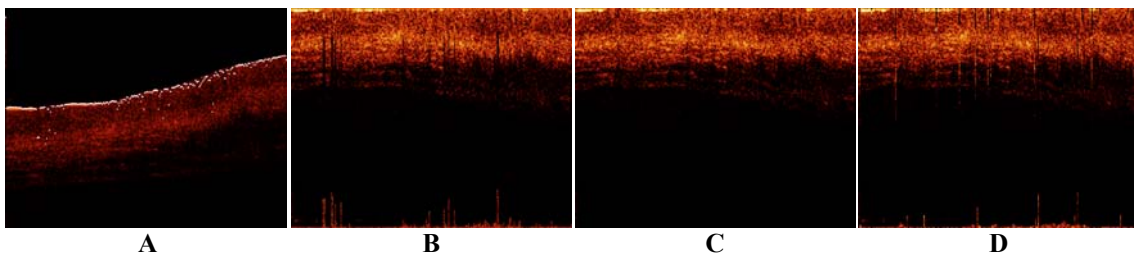


Figure B.1: A) Image with recognised surface (quantile 0.95 for binary mask). B) Applied surface. C) Refined surface (sliding median of 7 pixels). D) 10 % corrupted surface vector.

Figure B.1 shows how the vector corruption creates visual impact. Figure B.1A shows the realistic situation of automated surface recognition – some pixels are not at the surface. When normalising to this surface vector, one receives a result as shown in

Figure B.1B. The amount of misalignment in this situation depends on original image quality and on the parameters for the threshold. Figure B.1C shows how a sliding median filter gets rid of outliers from the surface vector. In Figure B.1D artificial 10 % misalignment have been added to the original surface.

While Figure B.1B shows a realistic situation, Figure B.1D is an artificial – hence reproducible – misalignment, which is used in this study.

Table B.1: Visual impact of histogram equalisation and median smoothing of the recognised surface. The circles shall help spotting some of the differences. The surface recognition was done by using a 0.95 quantile binary threshold filter.

Intensity	Normalised to maximum		Histogram equalisation	
	Surface	Overlay	Surface	Overlay
0				
3				
5				
7				

All deviations were generated with a seeded random number generator. In random number theory, “seeding” is the term for setting initial parameters for a random number generator in order to achieve reproducibility of the set of random numbers. (Such numbers are also called semi-random numbers). This ensured two requirements:

1) Random misalignment values were assigned for random positions of the surface vector; however 2) because of the seeding value, the same random values can be used for comparing different pre-processing techniques on effectively the same image sets. In order not to introduce a statistically relevant feature by misaligning the same amount of A-scans per sub-volume, also the amount of A-scans and the extent of misalignment were subject to randomisation: the parameters were therefore not the amount of deviation, but the maximum range of deviation, or maximum number of misaligned A-scans. Table B.1 shall help to compare these approaches by giving a visual impression.

Appendix B.1.3 - Axial cut-off

Assuming in imaging depth of 1 mm, rows were cut off at different steps between 30 % and 70 %, resulting in the following image dimensions: 300×150 , 300×200 , 300×250 , 300×300 , 300×350 , in addition to the original dimensions of 300×500 pixels.

Appendix B.2 - Results

Appendix B.2.1 - Intensity normalisation vs. histogram equalisation

Images which had undergone histogram equalisation for intensity normalising did not perform as well as intensity normalised scans.

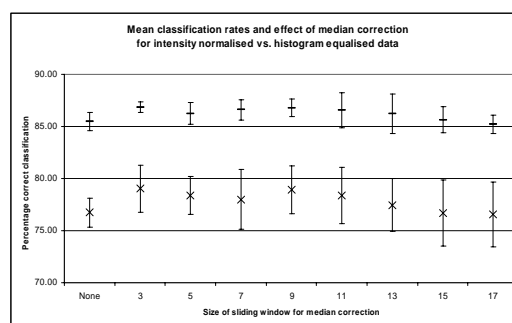


Figure B.2: Comparison of mean classification rates for various threshold levels without artificial misalignments. Upper values (-) from intensity normalised data. Lower values (x) histogram equalised. Intensity normalised data has better classification values.

Figure B.2 shows that for histogram equalised data, a sliding median window of 3-9 pixels can be of benefit. It has to be mentioned that for histogram equalised images,

median correction did not always have a benefit and in selected cases could result in a decreased classification performance.

Appendix B.2.2 - Effect of threshold value

No artificial surface misalignment

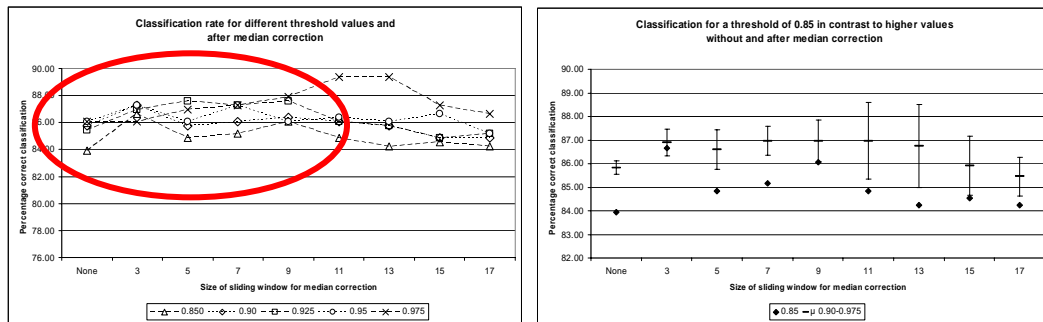


Figure B.3: Classification rates for the test image set for using different threshold levels for creating the binary image which is used for surface recognition. Please note that the dotted lines are for visualisation only and shall not indicate any trend.

Median correction of artificial surface misalignment

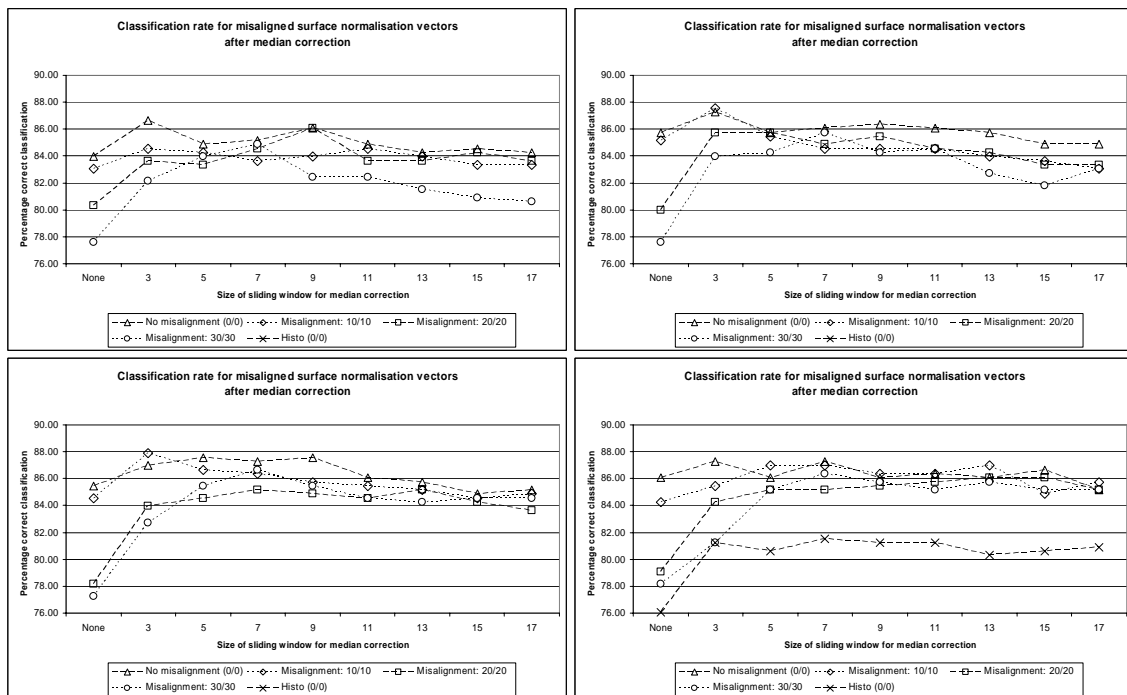


Figure B.4: Classification rates for the test image set for various amounts of misalignment and subsequent correction. Please note that the dotted lines are for visualisation only and shall not indicate any trend or interpolation.

Figure B.3, Figure B.4 and Figure B.5 show that stronger misalignment rates result in worse classification, when being uncorrected for. Using a sliding median filter, even of only 3 values, for correction of outliers does improve results.

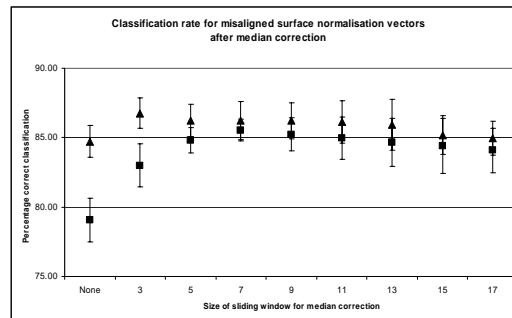


Figure B.5: Effect of median correction on classification rate, mean and standard deviation over different threshold levels. The difference between little (▲) and strong (■) misalignment could show that 1) Larger median windows are required when strong misalignment is expected; 2) Median windows larger than 11 do not further increase the results; 3) These effects are consistent for different threshold levels.

Axial cut-off

The results show the effect of cutting off rows of low SNR on the time required for classification (Figure B.8) and on the quality of classification (Figure B.6, Figure B.7).

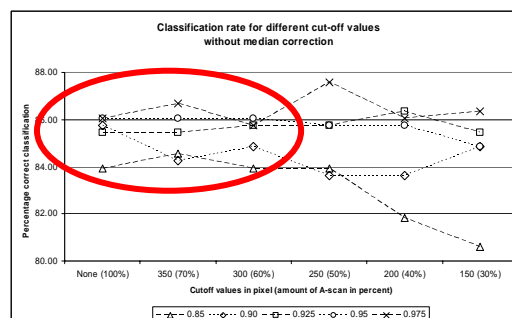


Figure B.6: Classification rate for different cut-off values. There is little variation for cutting off 100-60 % (red circle).

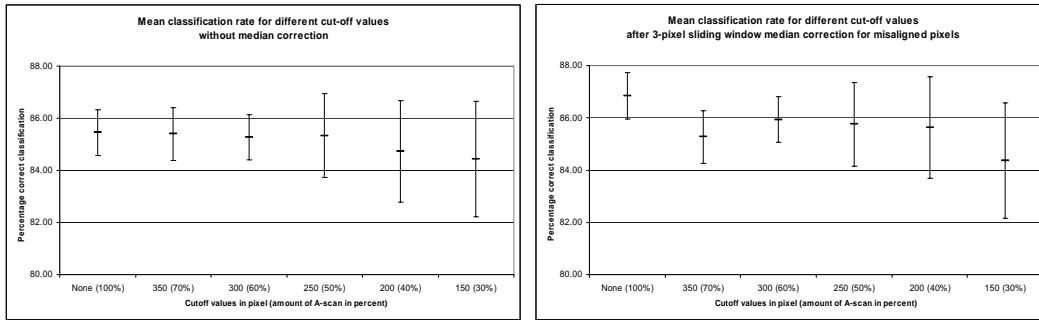


Figure B.7: Classification rates for different cut-off values.

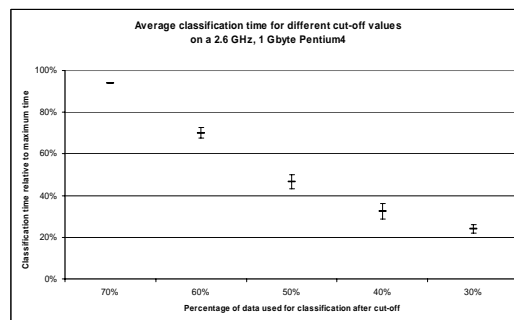


Figure B.8: Average classification time for different cut-off values. Times were measured using Matlab’s tic-toc functions. The small standard deviations (10 measurements) show high confidence.

Appendix B.2.3 - Amount of principal components

After sorting PCs in order of their significance, their impact can be displayed by plotting the classification performance as a function of the amount of PCs fed into the analysis. The results were ambiguous: Figure B.9 shows examples of both cases, where the amount of PCs either does not seem to have a strong impact, or a clear performance drop after a maximum amount of PCs.

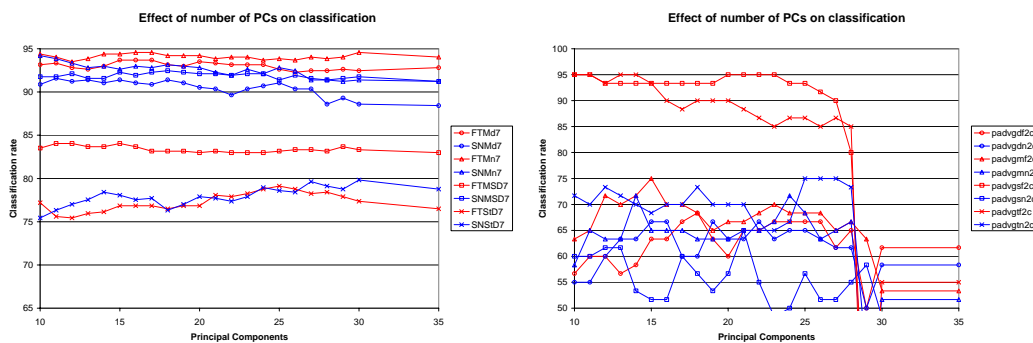


Figure B.9: Left: example of little impact of the amount of PCs. Right: example of significant impact of the amount of PCs and a performance drop with more than 27 PCs.

Appendix B.3 - Discussion and suggestions

Appendix B.3.1 - Threshold for binary mask

Different threshold values give only slightly different results. The lowest threshold value from the investigated range, 0.85, did give less optimal results (still a classification performance of 84 % without median correction), than the other values (ranging from 0.90-0.975). There is no trend which could be used for a recommendation. Within reason, it does not matter which threshold value is chosen for binary thresholding.

Appendix B.3.2 - Median correction of surface misalignment

Sliding windows larger than 9 pixels tend to have no benefit. In only one single case, the median of 11 or 13 was an improvement. When little or no noise is expected, a sliding window of 3 or 5 does significantly improve, whereas for stronger misalignments a median of at least 5 is needed, as displayed by Figure B.5. However it seems not necessary to assume a real situation where more than 10 % of A-scans are misaligned; in such cases an operator would probably retake the scan. As a general rule for unknown data, we recommend a median of 5 values.

Crucially, using a median filter does not worsen results for situations with no misalignment originally.

Appendix B.3.3 - Axial cut-off

The rationale behind cutting off rows was the benefit of a reduced computing time. Figure B.8 shows how computing time decreases for reduced datasets.

For original data (Figure B.7), there is little influence after truncating A-scans, and hence this would be beneficial as it reduces computational effort. However, for median smoothed data, truncating A-scans slightly decreases classification performance. This is unexpected but can be explained if the surface roughness and texture of a sample type (which define the reflection behaviour) plays a role for classification. Hence correcting for these misalignments would reduce a classification criterion, which could explain this result.

Appendix B.3.4 - Amount of PCs

The amount of PCs has an impact on the discrimination performance: PCs are defined by their orthogonality. That means that the variance described by one PC is not covered by another PC. Some PCs will describe low variation and variation of low significance. It is important to keep these two apart: PCs of lower unsupervised variance may still have a strong correlation with the feature that is significant for discrimination. One PC will always contain the remaining variation content, and therefore be – in essence – condensed noise. It is comprehensive that this component will have little impact on discrimination. From the results found here, trends depend however on the respective model and it is therefore difficult to summarise this in a general conclusion. If computationally possible, the optimal amount of PCs should be determined separately for each model.

Appendix B.3.5 - Outliers

For some combinations, e.g. using an 11-pixel sliding window for median correction on a surface vector form a 0.975 binary threshold, we achieved the highest classification values of nearly 90 %. At this point, we are aware of this performance but are not able to use this as a basis for a recommendation, as this is not part of a trend within the results we have obtained.

Appendix C - Ethics files

Appendix C.1 - Study protocol

STUDY PROTOCOL

**Feasibility study for use of
Optical Coherence Tomography (OCT)
and Spectroscopy
for Early Detection of Human Cervical Cancer**

PILOT STUDY

Lead Clinician: Mr Mike Read
Consultant Gynaecologist
Gloucestershire Hospitals NHS Foundation Trust
Gloucester GL1 3NN

Supervisor: Dr Nicholas Stone
Head of Biophotonics Research Group
Cranfield Postgraduate Medical School
Gloucestershire Hospitals NHS Foundation Trust
Gloucester GL1 3NN

Chief Investigator: Dipl.-Ing. Florian Bazant-Hegemark MSc
Biophotonics Research Group
Cranfield Postgraduate Medical School
Gloucestershire Hospitals NHS Foundation Trust
Gloucester GL1 3NN

Study Centre: Gloucestershire Hospitals NHS Foundation Trust

Table of contents

1.	Cervical cancer.....	3
1.1.	General.....	3
1.2.	Architecture of epithelial layers.....	3
1.2.1.	Important characteristics.....	3
1.2.2.	Relevance for cancer.....	3
1.3.	Progression of pathology of cervical cancer.....	3
2.	Present assessment methodology.....	4
2.1.	Deficiencies of present assessment.....	4
3.	Novel Technologies.....	5
3.1.	Optical Coherence Tomography.....	5
3.1.1.	Prospective fields of application.....	5
3.1.2.	Summary: Why OCT?.....	5
3.1.3.	Issues to be assessed.....	5
3.2.	Spectroscopy.....	5
3.2.1.	Prospective fields of application.....	6
3.2.2.	Summary: Why spectroscopy?.....	6
3.2.3.	Issues to be assessed.....	6
4.	This Study: OCT and spectroscopy for early cancer detection.....	7
4.1.	Aims of the study.....	7
4.2.	Methodology.....	7
4.2.1.	Sample types.....	7
4.2.2.	Study design.....	8
4.2.3.	Outline.....	8
4.3.	Patient population.....	9
4.3.1.	Study size & inclusion criteria.....	9
4.3.2.	Exclusion criteria.....	10
4.3.3.	Patient registration.....	10
4.4.	Study endpoints.....	10
5.	Data handling.....	10
5.1.	Data acquisition.....	11
5.1.1.	Data storage.....	11
5.2.	Data processing.....	11
5.2.1.	Anonymisation.....	11
5.2.2.	Quantification.....	11
5.3.	Publishing.....	11
6.	Risks and side effects.....	12
6.1.	Adverse effects.....	12
6.2.	Review points.....	12
7.	References.....	12

1. Cervical cancer

1.1. General

Cancer is a major cause of morbidity in the United Kingdom (UK). Each year more than a quarter of a million people are newly diagnosed with cancer. Overall it is estimated that one in three people will develop some form of cancer during their lifetime.

In the UK, where screening programs are in place, cervical cancer is the third most common gynaecological cancer (after ovarian and endometrial cancer). Worldwide, cervical cancer has the second highest prevalence and third highest mortality of cancers among women.

1.2. Architecture of epithelial layers

Healthy cervical tissue is covered by two types of epithelium:

- Several layers of squamous cells for ectocervical epithelium (on the os)
- A single layer of columnar cells for endocervical epithelium (within the canal)

1.2.1. Important characteristics

- Columnar tissue can change to squamous metaplasia, which is a normal process in an area called the transformation zone (TZ).
- The border between those two types, the squamocolumnar junction (SCJ), is an essential feature for visual examination (colposcopy).

1.2.2. Relevance for cancer

Those tissue types which are inherently able to remodel from columnar to squamous metaplasia are suspected to be susceptible to carcinogenesis: The two relevant types of cervical cancer, squamous cell carcinoma (75%) and adenocarcinoma (25%) develop within the TZ [2].

1.3. Progression of pathology of cervical cancer

Cervical cancer is characterised by a long phase of preclinical disease, meaning that symptoms occur late, when cells already have invaded adjacent tissue - and prognosis is poor. The time it takes to become life threatening is usually several years, but can be a few months in some cases. By the time symptoms are present tumours are usually of a significant size, and it is often too late to achieve a complete cure.

Current methods of detecting early malignancies rely upon screening (surveillance of risk populations) or diagnostic investigations following presentation with suspicious symptoms.

- The primary requirement for successful treatment of malignancy is early detection.
- Treated during early stages, cure rates are estimated to reach almost 100%.

2. Present assessment methodology

Early detection is attempted by screening, before presentation of symptoms. Cervical screening is routinely performed using a spatula to make a 360° sweep of the visible cervix in order to collect the surface cells. It is these cells, which are screened for the abnormalities, which may lead to squamous cell carcinoma of the cervix.

If results are not clearly negative, patients are referred to colposcopy. This is a visual examination by the gynaecologist utilising contrast agents such as acetic acid or iodine solution.

Suspicious lesions are examined through biopsy and subsequent histopathology.

If an abnormality (called dysplasia) is detected, it can be removed by large loop excision of the transformation zone (LEETZ), to prevent development of cancer.

2.1. Deficiencies of present assessment

In short, present assessment is invasive, and not in real time, tissue removal is necessary even in healthy individuals.

The crucial steps, biopsy and histopathology, do have disadvantages:

Procedure	Complication	Consequences
Biopsy	Invasive	
	False Positives	Unnecessary removal
Histopathology	Time-consuming	For healthy individual: Stress For patient: Delayed treatment
	LLETZ	Invasive
Destructive		Complications with previous biopsies Stress Potentially unnecessary

3. Novel Technologies

We wish to assess a new technique, called Optical Coherence Tomography (OCT) and spectroscopy, as a possible alternative to the present methods of detection.

3.1. Optical Coherence Tomography

3.1.1. Prospective fields of application

The ability of OCT to distinguish between healthy and malignant tissue has already been displayed [1]. The scope of this study is to determine how far signs of *early* malignancies can reliably be spotted. This would allow ‘catching’ cancer when disease is certain and treatment justified.

In the case of cervical cancer assessment, the benefits listed by Brezinski & Fujimoto 1999 [1] apply:

- OCT is not hazardous, as conventional biopsy
- OCT reduces false negative rates
- OCT allows guidance of microsurgical procedures

3.1.2. Summary: Why OCT?

Promising OCT facts:

- Resolution (~10 μm) allows resolving tissue architecture at histological level.
- Results are obtained in near real time
- Minimally invasive
- Applicable in vivo

3.1.3. Issues to be assessed

- Consistency with histopathology
- Accurate differentiation of precancerous lesions
- Determination of margins in surgical interventions
- Improvements in present assessment (time, pain caused, accuracy, and cost efficiency)

3.2. Spectroscopy

The interactions of light with matter are dependent on its properties and composition. Even subtle changes between the light before and after the interaction process can be measured. These

changes in light spectra can be recorded as a pattern and a molecular fingerprint unique to a tissue detected.

It can potentially be used to provide diagnostic information about the chemical and morphological structure of tissue *in vivo* and in near real time.

3.2.1. Prospective fields of application

Spectroscopy has in many studies been shown to have the ability to differentiate between benign and other tissues with good correlation to the gold standard histopathology. This diagnostic potential has been demonstrated in tissue from many organs for e.g.: oesophagus, prostate, bladder, breast, and cervix [10], [9], [6]. However an understanding of the biochemical changes relating to the disease progression that may be indicated by spectroscopy is required.

3.2.2. Summary: Why spectroscopy?

- Highly sensitive to minor biochemical changes
- Results are obtained in near real time
- Minimally invasive
- Applicable in vivo

3.2.3. Issues to be assessed

- Consistency with disease progress
- Accurate differentiation of precancerous lesions
- Determination of margins in surgical interventions
- Improvements in present assessment (time, accuracy, and cost efficiency)

4. This Study: OCT and spectroscopy for early cancer detection

4.1. Aims of the study

From what we know, OCT and spectroscopy have high potential for becoming a highly reliable tool for early detection on histological (tissue architecture) level, rather than Pap smear screening on a cytological (cell morphology) level.

Due to the advantages it would indeed seem unethical not to investigate further [15]. The study milestones are:

1. To assess the ability of OCT and spectroscopy to accurately identify different subgroups in the progression to malignancy.
2. To correlate OCT and spectroscopy tissue analysis with the current histological gold standard.
3. To demonstrate the ability of OCT and spectroscopy to differentiate between different pathological groups and show pathways of progression.
4. To develop a predictive model based on OCT and spectroscopy analysis.

4.2. Methodology

In brief: All data will be obtained from specimens removed for diagnostic or therapeutic reasons. The patients will have given informed consent, but from their point of view, there will be no alteration of treatment.

4.2.1. Sample types

For the purpose of this project, specimens or samples fall into the following categories:

Sample type	Expected disease state	Reason for removal
Hysterectomy	Any state, including 'normal'	Specimens from cancer or non-cancerous diseases of the uterine corpus
Small Biopsy	Abnormal	Taken for the tissue diagnosis of suspected ectocervical dysplasia or malignancy
LLETZ	Abnormal	Large Loop Excision of the Transformation Zone, performed for suspected ectocervical dysplasia

In the following, all these approaches will be referred to as "tissue sample" or "tissue sampling".

4.2.2. Study design

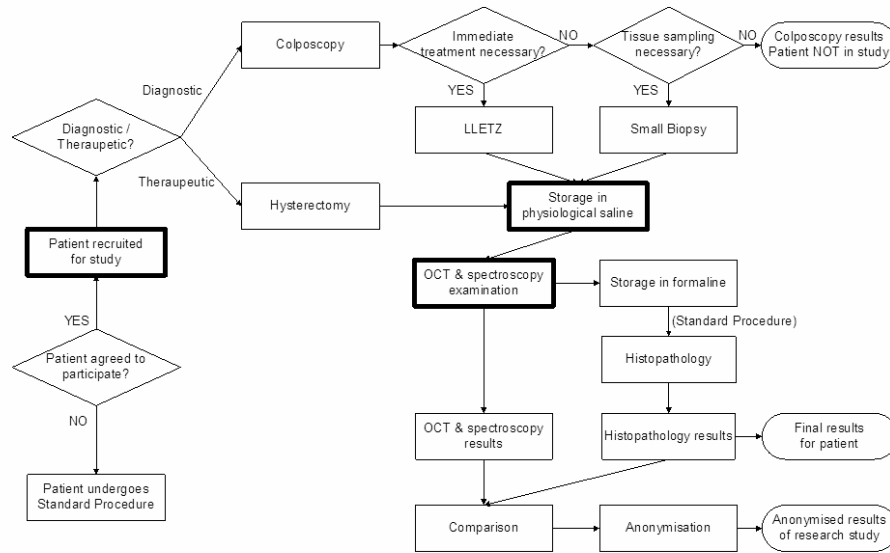


Figure 1: Study design. The boxes with thick border are essential steps of the study. They do not alter or delay the normal assessment.

Normally, tissue specimens are put immediately into formalin in the outpatient clinic or operating theatre, and subsequently transported to the histopathology laboratory.

For this project, the specimens will be placed in buffered saline. The spectroscopy readings will be taken, a process which takes a few minutes. Then the biopsy specimens will be placed in formalin, and be processed in the histopathology laboratory exactly as per normal procedure.

This method of scanning a routine biopsy before it is undergoing histology has been given the all clear by the pathology teams in terms of not having any likely effect on patient diagnosis and care.

The readings will then be correlated with the histopathological data obtained.

4.2.3. Outline

1. Written information about the study will be sent to potentially suitable patients with their Colposcopy clinic appointments along with the clinic information booklet so that they will have time to read the information prior to attending the clinic.
2. The study will be discussed at the clinic attendance and informed written consent obtained by the clinician who is performing the Colposcopy and taking the tissue samples.
3. The tissue sample is labelled with a unique and pseudo-anonymous identification code. This is stored in a password-protected database. Only researchers who are involved in comparison with pathology do have access to the patient number.

4. The sample is stored in saline and transferred to the Biophotonics Lab (GRH).
5. OCT and spectroscopy scanning are undertaken.
6. The sample is placed in formalin and undergoes routine histology.

4.3. Patient population

The population should cover the full spectrum of abnormalities sufficiently for determining a required statistical power in further studies.

4.3.1. Study size & inclusion criteria

The study will allow patients to enter if they meet the following criteria, in order:

1. Within NHS Cervical Screening Programme (NHSCSP)
2. Within Gloucestershire Cervical Cancer Screening Programme
3. Having either
 - a. Pap smear with abnormalities OR
 - b. LLETZ Treatment OR
 - c. Hysterectomy
4. Has given informed written consent

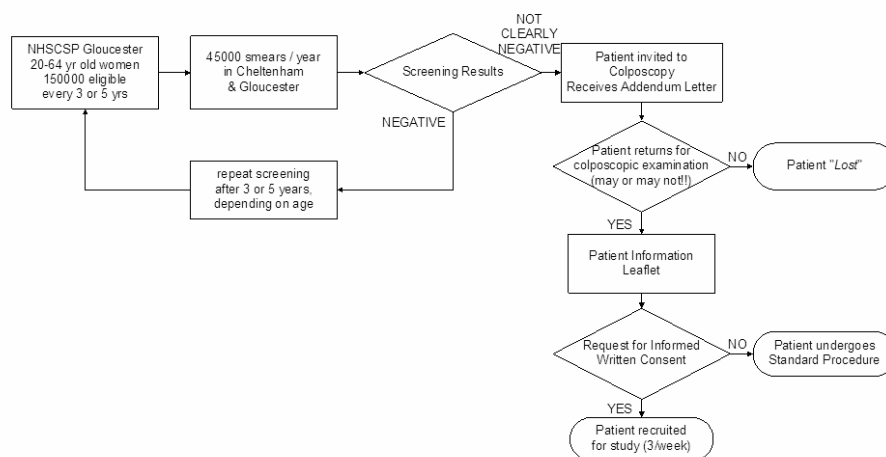


Figure 2: Flow chart showing how the group will be formed.

From previous experience, we estimate approximately 3 patients per week to participate. These specimens would be collected over a period of up to 2 years and subjected to this study. The sample size of the study is calculated to be 200.

4.3.2. Exclusion criteria

Patients are excluded from the study if:

1. They are unable to give a written informed consent.
2. They withdraw their previous consent, which is possible any time without giving reason.
3. Any patient who during the course of examination is found not to require having any tissue removed will not participate in study, even if written consent was given beforehand.

4.3.3. Patient registration

Approximately 2 weeks before colposcopy/hysterectomy, patients will be sent the usual **Letter of Invitation** along with an **Addendum**, which is the first approach to them whether they generally would mind participating in research.

If they opt for getting more information, they will be given the **Patient Information Leaflet** covering a variety of relevant questions. At the time of the procedure the clinician will allow time for discussion and obtain written consent.

The crucial points are listed separately in the **Consent Form** and need to be understood and agreed by the participant before signing the consent. This is checked by the responsible clinician.

There are three copies of the Consent Form, of which one is given to each:

- the patient
- the patient's hospital records
- the research file

Patient Information Leaflet and Consent Form are provided with a corresponding version number and date for referral.

4.4. Study endpoints

The project will end either after two years or when there is sufficient data to show that OCT and spectroscopy are a useful tool in detecting ectocervical dysplasia and/or malignancy, whichever is the sooner.

5. Data handling

The information acquired is of qualitative kind (images).

5.1. Data acquisition

The image acquisition is achieved by using the LabView, a reliable widely used software package for scientific data handling.

5.1.1. Data storage

Data is stored on hard drive and data CDs. These technologies are established and safe, and compatible with emerging data storage systems. This makes them readable for at least the next 5-10 years.

The non-moveable storage items are locked by restricted access (physically locked rooms, password protection). The moveable items are locked in special containers (lockable file cabinets, in physically lockable rooms).

5.2. Data processing

Data processing is performed by Matlab, a software package for optimised mathematical operations. This package includes “tool boxes”, programs for statistical evaluation and image processing.

5.2.1. Anonymisation

The data after comparison with histology no longer needs to be linked to a certain patient. Data is sorted according to disease state, and the case numbers are not corresponding to a person in any way.

Tracking is maintained on a password protected database in case a patient should decide to withdraw participation. However no publication will hold any recognisable feature.

5.2.2. Quantification

- Structural features are marked by image recognition algorithms.
- Backscattering features are quantifiable and will be grouped if agreeing features in histology exist.
- Spectral information will be exploited to provide biochemical signatures of pathological states.

5.3. Publishing

Results are intended to be published in peer reviewed scientific journals.

They should further allow design of subsequent studies and probes for in vivo application. Basically, the more evident differences are, the smaller the next study would be expected to be.

6. Risks and side effects

- The study will in no way influence either the surgical procedure carried out or the subsequent management of the participant.
- No additional procedures will be carried out or withheld.

6.1. Adverse effects

In the case of adverse effects or occurrence of any evidence for redundancy of this study, the recruitment for this study will cease immediately.

6.2. Review points

Review points will be set at the Biophotonics Research group meeting, which usually takes place on Mondays.

Progress reviews are set to be every 6 months (presentation of Mr F Bazant within group and researchers), starting with 4th October 2005 (April 2006, October 2006, April 2007).

7. References

- [1] Brezinski ME, Fujimoto JG: Optical Coherence Tomography: High-Resolution Imaging in Nontransparent Tissue. *IEEE Journal of Selected Topics in Quantum Electronics*, 1999;5(4):1185-1192.
- [2] Euscher B, Morisson C, Nuovo G: *The Cancer Handbook*; Chapter 43 - Female Reproductive System.
- [3] Fercher AF, Drexler W, Hitzenberger CK, Lasser T: Optical Coherence Tomography – principles and applications. *Reports On Progress In Physics*, 2003;66:239-303.
- [4] ISD Online: Information and statistics Divisions, NHS Scotland, 2003. <http://www.show.scot.nhs.uk/isd/cancer/cancer.htm>
- [5] Mahadevan-Jansen A, Richards-Kortum R: Raman Spec. for the Detection of Cancers and Precancers. *J of Biomed. Optics* 1(1), 31-70, Jan 1996.
- [6] Manoharan R, Wang Y, Feld MS: Histochemical analysis of biological tissues using Raman spectroscopy. *Spectrochimica Acta Part a* 52 1996,215-249
- [7] Northern Ireland Cancer Registry. Cancer Incidence and mortality; <http://www.qub.ac.uk/nicr/commoncan.htm>
- [8] Office for National Statistics: registrations. Registration of cancer diagnosed in 2000. England. Series MBI no31. London National Statistics, 2003. <http://www.statistics.gov.uk>
- [9] Shafer-Peltier KE., Haka AS, et al.: Raman microspectroscopic model of human breast tissue: Implications for breast cancer diagnosis in vivo. *J of Raman Spectroscopy*, 33,7,552-563 (2002)
- [10] Stone N, Kendall C et al.: Near-infrared Raman spectroscopy for the classification of epithelial pre-cancers and cancers. *J of Raman spec.*33, 7,564-573(2002).
- [11] Tomlins PH, Wang RK: Theory, developments and applications of OCT. (2005, in press)
- [12] Walker DC, Brown BH, Blackett AD, Tidy J & Smallwood RH: A study of the morphological parameters of cervical squamous epithelium. *Physiol. Meas.*, 2003;24:1-15.
- [13] Welsh Cancer Intelligence and Surveillance unit. Cancer Incidence in Wales. 1992-2001. 2002. <Http://www.wales.nhs.uk/sites/home.cfm? OrgID=242>
- [14] Welzel J: Review article: Optical coherence tomography in dermatology: a review. *Skin Research and Technology* 2001;7:1-9.
- [15] Zuluaga AF, Richards-Kortum R et al.: OCT: A pilot study of a new imaging technique for non-invasive examination of cervical tissue. *American Journal of Obstetrics and Gynecology*, 2005;193;83-8.

Appendix C.2 - Addendum to standard clinic invitation letter



Biophotonics Research Group
Cranfield Postgraduate Medical School
Gloucestershire Royal Hospital
Great Western Road
Gloucester
GL1 3NN

ADDENDUM TO INVITATION LETTER

Dear Patient,

Gloucestershire Cervical Screening sometimes takes part in research, which, we hope, will improve treatments for women with cervical diseases in the future.

Women attending our clinics are occasionally asked if they would like to take part. A full explanation would be given to you during your visit, and you would be able to choose to take part or not.

Taking part would **not** involve any extra visits or test, and most importantly, it would **not** alter the care you receive in any way.

With kind regards
Biophotonics Research

Appendix C.3 - Study participant information letter



Gloucestershire Hospitals



NHS Foundation Trust

Biophotonics Research Group
Cranfield Postgraduate Medical School
Gloucestershire Royal Hospital
Great Western Road
Gloucester
GL1 3NN

PATIENT INFORMATION LEAFLET

Thank you for taking time to read the following information carefully.

You are invited to take part in a research study being carried out at the Gloucestershire Royal Hospital. Before you decide whether you would like to participate it is important for you to understand **why** the research is being done, and what it will involve.

Please do not hesitate to contact us if there is anything unclear or if you would like further information.

Please take time to decide whether or not you wish to take part.

What is the aim of the study?

Development of a new, non-invasive tool for use in the diagnosis of diseases of the cervix.

What is the purpose of the study?

The cervix is the neck of the womb. The cervical screening programme is very successful in detecting early signs of diseases of the cervix, but results still need to be confirmed usually by taking small tissue samples (called a biopsy). It takes some time for the tissue sample to be processed in the laboratory (a process called histology), the results to be obtained and any necessary treatment to be arranged. For some women this can be an anxious time.

We wish to develop a new technique which we hope will improve this situation. This technique uses special imaging technologies, called **optical coherence tomography**, which shines infrared light onto the surface of the cervix, and **spectroscopy**, which analyses light that is shone into the tissue and carries subtle information about its composition.

By analysing the light that is reflected back, we hope to detect painlessly and harmlessly any early diseases, which may be developing, without the need to take small samples of tissue as it has to be done today. It principally works like ultrasound for a pregnant woman.

Before we can apply the device in the human body, we need to show that it doesn't cause harm, that it produces reliable results and that it is of benefit. That's why we first do trials on tissue samples.

For the purposes of this project we will be looking at your sample after it has been removed for therapeutic or diagnostic reasons.

Who is organising the study?

This study is a collaborative project between the Gloucestershire Hospitals NHS Foundation Trust and Cranfield University Postgraduate Medical School.

Why are you asking me?

You are being asked to take part in the study because you are attending the colposcopy clinic for more tests. Just because you are being asked to take part does not necessarily mean that you have a disease of the cervix.

Do I have to take part?

It is entirely up to you to decide whether or not to take part. If you decide to participate you will be given this information sheet to keep and will be asked to sign a consent form.

Can I withdraw if I should change my mind?

Even after opting for participation you are free to withdraw at any time and without giving a reason (your contact person is Mr F Bazant, 08454 22 5717). Your care, treatment or method of assessment will not be altered in any way by your decision.

Will there be additional tissue taken just for this study?

No. You will notice no difference in the way that you are investigated and treated. We will only use material that has been obtained for routine investigation and diagnosis in the normal way. Taking part will in no way alter or worsen your medical care. There is no need for any lifestyle restrictions to participate in this study.

What will happen to my tissue in the study?

Before your tissue is sent to histopathology, as normal, we would like to scan your tissue with the infrared light. This will cause no delay for treatment of your sample in histopathology and will not damage your sample in any way.

What are the possible benefits in taking part?

It is unlikely that there will be any direct benefits to you, but this study may help us to develop a new, non-invasive technique, which will improve our ability to detect diseases of the cervical canal before they become serious. This could prove to be of benefit to future patients.

Are there any disadvantages in taking part in this trial? No.

What are the possible risks in taking part? There are no risks.

Is my doctor being paid for including me in this study? No.

Confidentiality – who will know I am taking part in the study?

All information which is collected about you during the course of this study will be kept strictly confidential. Apart from the medical staff, research staff will need to have access to your medical data. If you consent to taking part in this study you will give permission to the following persons to have access to your pathology results, **only** where it is relevant to your taking part in research:

Dr N Stone, Dr C Kendall, and Mr F Bazant-Hegemark (Biophotonics Research Group).

What will happen to the results of the research study?

The results of this research will be published in scientific journals. If this is the case there will be no way that you can be identified from any results published.

By the end of this study, which is deemed to be November 2007, we expect to have suitable data for publication. It is good practise to inform you of such articles. If you'd be interested in receiving notification, please feel free to contact us (details below).



Gloucestershire Hospitals



NHS Foundation Trust

Ethics Committee Approval:



This study has been approved by Gloucestershire Local Research Ethics Committee.

Consumers for Ethics in Research (CERES) publish a leaflet entitled 'Medical Research and You'. This leaflet gives more information on medical research. A copy can be obtained from CERES, PO Box 1365, London N16 0BW.

Contact for further information:

Thank you for considering to take part in the study. If you have any questions with regard to this trial or above mentioned publications, please feel welcome to contact either Dr Nick Stone, 08454 22 5712, or Mr Florian Bazant, 08454 22 5717, or email to f.bazant@medical-research-centre.com.

Appendix C.4 - Consent form

	<p>Gloucestershire Hospitals</p> <p>NHS Foundation Trust</p>	
<p>CONFIDENTIAL</p> <p>3 copies: - 1 for patient - 1 for patient's hospital records - 1 retained with research file</p>		
<p>Biophotonics Research Group Cranfield Postgraduate Medical School Gloucestershire Royal Hospital Great Western Road Gloucester GL1 3NN</p>		
<p>CONSENT FORM</p>		
<p>Centre Number:</p> <p>Study Number:</p> <p>Patient Information Number for this trial:</p>		
<p>Title of Project: Feasibility study for use of Optical Coherence Tomography (OCT) and Spectroscopy for Early Detection of Human Cervical Cancer</p> <p>Name of Researchers: Mr Florian Bazant-Hegemark, Dr Nick Stone, Dr Catherine Kendall Dr Keith McCarthy, Mr Mike Read.</p>		
<p>Please initial box if you agree:</p> <p>1. I confirm that I have read and understand the Patient Information Leaflet for the above study (Version 2, dated 15 Nov 2005). I discussed it with the clinician and fully understand what is required..... <input type="checkbox"/></p> <p>2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving reason, without my medical care or legal rights being affected. I know that if I wish to withdraw from this study, data collection will cease immediately and I can request that all data collected is destroyed. I can contact Mr F Bazant (08454 22 5717) even after the biopsy has been taken..... <input type="checkbox"/></p> <p>3. I understand that researchers from the Biophotonics Research group may need to access relevant sections of pathology reports linked to biopsies taken..... <input type="checkbox"/></p> <p>4. I agree to take part in the above study, as described in the Patient Information Leaflet (Version 2, dated 15 Nov 2005). <input type="checkbox"/></p>		
Name of patient	Date	Signature
Name of person taking consent (if different from researcher)	Date	Signature
Researcher	Date	Signature
Consent Form	Page 1/1	Version 2.0, 15 Nov 2005

Appendix D - PC-fed LDA for OCT in dermatology

Optical coherence tomography: a potential tool for unsupervised prediction of treatment response for Port-Wine Stains

Florian Bazant-Hegemark^{1,2}, Igor Meglinski^{1,3}, Nanda Kandamany³, Barry Monk³, and
Nicholas Stone^{1,2}

1) Cranfield Health, Cranfield University at Silsoe, Bedfordshire MK45 4DT, UK

2) Biophotonics Research Group, Gloucestershire Royal Hospital, Great Western Road, Gloucester GL1 3NN, UK

3) Laser Treatment Centre, Bedford Hospital NHS Trust, Kempston Road, Bedford MK42 9DJ, UK

Modified[†] from a manuscript that was published in
Photodiagnosis and Photodynamic Therapy, 5(3), 191-197, (2008)

Appendix D.1 - Abstract

Background: Treatment of Port-Wine Stains suffers from the absence of a reliable real-time tool for monitoring a clinical endpoint. Clinical response varies substantially according to blood vessel geometry. Even though optical coherence tomography (OCT) has been identified as a modality to suit this need, it has not been introduced as a standard clinical monitoring tool. It seems that – although OCT is capable of acquiring data in real-time – gigabyte data transfer, processing and communication to a clinician impede the implementation as a clinical tool.

Objectives: We investigate whether an automated algorithm can address this problem.

Methods: Based on our understanding of pulsed dye laser treatment, we present the implementation of an unsupervised, real-time classification algorithm which uses principal components data reduction and linear discriminant analysis. We evaluate the algorithm using 96 synthesised test images and 7 clinical images.

[†] Modifications include: British English spelling; cross-referencing; removing redundancies; and a greater level of detail.

Results: The synthesised images are classified correctly in 99.8 %. The clinical images are classified correctly in 64.3 %.

Conclusions: Principal components-fed linear discriminant analysis is a valuable method to classify clinical images. Larger sampling numbers are required for a better training model. These results justify undertaking a study involving more patients and show that disease can be described as a function of available treatment options.

Keywords: Discriminant Analysis, Optical Coherence Tomography, Port-Wine Stain

Appendix D.2 - Introduction

Appendix D.2.1 - Port-Wine Stains

Capillary haemangiomas (Port-Wine Stains, PWS) are congenital malformations of dermal capillaries. In the UK, PWS are prevalent in 3 of 1,000 inhabitants [222]. Lesions, particularly in facial areas, may have an impact on self esteem of patients. The development of the pulsed dye laser marked a significant advance in treatment; its mechanism is believed to involve coagulation of dermal capillaries via the absorption of laser energy by oxyhaemoglobin. However, blood vessels have varying diameter, wall thickness, flow rates, depth, and carry blood with different oxygenation states. For a successful response to treatment, a lesion needs to be within the range of the laser. The most common PDL lasers operate at yellow (577 or 585 nm) wavelengths [223] which penetrate not more than 1 mm into skin [224]. Blood vessels must be destroyed entirely to prevent regeneration and revascularisation. The capillaries need to carry enough oxyhaemoglobin to absorb energy, vessel thickness <50 µm do not provide sufficient blood volume for a good response [223]. With standard PDL lasers, good response is expected for blood vessels in a depth up to 1 mm and a diameter of 55-150 µm, depending on body site and skin type [225]. Thickness of overlying skin layers and pigmentation can reduce the flux delivered to the capillary. Response to treatment can be improved by altering pulse length, wavelength, energy, vessel dilation (thermal or drug induced), or by cooling surrounding tissue to allow longer pulses [223],[226],[227]. These treatment options carry however an increased risk of complications.

An important practical problem is the lack of objectivity in evaluating clinical response, and the difficulty of predicting which patients will benefit from treatment or further treatment. On average, skin lightens 12 % per treatment [225]. Complete lightening can be achieved in less than 25 % of the lesions, and 70 % of PWS lighten at least 50 % [228]. Purpura, crusting and bleeding are common side effects which can last between 2 and 14 days. After about 6 weeks, a reliable assessment of treatment outcome is possible. Protocols usually require 6 to 10 sessions in 6 to 12 weeks intervals. For a vast majority, “excellent” results in terms of patient satisfaction can be achieved; however this is only established after repeated sessions with no progress in outcome. In up to 20 %, no improvement or darkening of the affected lesion is observed [228]-[230]. For these patients, unsuccessful PDL sessions present therefore a rather ineffective way of determining responsiveness to treatment.

Appendix D.2.2 - Optical coherence tomography

Optical coherence tomography (OCT) is a non-destructive imaging modality. It uses laser light of low coherence to localise reflection events (similar to echoes in ultrasound) within a specimen. For typical infrared sources in dense human tissue, the characteristic penetration depth is around a millimetre and the resolution is around 10 μm [42]. Conventional OCT probes for variations of the refractive index and renders structural information. It is therefore widely investigated for replacing or guiding destructive tissue sampling (biopsies) [89]. Today, fast systems can acquire about a million A-scans per second, in gigabytes of data, which would correspond to about a cm^2 per second at a lateral resolution of 10 μm . For PWS, OCT seems to be able to image the geometry of blood vessels in the relevant depths and magnitudes. Penetration depth, usually a limit in dense tissue, is not so much a problem for this application: A PWS that cannot be displayed with OCT will be unsuitable for PDL of shorter wavelength.

Skin lesions are usually easy to access for existing OCT probes. Studies towards characterising blood vessel formation and flow using OCT have been performed on small animals [231],[232] and in ophthalmic imaging [233], and *in vivo* human skin [234], and even for monitoring treatment response *in vivo* [235]. However, despite

the initial success and potential which has been identified by these studies, it seems that OCT is not yet available as a clinical modality for standard treatment assessment. There is research into using Doppler OCT for monitoring the cessation of blood flow, but to our knowledge that is not widely applied in clinics. We believe that computing accurate velocity models is time intensive: For instance, the angle between the beam and the vessel is integral to the Doppler vector calculations; a velocity vector orientated 90 degrees to the probing beam will show no Doppler signal [165]. Hence the volumetric vessel orientation needs to be established, adding to the overall processing time.

We believe the reason for this to be that data handling and communication, which are common problems identified with OCT: Here, the usually subtle difference between “real-time” and “near real-time” becomes crucial. Although imaging is straightforward, the high scanning speed can only be exploited fully if data can be retrieved at similar speed (Chapter 5). True real-time imaging is therefore difficult to achieve for larger areas. Human operators are not likely to be able to keep up with this data transfer volume. Further, one disadvantage of PDL treatment for PWS is the required experience, and raw OCT images might call for a similar amount of training, artefact knowledge, and data visualisation skills. Vessel mapping and displaying require time and computational effort, let alone programming skills. For a clinical setting however, it is not the astonishing and colourful 3D rendering that is important, but a quick and reliable support for the choice of treatment.

Such can be given by unsupervised classification algorithms, which consequently are an active and emerging research field for OCT data (Chapter 5). Specifically for keratinised layers of skin, or cartilage, procedures to automatically retrieve layer thickness have been investigated [191]-[193]. All these algorithms have certain requirements to image quality and clinical markers, and are usually iterative – which means slow. A quick method is principal components fed Linear Discriminant Analysis (PC-fed LDA). It has been displayed to perform well for classifying test images and porcine tissue (Chapter 6), but requires surface normalised A-scans of the classification groups to be different. As OCT images of PWS seem to show vessel diameter and depth

with clear contrast, such images seem a promising candidate for automated classification.

Appendix D.2.3 - Research questions

In this communication we investigate how well classification algorithms developed for OCT data can be applied to clinical images of PWS, in order to provide a quick and objective aide for diagnosing and guiding treatment.

Appendix D.3 - Methods

Due to the low number of clinical images available to us, we designed a set of synthesised images which was used to test the algorithm with a cross-validation approach using combinatory rules. The clinical images, which were not enough for training an algorithm properly, were tested using leave-one-out cross-validation.

Appendix D.3.1 - Synthesised images

The test images are designed to represent PWS of 3 depths – as approximate treatment categories, e.g. “Treatment option 1”, “Treatment option 2”, or “Untreatable”, as shown in Figure D.1. These 256 bit grey-level images measure 80×512 pixel and have logarithmic intensity gradients representing different signal-to-noise gradients in dense tissue. For each gradient type, there are 8 possibilities of PWS: 3 variations of depth and combinations as shown in Figure D.1. Assuming that 512 pixels are equivalent to 1 mm penetration depth, a disk of 64 pixels would represent a diameter of $125 \mu\text{m}$. In analogy to the standard processing, the intensity values were normalised to fit to unity values (i.e., ranging from 0 to 1).

The 3 depth options are presented by positioning the centre of this disk at pixel 164 ($320 \mu\text{m}$), 210 ($410 \mu\text{m}$), 256 ($500 \mu\text{m}$), and 384 ($750 \mu\text{m}$). Artificial speckle noise was added to the images using the Matlab “imnoise” function. The diameter of the vessels was randomly varied for ± 10 pixel ($\sim 20 \mu\text{m}$), and the depth position was randomly varied for ± 20 pixel ($\sim 40 \mu\text{m}$). For reproducibility, the function for producing random numbers was seeded before creating a set of images using “rand('state',0)”.

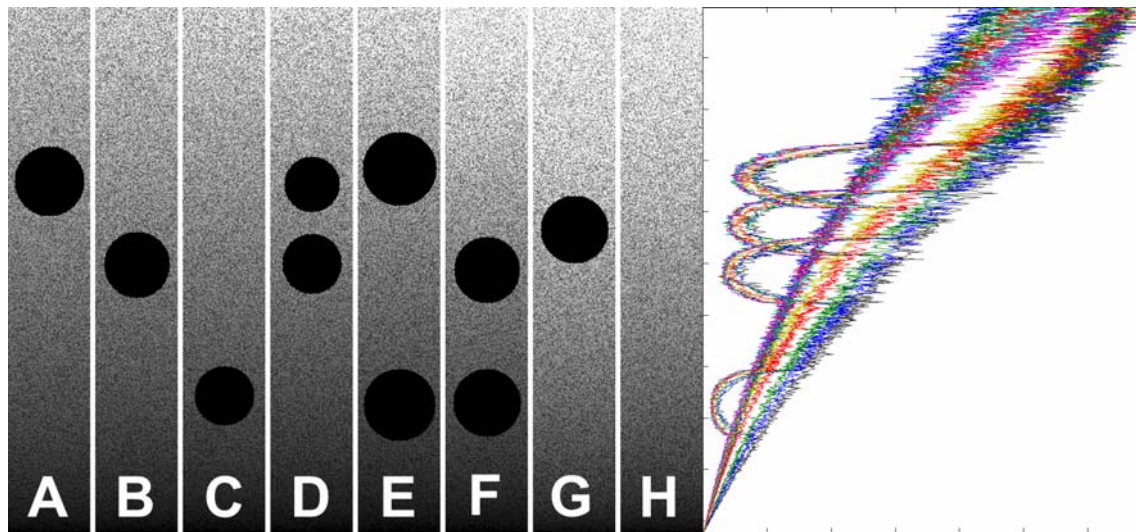


Figure D.1: Synthesised test images for assessing the fundamental possibility of classifying PWS as a function of treatment parameters. Different skin types are represented as different grey scale gradients (8 of 12 gradients are shown). The images here are shown as fed into analysis, with random depth and size variations, and with artificially created speckle noise. The classification model is tested on its ability to classify into “Treatment option 1”, “Treatment option 2”, or “Untreatable” (1, 2, or X). The expected outcome is listed in Table D.1. Right graph: Mean A-scans of these 96 synthesised images.

Eight variations of depth, representing vessels, and 12 background gradients, representing tissue type - i.e. 96 images in total - were used for analysis. Table D.1 shows the design of classification groups for treatment suggestion and expected response. Every model was trained from 6 tissue gradients and cross-validated with the remaining 6 gradients. The mean of the result from 924 permutations (6 from 12), as shown in Figure D.2, was taken as classification result. At this stage, we were satisfied with a result to be either correct or not, and did not calculate specificities or sensitivities.



Figure D.2: Assigning data to test set or prediction set for cross-validation: Representation of 924 permutations for a selection of 6 out of 12, shown in four rows of 231 columns. Each column has 12 pixels (6 grey and 6 black). None of the 924 columns equal each other.

Table D.1: Desired classification performance in designed model. Vessel combinations are shown in Figure D.1.

Classification design	Vessel combination							
	A	B	C	D	E	F	G	H
Vessel depth(s) (μm)	320	500	750	320, 500	320, 750	500, 750	410	-
Treatment option(s)	1	2	none	1&2	1	2	1&2	none
Treatment response	full	full	none	full	partial	partial	full	none

Appendix D.3.2 - Measurements

OCT device

The OCT device used in this study was a time domain bench top system which was similar in design to the system described previously (Chapter 2). The light source is a super luminescent diode (SLD, Superlum) with a central wavelength of 1310 nm and provided an axial resolution of $\sim 15 \mu\text{m}$ in air and a $10 \mu\text{m}$ lateral resolution. The rapid delay line is realised using a double pass mirror and a reference mirror mounted on a galvanometer (Cambridge Instruments). The data acquisition was performed using Labview. The actual acquisition of 333 A-scans per millimetre meant an oversampling of the beam resolution.

Tissue samples

Images were obtained in agreement with the guidelines by the local research ethics committee of the Bedford Hospitals NHS Trust. 7 patients were enrolled in this preliminary study. The measured sites were chosen due to their cosmetic relevance: Forehead, upper and lower cheek, and forearm. The measurements were performed on site by clinicians of the Bedford Hospitals NHS foundation trust. The protocol did not involve any contrast agents or other preparation. At this stage, the samples were confirmed visually by the experience of the clinician. No biopsies were taken for exact matching of the OCT images.

Appendix D.3.3 - Pre-processing

Synthesised images did not require pre-processing, as their surface equalled the top image row. All measured images however, needed to be pre-processed so that the depth pixel of each A-scan would represent the same distance.

The clinical images in this study measure 512×1000 pixels for a width of 3 mm (Figure D.3). Every image depicts the full technical penetration depth of the sample, and there are no excessive reflections. The log of the intensity data (time domain signal) was stored as a matrix. Values were handled in double precision format. The upper part of the image which depicts air was regarded as background noise level. The mean of this plus $7 \times$ the standard deviation of this background signal was used as the intensity threshold. After thresholding, intensity values of one B-scan were normalised to the maximum intensity peak. A binary thresholding was used to determine the sample surface, as shown in Figure D.4A and Figure D.4B. Some artefacts can cause the surface recognition to fail, e.g. minor reflections and hair which is present as small spots above the surface (cf. Figure D.3 & Figure D.4). These were reduced by a sliding median filter. Although the settings for this processing were determined manually, the actual processing was performed fully automated. For the sake of this full automation, we accepted slight misrecognition of the surface as shown in Figure D.4B. After surface normalising data of low signal-to-noise ratio was removed by cropping data below 300 pixels (Figure D.4D).

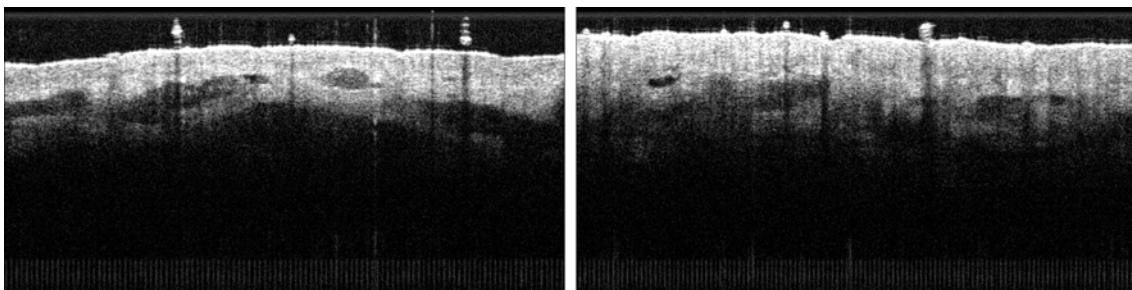


Figure D.3: Example OCT images of PWS, of the lower cheek (left image) and mid cheek (right image).

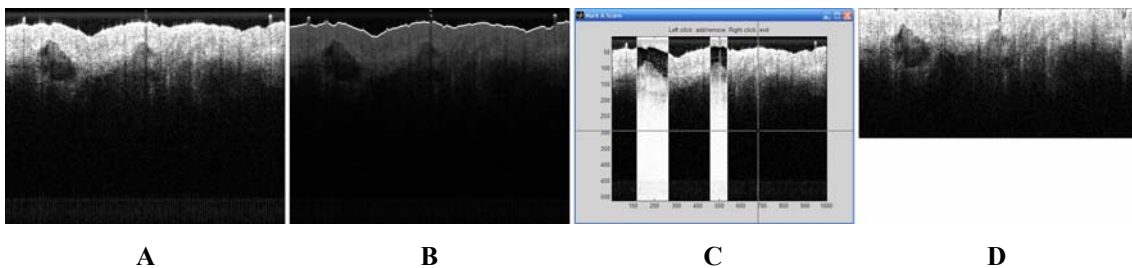


Figure D.4: Pre-processing for clinical images. (A) Original scan. (B) The recognised surface as overlay over the image (The image is darker for visualisation purposes only). (C) Screenshot of the program for manually selecting vessel A-scans. (D) Surface normalised and cropped image.

From the 7 available images, areas (A-scans) containing blood vessels were identified and marked manually using in house written Matlab code as shown in Figure D.4C. Although this is observer biased and based on the assumption that blood vessels show up as darker areas, we believe this to be an acceptable approach for a proof of principle. Of the 7000 A-scans, 2864 were selected as vessel-free, the remaining were classified as containing vessels. No A-scan was excluded from analysis. A representative measure was taken of all A-scans of one sample, thereby providing 14 measures (7 each for vessel and non-vessel images). We want to emphasise that this manual A-scan selection was purely for training purposes, and that no manual step is required for the classification algorithm.

Appendix D.3.4 - Data analysis

Data handling was performed as described previously (Chapter 6), in short: Data handling, pre-processing, and PC-fed LDA were performed using Matlab V6.1.0.450, Release 12.1 (The MathWorks, Inc.), extended by the “PLS_Toolbox” (Eigenvector Research, Inc.) and our in house tools (“Classification_Toolbox”).

Using principal components analysis, the data was reduced to the first 10 principal components (PCs) [210],[211]. This amount of PCs is an arbitrary number and likely to include PCs of low significance to this discrimination. The strongest loadings of these components were determined by characterising the respective F values using ANOVA. An example for the loadings is shown in Figure D.5 and Figure D.6. The PCs were used for linear discriminant analysis, which is inherently able to suppress loadings of lower significance for the model. The cross-validation for the model from synthesised images was performed as explained above using 924 permutations. The clinical images, which provided only 14 measures, were too few for training a strong classification model, and were tested using leave-one-out cross-validation.

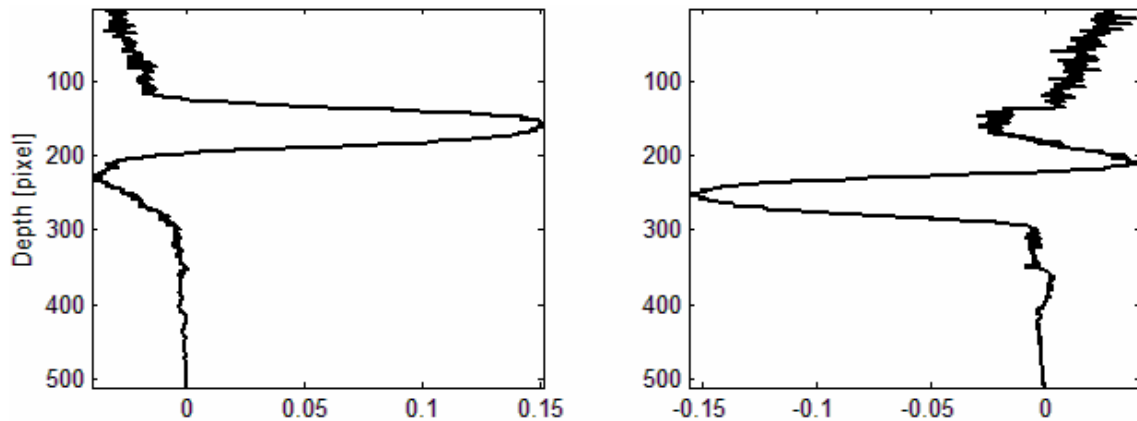


Figure D.5: The 2 strongest loadings for LDA model of the synthesised images. Note that extrema correlate with vessels depths Figure D.1.

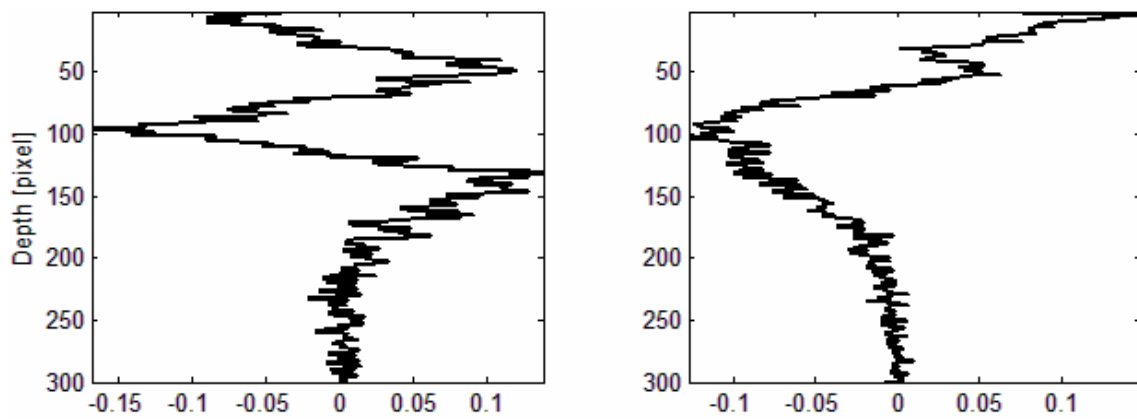


Figure D.6: The 2 strongest loadings for LDA model of tissue data.

Appendix D.4 - Results

The synthesised images were classified correctly in 99.8 %. This is the mean of 924 cross-validations (from all permutations which are possible when taking random 6 out of 12 tissue gradients) for training a classification model.

The training model based on the manual selection of vessel/no vessel, although for a small sample size of clinical images, had a performance of 100 %. The leave-one-sample-out cross-validation was successful in 9 out of 14 images, and therefore had a performance of 64.3 %.

Appendix D.5 - Discussion

Appendix D.5.1 - Procedure and limitations

There are three limitations which can be immediately identified from these specimens: Because of this low sample number, the results of this classification can only be regarded as preliminary. However, within clinical OCT, where larger trials require approval by ethics boards and are timely, it is common to demonstrate the general applicability of an algorithm on only a few samples [198],[204],[206],[235].

Secondly, we assume blood vessels to show up with clear contrast. Believing images from some research groups, this might not always be the case [235]. In such cases, the velocity profiles obtained by Doppler OCT allow a clearer contrast, and therefore also deeper penetration depths. For the presented algorithm, we expect no fundamental difference in distinguishing the clinical marker, be it an absorption maximum or a velocity profile, as long as such a marker can be clearly correlated with the vessel position.

Further, and more crucially, we were not provided with data on treatment outcome. With the present data, we are able to classify normal tissue from blood vessels, measured from PWS lesions which were visibly identified by eye. Exact correlation with histopathology is a major challenge [193]. Our intention is to provide a prognosis on treatment response. Using designed images which we calculated to show clinically relevant geometries, we attempted to display whether such a classification would be possible in principle. Problems in a real clinical situation which we were not able to address in this paper are the impact of skin tone, composition of layers covering the PWS, or dynamic effects like vessel dilation. However, using a large enough population for training a classification model, we believe that results can aid both clinician and patient in the treatment of PWS.

We also have to admit that our assumptions for vessel geometry and depth might not exactly describe the real situation. We believe that for this initial study these assumptions are justified, as laid down in the introduction, and are convinced that - if necessary - the model can be tweaked to more realistic parameters.

Our results were obtained in “near” real-time. The cross-validation time for 924 permutations on a 2.8 GHz Pentium 4, 1 gigabyte RAM PC was 93 seconds (measured

using the Matlab “tic” – “toc” functions), each process of creating a test model and classifying took therefore less than 0.1 seconds. In a clinical situation, the amount of data should be expected to be significantly larger than in our model, however when being used in a clinical setting, the development time for building a classification model and the cross-validation do not contribute to the processing time. As our set of synthesised images consisted of 7680 (80×96) A-scans, we estimate that a square centimetre imaged with $10 \mu\text{m}$ resolution, i.e. 10^6 A-scans, will be classified in less than 13 seconds.

The model trained by clinical images reaches 64 % correct classification. In a clinical setting, we expect that intra- and inter-patient variation will inevitably result in a certain degree of misclassification. The data was collected from a small population and from different sites of the body; this certainly had an impact on the training and testing performance. A model based on a larger pool of data can be expected to perform better. This algorithm is based on taking a representative measure, e.g. the mean, of several A-scans. At a lateral resolution of $10 \mu\text{m}$, a mean of 10000 A-scans would represent an area of 1 mm^2 . This is well below the spot diameter of treatment lasers, which cover a few square millimetres. While we cannot rule out an occasional misclassification, we would expect it to stand out of the surrounding classification. The general impression would be a reliable support for the treatment decision.

Appendix D.5.2 - Summary and conclusions

For a further study design, lesions should be imaged with OCT repeatedly during the intervals between treatment sessions, in order to observe the response after PDL. There are problems which can be imagined because of the dynamics of the healing process, e.g. the impact of crust on image quality. We envisage a protocol which provides a means to co-register scanned sites in order to allow a reliable timeline, and the design of a training model based on the actual treatment outcome of study participants.

A classification algorithm for OCT opens the vast potential not only for characterising a site by imaging, but also to determine whether a site is suitable for enhanced treatment options. Optical clearing agents increase penetration depth [236]. Although the clearing effect depends on the wavelength, effective clearing agents can be expected to impact on treatment parameters like focusing or flow dynamics [237].

OCT would be able to observe, in real-time, when an optimal condition from a treatment preparation has been achieved. We have displayed an algorithm robust enough to cope with slight variations in depth location and vessel diameter. Notably, this algorithm works fully automated and is fast on a current standard computer. As such it could play a role in categorising blood vessel depth according to treatment options, further assessment of this approach – for real-time monitoring of treatment response and prognosis of PWS – is necessary in the scientific and clinical communities.

Appendix D.6 - Acknowledgements

This work is funded by the Pump Prime Fund of Cranfield University and Gloucestershire Hospitals NHS Foundation Trust.

The work presented in this article would not have been possible without the initial support of Prof. Ruikang K Wang, Dr. Shing Cheung, and Dr. Laurie Ritchie.

Dr. Nicholas Stone holds a Career Scientist Fellowship which is funded by the UK National Institute of Health Research.






Appendix D.7 - Additional references

- [222] British Association of Dermatologists. BAD Patient Information Leaflet, [last update 2004 October, cited 2007 December 29]. Available from: <http://www.bad.org.uk/patients/leaflets/birthmarks.asp>, (2004).
- [223] Kelly KM, Choi B, McFarlane S, *et al.* Description and analysis of treatments for port-wine stain birthmarks. *Arch Facial Plast Surg.* **7**, 287-294 (2005).
- [224] Bashkatov AN, Genina EA, Kochubey VI, Tuchin VV. Optical properties of human skin, subcutaneous and mucous tissues in the wavelength range from 400 to 2000 nm. *J Phys D Appl Phys.* **38**, 2543-2555 (2005).
- [225] Smit JM, Bauland CG, Wijnberg DS, Spauwen PHM. Pulsed dye laser treatment, a review of indications and outcome based on published trials. *Br J Plast Surg.* **58**, 981–987 (2005).
- [226] Chiu CH, Chan HHL, Ho WS, *et al.* Prospective study of pulsed dye laser in conjunction with cryogen spray cooling for treatment of port wine stains in Chinese patients. *Dermatol Surg.* **29**: 909-915 (2003).
- [227] Lanigan SW, Taibjee SM. Recent advances in laser treatment of port-wine stains. *Brit J Dermatol.* **151**: 527-533 (2004).
- [228] Jasim ZF, Handley JM. Treatment of pulsed dye laser-resistant port wine stain birthmarks. *J Am Acad Dermatol.* **57**: 677–682 (2007).
- [229] Soueid A, Waters R. Re-emergence of port wine stains following treatment with flashlamp-pumped dye laser 585 nm. *Ann Plas Surg.* **57**: 260-263 (2006).
- [230] Huikeshoven M, Koster PHL, de Borgie CAJM, *et al.* Redarkening of Port-Wine Stains 10 Years after Pulsed-Dye–Laser Treatment. *New Engl J Med.* **356**: 1235-1240 (2007).



- [231] Barton JK, Izatt JA, Kulkarni MD, *et al.* Three-Dimensional Reconstruction of Blood Vessels from in vivo Color Doppler Optical Coherence Tomography Images. *Dermatology*. **198**: 355-361 (1999).
- [232] Wang RKK, Jacques SL, Ma Z, *et al.* Three dimensional optical angiography. *Opt Express*. **15**: 4083-4097 (2007).
- [233] Makita S, Hong Y, Yamanari M, *et al.* Optical coherence angiography. *Opt Express*. **14**: 7821-7840 (2006).
- [234] Zhao Y, Chen Z, Saxer C, *et al.* Doppler standard deviation imaging for clinical monitoring of in vivo human skin blood flow. *Opt Lett*. **25**: 1358-1360 (2000).
- [235] Nelson JS, Kelly KM, Zhao YH, Chen ZP. Imaging blood flow in human port-wine stain in situ and in real time using optical Doppler tomography. *Arch Dermatol*. **137**: 741-744 (2001).
- [236] Proskurin SG, Meglinski IV. Optical coherence tomography imaging depth enhancement by superficial skin optical clearing. *Laser Phys Lett*. **4**: 824-826 (2007).
- [237] Vargas G, Readinger A, Dozier SS, Welch AJ. Morphological Changes in Blood Vessels Produced by Hyperosmotic Agents and Measured by Optical Coherence Tomography. *Photochem Photobiol*, **77**: 541-549 (2003).

Appendix E - Publications

Appendix E.1 - Peer reviewed



	<p>Bazant-Hegemark F, Edey K, Swingler G, Read MD, Stone N: <i>Optical Micrometer Resolution Scanning for Non-invasive Grading of Pre-cancer in the Human Uterine Cervix</i>. Technology in Cancer Research and Treatment, 7(6), 483-496, (2008). http://www.tcrt.org/index.cfm?d=3032&c=4274&p=16972&do=detail</p>
	<p>Bazant-Hegemark F, Stone N: <i>Towards automated classification of clinical optical coherence tomography data of dense tissues</i>. Lasers in Medical Science (in press, 2008). http://dx.doi.org/10.1007/s10103-008-0615-6</p>
	<p>Bazant-Hegemark F, Meglinski I, Kandamany N, Monk B, Stone N: <i>Optical coherence tomography: a potential tool for unsupervised prediction of treatment response for Port-Wine Stains</i>. Photodiagnosis and Photodynamic Therapy, 5(3), 191-197, (2008). http://dx.doi.org/10.1016/j.pdpdt.2008.09.001</p>
	<p>Bazant-Hegemark F, Stone N: <i>Near real-time classification of optical coherence tomography data using principal components fed linear discriminant analysis</i>. Journal of Biomedical Optics, 13(3), 034002, (2008). http://dx.doi.org/10.1117/1.2931079</p>
	<p>Barr H, Kendall C, Bazant-Hegemark F, Stone N: <i>Endoscopic photodynamic therapy for oesophageal disease</i>. PDPDT, 3, 102-105 (2006).</p>


Appendix E.2 - Non peer-reviewed

	<p>Holmes J, Hattersley S, Stone N, Bazant-Hegemark F, Barr H: <i>Multi-channel Fourier domain OCT system with superior lateral resolution for biomedical applications</i>. Proc. SPIE 6847-23, Vol. 6847, (2008). http://dx.doi.org/10.1117/12.761655</p>
	<p>Bazant-Hegemark F, Stone N, McKenzie G, Holmes J: <i>Optical Coherence Tomography Aids Cancer Detection</i>. Biophotonics International, 14(11), 46-47, (2007).</p>

	Bazant-Hegemark F, Stone N, Read MD, McCarthy K, Wang RK: <i>Optical Coherence Tomography (OCT) imaging and computer aided diagnosis of human cervical tissue specimens</i> . Proc. SPIE 6627-65, Vol. 6627, (2007). http://dx.doi.org/10.1117/12.728366
	Bazant-Hegemark F, Stone N, Read MD, McCarthy K, Ritchie LJ, Wang RK: <i>Analysing OCT images of epithelium progressing through precancerous stages</i> , Institute of Physics and Engineering in Medicine (IPEM) Lasers and Optical Science Conf. Proc. (2006).
	Bazant-Hegemark F, Stone N, Read MD, McCarthy K, Ritchie LJ, Wang RK: <i>Optical Coherence Tomography (OCT) Imaging of Human Cervical Tissue Specimens</i> , BMLA Conf. Proc. (2006).
	Barr H, Kendall C, Bazant-Hegemark F, Moayyedi P, Shetty G, Stone N: <i>Endoscopic screening and surveillance for Barrett's esophagus - clinical implications</i> . MedGenMed, 8(2), 88 (2006).
	Bazant-Hegemark F, Stone N, Read MD, Barr H, McCarthy K, Ritchie LJ, Wang RK: <i>Feasibility of using Optical Coherence Tomography (OCT) for Early Detection of Malignancies</i> , BMLA Conf. Proc. - LMS Suppl. B9 (2005).
	Bazant-Hegemark F, Cheung S, Stone N, Read MD, McCarthy K, Wang RK: <i>Feasibility study of application of Optical Coherence Tomography (OCT) for early cancer detection of human cervical cancer</i> , OFMC Proc., 135-141 (2005).

Appendix E.3 - Book chapters / Uploads

	Barr H, Hutchings J, Kendall C, Bazant-Hegemark F, Stone N: <i>Endoscopic Approaches to the treatment of early malignancy and their relationship to clinical outcome</i> . Chapter in: UK Key Advances in Clinical Practice Series, The effective management of Upper Gastrointestinal malignancies, (2007).
	OCT News, webpage, <i>Image of the week</i> , http://octnews.hivefire.com/ , Screenshot from 31 March 2008.

	<p>Bazant-Hegemark F: <i>Rotating Kernel Transformation</i> - algorithm upload, Matlab Central file exchange. http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=14977</p>
---	---

Appendix E.4 - Enclosed copies

Appendix E.4.1 - Chapters in this thesis

- Chapter 4 Technology in Cancer Research and Treatment, 7(6), 483-496, (2008).
<http://www.tcr.org/index.cfm?d=3032&c=4274&p=16972&do=detail>
- Chapter 5 Lasers in Medical Science (2008).
<http://dx.doi.org/10.1007/s10103-008-0615-6>
- Chapter 7 Journal of Biomedical Optics, 13(3), 034002, (2008).
<http://dx.doi.org/10.1117/1.2931079>
- Chapter 8 Proc. SPIE 6627-65, Vol. 6627, (2007).
<http://dx.doi.org/10.1117/12.728366>
- Appendix D Photodiagnosis and Photodynamic Therapy, 5(3), 191-197, (2008).
<http://dx.doi.org/10.1016/j.pdpdt.2008.09.001>

Appendix E.4.2 - Other

- Photodiagnosis and Photodynamic Therapy, 3, 102-105 (2006).
- OCT News [webpage], Image of the week, <http://octnews.hivefire.com/>, Screenshot from 31 March 2008.
- Rotating Kernel Transformation - algorithm upload, Matlab Central file exchange.
<http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=14977>

Review: Optical Micrometer Resolution Scanning for Non-invasive Grading of Precancer in the Human Uterine Cervix

www.tcrt.org

Management of cervical precancer is archetypal for other cancer prevention programmes but has to consider diagnostic and logistic challenges. Numerous optical tools are emerging for non-destructive near real-time early diagnosis of precancerous lesions of the cervix. Non-destructive, real-time imaging modalities have reached pre-commercial status, but high resolution mapping tools are not yet introduced in clinical settings. The NCBI PubMed web page was searched using the keywords 'CIN diagnosis' and the combinations of 'cervix {confocal, optical coherence tomography, ftir, infrared, Raman, vibrational, spectroscopy}'. Suitable titles were identified and their relevant references followed. Challenges in precancer management are discussed. The following tools capable of non-destructive high resolution mapping in a clinical environment were selected: confocal microscopy, optical coherence tomography, IR spectroscopy, and Raman spectroscopy. Findings on the clinical performance of these techniques are put into context in order to assist the reader in judging the likely performance of these methods as diagnostic tools. Rationale for carrying out research under the prospect of the HPV vaccine is given.

Key words: Confocal microscopy; Optical coherence tomography; Infrared spectroscopy; Raman spectroscopy; Epithelial precancer; Cervical cancer.

Introduction

Management of cervical cancer is archetypal: In those countries that have established programs for screening, early detection, and preventive treatment, cancer rates have dropped (1). In the UK, the Cervical Screening Programme of the National Health Service (NHSCSP), which was introduced in 1998, is estimated to save 4,500 lives a year (2). This is possible because squamous cervical carcinoma, the most prevalent form of cervical cancer, is preceded by a long phase of pre-clinical alterations, which are commonly referred to as cervical intraepithelial neoplasia (CIN), and graded into mild (CIN1), moderate (CIN2), or severe (CIN3) (3). The following paragraphs shall provide the common ground of the present understanding of early changes that can serve as surrogate markers for CIN. Further, diagnostic and logistic challenges of current management are discussed in order to support the motivation for improving this situation.

Florian Bazant-Hegemark, MSc^{1,2}
Katharine Edey, MRCOG³
Gordon R. Swingler, FRCOG³
Mike D. Read, MD³
Nicholas Stone, PhD^{1,2,*}

¹Cranfield Health

Cranfield University at Silsoe
Bedfordshire MK45 4DT, UK

²Biophotonics Research Group
Gloucestershire Royal Hospital
Great Western Road
Gloucester GL1 3NN, UK

³Women's Health Directorate
Gloucestershire Royal Hospital
Great Western Road
Gloucester GL1 3NN, UK

*Corresponding Author:
Nicholas Stone, PhD
Email: n.stone@medical-research-centre.com

Abbreviations: CIN, Cervical intraepithelial neoplasia; CM, Confocal microscopy; HPV, Human papilloma virus; IR, Infrared spectroscopy; MMP, Matrix metalloproteases; NHS, National Health Service; NHSCSP, Cervical Screening Programme of the NHS; OCT, Optical coherence tomography; RS, Raman spectroscopy.

Early Changes

What are the characteristics of the different steps in precancer development? Cellular makeup and appearance provide markers for increased proliferation, but specifically early cancer is similar to normal proliferation. The location of proliferating cells within the histological structure is therefore relevant information.

Histological Level: Epithelial cells originate from stem cells that are lying amongst the basal layer of cells on a supporting molecular network of collagen, which also forms the boundary to the underlying connective tissue (basal lamina). The sophisticated regulatory system in the healthy body achieves a smooth gradient of cell maturity from the stem cell level towards the outside throughout the several hundred micrometer thick epithelium. The development from a freshly divided unit to a mature cell can roughly be correlated with the distance from the basal layer. CIN is correlated with subcellular alterations, which result in divergence from this scheme.

Cytological Level: Mature cells have a reduced nucleus-to-cytoplasm ratio and are flat (“squamous”), thereby forming a protective layer. This goes in hand with biochemical changes: glycogen accumulation due to a cease in energy demand, and depletion of metabolites with a relation to cellular activity, e.g., RNA or phosphorylated molecules. Deviations in cell shape and sub cellular architecture (dyskaryosis) and metabolic activity in surface layers may indicate problems with the proliferation at the basal level. However, such changes are not exclusive to precancer, benign conditions (e.g., inflammatory processes) can show similar deviations.

The Lower Boundary: Growth of cells below the basal lamina is believed to be facilitated if this layer has been decomposed by proteins such as the matrix metalloproteases (MMP). Viewed under the microscope, this can be noticed when the usually sharp border between epithelium and underlying tissue is interrupted. Expression levels of MMPs allow correlation with CIN staging (4), and the ability to transgress through tissue boundaries is a feature for determining severity. Cancer prevention aims at catching cells before they gain motility (5).

Diagnostic Challenge

Progression of CIN1 to more severe stages is not definite: Rather, CIN stages serve as a marker for a likelihood of progressing to cancer, and a decreased likelihood of reverting back to normal without intervention. Certain strains of the human papilloma virus (HPV) have been shown to be involved and to be a necessity for cancer progression (6). Large evidence supports that the cancer risk is a consequence

of the body’s inability to develop immunity against HPV, thereby failing to rid itself of the virus. This has led to the development of HPV vaccines.

The key issue in CIN management today remains: Justifying preventive treatment, without missing a lesion and without over treating healthy patients, in as few diagnostic procedures as possible. Diagnosis in modern screening programs is initiated by cytological examination of exfoliated cells (superficial), colposcopy (superficial), and histopathologic assessment of biopsies (sub-surface) (7), as indicated in Table I. Findings of moderate or severe dyskaryosis on a smear or consecutive events of mild abnormalities prompt a referral to colposcopy (8). Current findings support that women should be invited for colposcopy even after only one mild event (9).

Of the adequate tests in 2002/3 in the UK, the follow up results for around 96.6% were negative or borderline for disease. Further surveillance was justified in 2.1% of cases with mild dyskaryosis and in 1.1% of cases with moderate or severe dyskaryosis the protocol recommended preventive treatment (10). While normal and progressed CIN stages are distinguishable, false positives and false negatives are higher for distinguishing CIN1 from CIN2, and are also higher for differential diagnosis of other benign conditions (11-14). Women with mild dyskaryosis who are immediately seen in colposcopy have a higher rate of CIN confirmed by histopathology compared to women who are invited 2 years later, suggesting that regression can happen without intervention (9, 15). However, larger intervals increase the risk of missing a developing high grade lesion; therefore present management has to subject every mild lesion to further diagnostic procedures. Immediate colposcopy referral has a better rate of women attending the follow-up (11). As a consequence, today’s clinical practice has to accept repeated tests, anxiety of patients, and potential over treatment of a lesion (16).

Logistic Challenge

In the UK, of 4.4 million women who are invited for screening, only about 3.6 million actually attend screening (10).

Table I

Management of cervical precancer in the UK (National Health Service Cervical Screening Programme, NHSCSP) involves various levels where the technologies discussed in this manuscript may be of benefit. The HPV vaccination has not yet made the NHSCSP redundant.

Level	Technology	Responsibility
Prevention	Public awareness HPV vaccination	Several bodies, media, NHS NHS, individual
Screening	Cytology (“Pap test”)	Provided via NHSCSP
Early detection	Visual assessment (colposcopy) Histopathology	Part of NHSCSP, Gynaecology
Treatment	Removal of precancerous lesion	Clinician with informed consent of patient
Monitoring	Shorter intervals for suspicious lesion(s), and after treatment	Follow-up as part of NHSCSP

About 10% of these tests are inadequate for evaluation, resulting in over 4 million tests performed per year, at an average cost of £38 (2) per test, or £157 million a year (2) for an estimated 4,500 lives saved due to cervical cancer as primary cause of mortality.

It takes time to communicate these results to the screening participants, thereby causing anxiety and treatment delay: 48% of the women are informed within 4 weeks and 74% within 6 weeks. These figures of 2006/7 are “the highest level recorded” (17) and mean that any woman waiting up to one month for her result still has to consider herself to be above the average. The treatment delay is tackled in the UK, by managing moderate and severe cases with higher priority. This short listing is however based on the subjective impression of the clinician at the time of the sampling, and inherently means a delay for women who do not seem to require urgent treatment. If a biopsy is then required for definitive diagnosis, this result reaches 25% of women within 2 weeks and 38% of the women within 4 weeks. This correlates with a recent study (18) on 600 women in England and Wales, which states that the median time to initiate treatment is 50 days. These figures show a potential for new diagnostic tools to improve the issues of time, cost, and anxiety, which are linked to precancer screening.

Motivation

Therefore, there are attempts to provide a clinical tool that is able to provide information about the tissue in near-real time, where necessary repeated over a longer surveillance time, and in a non-destructive fashion. The motivation is replacing biopsies in order to monitor suspicious lesions and schedule preventive treatment (11). More realistic is the aim of guiding and reducing biopsies, improving the prognostic value of the findings, and better targeting treatment. The paragraphs above may serve to show that this would allow a reduction of healthcare cost and patient distress, as participants can be expected to prefer a “see-and-treat” protocol to a “diagnose-and-defer” approach (19-22).

A further consideration for researching new diagnostic tools is that treatment today is limited to methods that leave the tissue suitable for margin assessment, as this is relevant for prognosis and follow-up: Margin involvement is correlated with higher recurrence rates (23-27), specifically at the endocervical margin (28). This margin assessment by histopathology is based on few sections (sampling a sample) and bears the possibility of misclassifying (29). Destructive treatment may have less pregnancy-related morbidity than excision treatment (30). If tissue can be characterized by a non-destructive tool, this provides destructive treatment options such as laser ablation, cryo surgery, or photodynamic therapy (31-34).

Method: Search Strategy and Selection Criteria

This review aims to cover optical tools for staging of cervical precancer, capable of high resolution imaging or mapping. Following an initial research, the NCBI PubMed data base (35) was queried for ‘CIN diagnosis’ and the results back to the year 1990 were sieved for titles that seemed relevant for non-destructive grading. Following this initial sifting, the database was queried for ‘cervix confocal’, ‘cervix optical coherence tomography’, ‘cervix ftir’, ‘cervix infrared’, ‘cervix Raman’, ‘cervix vibrational’, and ‘cervix spectroscopy’. These high spatial resolution mapping technologies were selected because they are being developed and/or subject to active research for their benefit in staging cervical precancer. This choice shall not be understood as exclusive in the sense that other technologies are not equally legitimate, but should be seen as a self-imposed limitation on what this manuscript aims to discuss. Most recent studies were accepted as such when no forward references were produced from the ISI Web of Knowledge web page (36). This report covers the key publications concerning high resolution cervical precancer mapping since 1975.

The emphasis was on high spatial resolution imaging at the micrometer scale, in contrast to wide-field imaging where some systems at pre-commercial stage are subject to advanced trials. The data on the performance of such systems is increasing rapidly and it seems appropriate to await a separate review of their performance in due course. For this reason, fluorescence spectroscopy as part of imaging technologies was excluded although fluorescence plays a role in generating contrast in some of the technologies which are discussed here. In this context we would like to refer to earlier reviews on optical diagnosis of the cervix (1, 37-39), and on other reviews on molecular markers (40, 41) and methods of detecting HPV (42, 43).

Results: Emerging Technologies for High Resolution Mapping of the Cervix

Confocal Microscopy

A confocal microscope is generally described as a microscope that collects light that is coming from the focal plane only. Mapping within a plane is achieved using galvanometer motors (44), altering the focal plane provides different slices for 3D stacks. Today, lasers allow resolutions of around 1 μm ; thereby, providing images as shown in Figure 1.

CM visualizes *en face* planes of microscopic structures: Studies on smear samples showed that subcellular structures, such as mitotic figures and chromatin patterns, can be resolved and hence contribute to diagnosis (45, 46). The light for detection is created by either reflection or fluorescence,

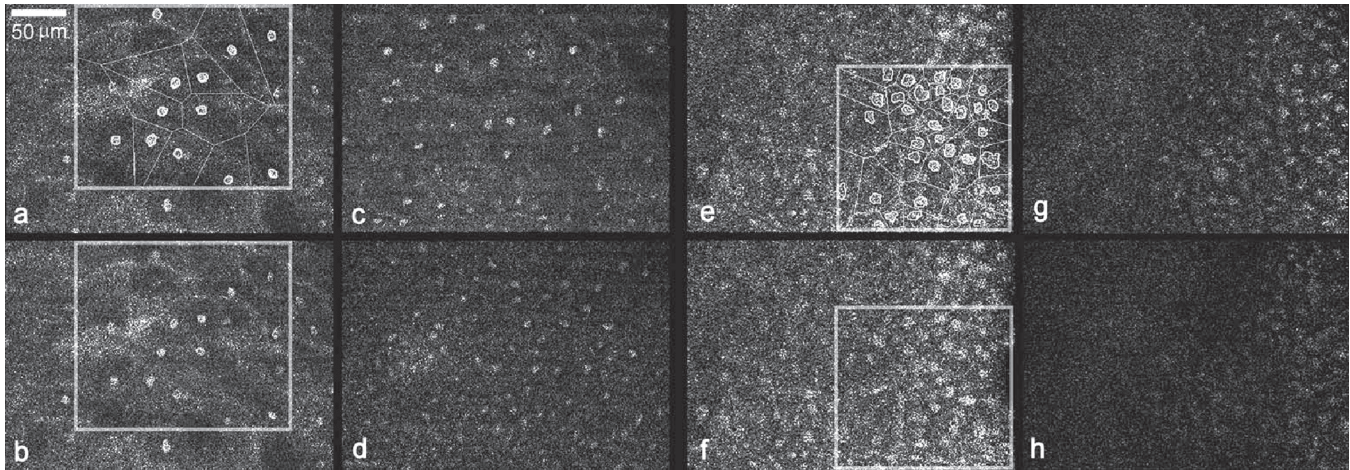


Figure 1: CM *en face* images of biopsy samples, with their nuclear contrast enhanced by acetic acid: Normal tissue at 50 μm below the surface (b), an intermediate layer (c), and close to the basement membrane (d), and of CIN3 tissue at the equivalent distances (f-h). (a) and (e) show an overlay of an automated segmentation algorithm. Reprinted with kind permission (56).

which opens the possibilities of this technology for functional imaging and double selectivity of target volume (47). Fluorescence CM allows mapping fluorophoric metabolites (NADH, FAD) within subcellular structures and distinguishes between cytoplasmic and peripheral fluorescence (48). Fluorescence has been shown to distinguish basal cells from parabasal, intermediate, and superficial cells (49).

As the CM allows description of cellular DNA appearance, risk factors for cancer progression such as morphological changes, nuclear-to-cytoplasm ratio, or irregularity of cell spacing, can be assessed (48, 50). Simple chemicals can serve as contrast agents, *e.g.*, acetic acid, which is a standard agent for colposcopy, improves the contrast for cell nuclei seconds after application (50, 51). However, contrast agents can cause tissue swelling and thus complicate focussing (52).

The field of view for CM images is in the magnitude of a few 100 μm , mapping of larger areas would require sophisticated stitching, or mosaicing, software (53). Although tissue surfaces are not completely smooth it is possible to image cells near the surface and near the basement membrane, the ratio between the two can indicate the CIN stage (54). The penetration depth ('working distance') of modern probes is up to 250 μm , depending on the amount of scatterers in the tissue above. This can be significant in CIN2/3 biopsies where the basal layer cannot always be visualised (55). However, as growth at the basement membrane would be difficult to distinguish from normal, 50 μm may be a sufficient depth for detecting progressed dysplastic features (56).

CM has been used for *ex vivo* sections in order to assess DNA aneuploidy as risk marker and also as measure for treatment outcome. Tissue sections for CM need to be considerably thicker (15-200 μm) than slices for routine histopathologic examination (~ 10 μm) (48, 50, 55). Typical systems are ca-

pable of imaging a 515 \times 660 μm area with 2 μm lateral resolution at 15 frames per second (57), or a 600 μm field of view with 1.3 μm lateral resolution at 12 frames per second (58), or at 6.5 frames/second (56).

Although the technical realization of an *in vivo* system has been demonstrated (58-60), there are also recent *ex vivo* studies on cervical tissue (56). For volumetric imaging stepper motors have to ensure accurate positioning. Studies that abstain from tissue embedding have to cope with handling problems, when trying to image the full epithelial thickness: In a recent study only 9 out of 52 biopsies could be evaluated successfully, 43 were too fragile or had insufficient contrast (56). However, handling of cervical biopsies can be tricky in general: The small size and fragile nature of biopsies can result in about 50% of samples not being available for CM imaging (55). Whether volumetric imaging is advantageous, has to be assessed: There are significant differences of nuclear volume, surface area, and spherical shape factor between normal and abnormal cells; these are more pronounced on volumetric images (61). Another study found that for classification, standardizing to epithelial thickness for sub-surface 2D slices gives better classification than from volumetric images, such a protocol seems, therefore, better for *in vivo* imaging (56). Volumetric data can, however, aide automated segmentation: When auto-detection fails (62), adjacent planes can be used for correction (56). The potential of volumetric CM seems to be best realized where sample stability is not an issue. Recently, the possibility of adapting conventional wide field microscopes to reach both a contrast and a section thickness comparable to CM has been displayed on a cervical cancer cell line (63), at lower cost than CM.

The majority of studies on the benefit of CM for cervical precancer assessment concerns establishing clinical features or system development. To the best of the authors' knowl-

edge, there is only one published study subjecting findings from CM to standard histopathologic assessment. It included 19 participants and allowed detecting the distinction of high grade dysplasia from normal tissue with 91% specificity and 100% sensitivity based on the nucleus-to-cytoplasm ratio (64). Studies for the relevant classification of CIN1 vs. CIN2/3 are anticipated.

Optical Coherence Tomography

Optical coherence tomography (OCT) is based on low coherence interferometry. It is often described as optical ultrasound for its working principle: The position of a backscattering event is determined to within a few micrometers of accuracy, along an axial scan line ("A-scan"). A crucial paper that employed a scanning setup to create 2-dimensional sections from adjacent A-scans was published in 1991 (65). OCT typically uses near infrared light sources in the range of 800-1500 nm (66, 67), at powers that are harmless for skin. The contrast for structural features is determined by changes in the refractive index. Contrast agents for OCT are an active research topic (68) but have not impacted on studies of the cervix. In contrast to the other modalities, conventional OCT does not acquire *en face* images [although such scanning regimens exist, (69)]. The depth (axial) resolution is decoupled from the lateral resolution and both are a crucial specification. Typical resolutions are in the magnitude of 5-15 μm laterally and axially in air; resolutions around and beyond 1 μm are referred to as 'ultra-high'. The penetration depth in dense tissue is around 1 mm. Strong absorbers in human tissue, *e.g.*, melanin (visible range) and water (infrared range) limit the choice of wavelength. Light sources for OCT have to fulfil stringent criteria in respect to partial coherence, in order to allow interferometric detection. The quality of a beam's 'pulse' directly determines the axial resolution. Ideal sources are not available for every wavelength and large efforts of OCT research concern laser sources. However, the advantage of the interferometric detection setup is that it is practically not confounded by ambient light.

Research systems today can measure 10^6 to 10^7 A-scans per second, which allows several volumes per second and so-called 4D OCT (70). Cervical studies with such systems are, however, still awaiting their publication. Issues with OCT are the low contrast and the amount of data generated, which stretch modern data storage systems and may create a need for computer-assisted data evaluation (71).

The non-invasive, real-time OCT was a diagnostic revolution for ophthalmology. First images displaying endoscopic OCT assessment of cervical malignancy have been made in 1997 and 1998 (72, 73). The resolution of OCT is smaller than large epithelial cells. In practice, sub-cellular structures or even single cells are almost impossible to resolve because of

speckle noise and weak contrast. However, there is a promise of being able to rapidly image cervical pathologies through features that indicate some form of structural alteration at cellular level. A first *ex vivo* study on tissue of the female genital tract included 32 cervical specimens (74). This study concluded that an intact basement membrane can be depicted as a feature of healthy epithelium, and included 21 healthy and 11 CIN3 or carcinoma specimens. Escobar *et al.* (75, 76) performed an *in vivo* study on 50 patients, and included the diagnostically more relevant samples, from CIN1, CIN2/3, and inflamed sites. It allowed concluding that the thickness of squamous epithelium is varying for healthy samples, and confirmed healthy epithelium as a marker for an intact basement membrane. This membrane itself is too thin for depiction on OCT images. Instead, the visible marker is a distinct boundary between squamous epithelium and underlying stroma. Sensitivity and specificity of the OCT measurement, when blinded to other results, was 56% and 59%, respectively (77). The only improvement seems to be in specificity when OCT is used as adjunct to visual inspection using acetic acid (VIA), where it improves the low specificity of VIA alone from 34% to 61% for VIA + OCT. Sensitivity then is 76% for VIA alone, and 53% for VIA + OCT. Zuluaga *et al.* (78) reported a correlation in the backscattering intensity of normal and abnormal tissue, but do not give detail on sensitivity or specificity of this marker. Findings of a study with data from 212 patients resulted in the suggestion to use OCT for confirming positive colposcopic examinations prior to treatment, in low resource settings (77). Most studies used 2D images, such as shown in Figure 2, whereas state-of-the-art OCT is volumetric. Imaging of the gastrointestinal tract has shown that 3 dimensional OCT gives a much better representation of the microscopic organisation of tissue structures (79). Lee *et al.* have undertaken a study on 18 subjects and imaged not only structural OCT volumes, but the polarization state. It seems that the degree of circular polarization decays faster in progressed CIN, and quantifying this decay allowed discrimination of low grade and high grade lesions with up to 94.7% sensitivity and 71.2% specificity (80). However, further structural markers that can be definitely correlated with CIN progression would be desirable. In ophthalmology, OCT development could be targeted to successfully resolve the retinal layers. In contrast, dividing the cervical epithelium into basal, parabasal, intermediate, and superficial layers is a stratification based on relative cellular appearance; it is not based on a boundary of the organism as such. In absence of clear structural markers, individual A-scans can serve as a measure for optical properties. Algorithms to extract such features have been investigated (81, 82). These algorithms require robust markers and need to be evaluated for larger sample sizes of clinically relevant groups (83).

In vivo probe design for volumetric data faces the challenge of guiding the laser beam accurately in 2-dimensions. For

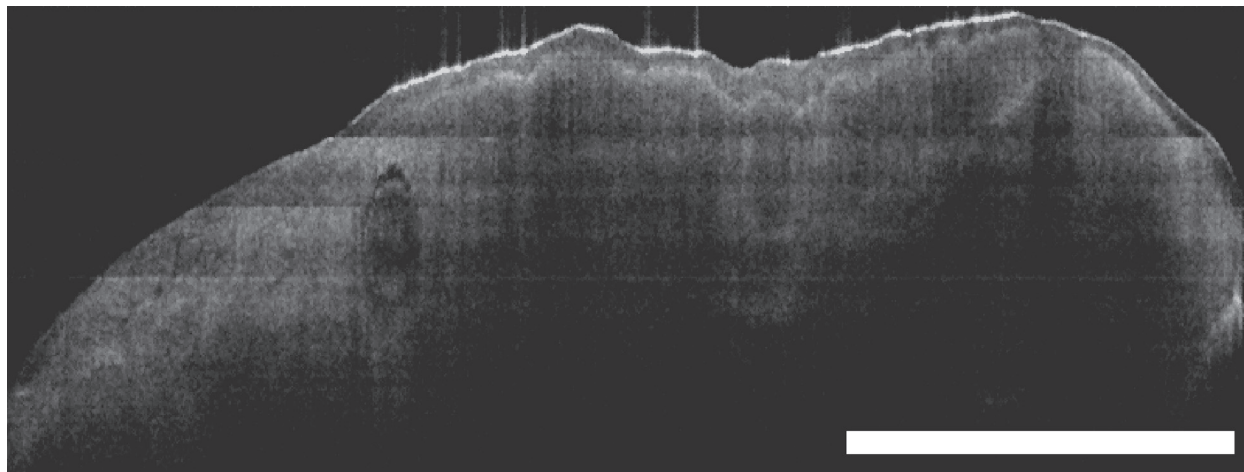


Figure 2: OCT cross section of cervical epithelium and stroma, obtained *ex vivo* in real time, using a 7.5 mm resolution frequency domain system operating at 1300 nm (Michelson Diagnostics EX1301 OCT Microscope, Michelson Diagnostics Ltd.). The left side shows glandular epithelium of the endocervical canal, and a Nabothian cyst is visible. Towards the right the squamous layer of the ectocervix is evident. The bar measures 3 mm.

cardiovascular settings and imaging of the gastro-intestinal tract, helical scanning protocols exist to acquire tubular volumes from within (84). This could provide useful for imaging of the endocervical canal. Such instrumentation has the focal point at a distance relative to the helix axis, and has to assume a tissue surface with this cylindrical shape. Where this is not the case, *e.g.*, at folds, data will, therefore, be of reduced quality or missing. This might be relevant when glandular carcinoma develops exactly in such folds.

Another approach, in addition to polarization sensitive OCT, to overcome the limited contrast for biochemical processes may be to combine OCT with fluorescence, and devices for such measurements have been built (85). The diameter of the endoscopic device is 2 mm. Technical challenges arise because OCT requires a focused beam, whereas fluorescence is designed to provide diffuse wide field illumination. *In vivo* measurements on a mouse colon take 45 seconds for a 6 mm stretch. The two modalities are not completely overlapping, a 1300 nm OCT beam allows a $\sim 15 \mu\text{m}$ resolution and penetration depth of about 1 mm, and the 325 nm excitation source for fluorescence penetrates approximately 200 μm into tissue. Housing both modalities in one system allows a reliable co-registration of measurements. This addresses a major issue for modalities at this level of resolution. Measurements for cervical tissue have not been published, but fluorescence-guided OCT has been attempted on other organs.

IR Spectroscopy

Mid-infrared spectroscopy (IR) measures wavelength-dependent absorption properties by probing vibrating energy levels of functional groups and intermolecular interactions. This is possible in molecules with an IR active dipole. Fourier transform infrared (FTIR) spectroscopy is the standard IR method today but still often referred to as 'FT'IR to empha-

size the method of spectral data collection using FT interferometry rather than dispersion gratings. For medical applications, near- or mid-infrared light (between 0.8 μm and 2 μm) sources give the best contrast. In contrast to analysis of pure substances, spectra from biological tissue represent a mixture of all *different* molecules in the interrogated micro-volume.

The typical mode of operation is absorption spectroscopy, and in order to obtain representative signal-to-noise ratios, transmission measurements have to be taken from thin tissue sections or thin cell layers. Initial publications are therefore on exfoliated cells (86). A review with an emphasis on different IR methodologies for cervical cancer screening has been published recently (87). Wong *et al.* measured spectra from 156 samples, categorising into normal, dysplastic, and malignant (86). They found that cancerous progression in mature epithelial cells, but also normal proliferation, go in hand with glycogen reduction, increased hydrogen bonding from phosphodiester groups, and a loss of hydrogen bonding in alcoholic groups of amino acids. Such features are intuitive as they represent energy demand and mitotic activity, and have been reconfirmed by subsequent studies on cytological smears (88-90), cancer cell lines (91, 92), and tissue (93-96).

While peak ratios of glycogen/phosphate (decrease) and RNA/DNA (increase) seem to correlate well with proliferation of cervical epithelial cells, detecting a malignant candidate is more difficult. An experienced spectroscopist might be able to interpret the shape of a spectrum or peak ratios, but this is not ideal for the diagnosis of precancer. Chemometrics can help giving a better estimation for classification (87, 88, 95, 97, 98). The investigation of IR for precancer assessment coincides with the evolution of personal computers that facilitate such analysis. Chemometrics provide objective and reproducible discrimination methods that are based on complex calculations, practically remove the need for staff

training, and allow a statistical estimation in border areas. An appropriately trained classification model may be able to separate inflammation, endocervical cells, and bacterial infection (88, 99). As with conventional cytology, debris, red blood cells, polymorph cells, endothelial cells, and other components can confound the analysis (91, 100).

Some of these findings deserve a closer look for their diagnostic value in addition to precancer staging. The type of cells and their maturity state is of great importance, and the spectral signatures of immature, benign cell types can obscure abnormal cells (99, 101, 102). Sampling plays an important role and issues concern areas such as the origin of a sample (101), the homogeneity within the sampling volume (97), cell handling (88), or cleaning protocols (103). Spectral signatures of cells change after two hours of being suspended in saline solution at room temperature (100). Cells would therefore have to be stored at a defined low temperature, which for a large scale program would pose a logistical challenge. Essentially, a protocol needs to ensure that it is valid to assume that: a proliferative signature within the sampling volume can be considered as a diagnostic factor for CIN. Still, in a study on 301 cytological samples, IR classified better than cytology (98.6% sensitivity and 98.8% specificity vs. 86.6% and 90.5%) (88). The main potential for improvement derived from a study on 2000 cells of 22 women seems to be that cytologically or morphologically normal epithelial cells from abnormal smears have an abnormal biochemical signature (93). In search for sampling types, lavage specimens have been investigated (104). The issues with sample collection and the performance were not convincing enough to replace smears; yet this sampling method was suggested as an alternative where smears are not possible.

Although IR is still being investigated for assisting cytology (105), it is not restricted to smear evaluation. In combination with a stage, tissue mapping is possible. IR systems can be designed for imaging in the sense that separate pixels are measured at the same time, rather than scanned sequentially. In order to be able to assume a homogeneous sampling volume, this is done on a few micrometer thin tissue sections. Typical systems take about 20 minutes for

500 × 500 μm area (IR microscopy or microspectroscopy). Preparation of thin tissue sections, which presents a main cause for delay in today's routine, requires an effort quite comparable to conventional histopathology: obtaining the biopsy, embedding or freezing, slicing, transferring to slide, fixing and staining (90, 94). Spectral differences in tissue are less pronounced than in exfoliated cells, but still allow good classification (102). On the other hand, mapping overrides some of the difficulties of correctly locating a cell in its histological environment. It allows characterization of the different cell and tissue types in cervical epithelium, such as cervical stroma, epithelium, inflammation, blood vessels, mucus, and abnormalities thereof (90, 94, 105, 107). The key advantage of IR spectroscopy is the objective evaluation to assist the histopathologist. The false positive rates can be as low as 9.04% for CIN1 and 0.01% for both CIN2 and CIN3 (108), or 0.04% (100) and can predict cancer in biopsies with 95% accuracy (95).

Wood *et al.* have taken this approach into the third dimension. By stacking maps from several consecutive tissue sections they created a 3-dimensional stack (Figure 3). The whole process from taking the measurement to multi-slice analysis and volume rendering was reported to last less than 1 hour for a four-section model covering 1,400 × 1,400 × 16 μm (109). Automation for cytological or histopathologic examination faces veritable challenges and requires a standardized sampling protocol. However, as a research tool it already has provided valuable evidence for understanding the biochemistry of precancer; and it can be imagined as an adjunct for *ex vivo* staging.

Raman Spectroscopy

Raman spectroscopy (RS) is the term for probing for inelastic scattered light, the Raman-scattering. RS differs from IR: IR spectroscopy probes for absorption caused by changes in dipole moment. RS characterises molecules according to their change in polarisability, and aims to measure deviations from elastic scatter. Despite of these differences and although Raman spectra exhibit more complexity, evaluation and presentation of Raman data are similar to IR spectroscopy.

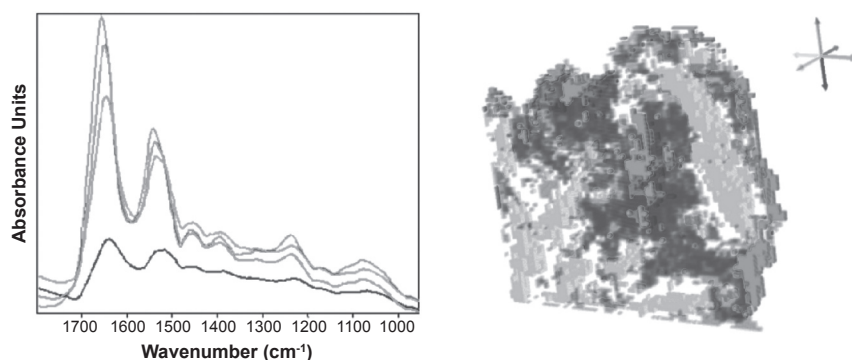


Figure 3: Left side: Mean IR spectra of different tissue types. These can serve as a reference for unknown spectra (multivariate models are more complex). Right side: Volumetric presentation of consecutive IR-maps. The shade of each voxel represents the classification likelihood for a type of cervical tissue. The volume measures 1,400 × 1,400 × 16 μm. Modified and reprinted with kind permission (109).

copy (Figure 3). RS and IR together are commonly referred to as vibrational spectroscopy (110).

Raman signals from tissue are weak and occur, for applied wavelengths and powers, statistically only one time for 10^{10} incident photons. Other effects easily cover a Raman signal, e.g., fluorescence is about a million stronger (111). A significant disturbance is fluorescence from biological tissues and from fiber optics (e.g., silica), and filter designs for contribution of optics are a crucial component of *in vivo* Raman systems (112). The near infrared range around 780 nm is typically used for fiber optics probes; but also *ex vivo* studies using UV light (254 nm) exist (113). Fluorescence, which occurs at the same wavelength range as the Raman signal, cannot be filtered out using normal filter design, and may require baseline subtraction (114). Cross talk is also the reason why 2-dimensional Raman data are mapped, rather than imaged. Lack of photons makes RS a much slower technology than IR, and novel approaches that remove delays in stage movement and CCD readout have a huge impact on scanning speed (115). Systems employing time resolved technology to 'time-gate' out the fluorescence signal require sophisticated hardware that can occupy a large room (116). Although near IR light would penetrate deeper into tissue, it is estimated that a good Raman signal-to-noise ratio from cervical tissue is only obtained from epithelial layers (117).

Raman active molecules and chemical groups allow characterizing the presence, quantity, and composition of tissue (118). Suggestions for using this technology to distinguish abnormal and normal tissue of the gynaecologic tract were mentioned in 1992 (119). Results from a system using a 254 nm light source on cell suspensions, allow distinction of normal from malignant cells due to peak ratios from different nucleic acid contents (113). Recent studies use near IR lasers. *Ex vivo* tissue samples (36 biopsies of 18 patients) could be classified as precancerous or normal with 82% sensitivity and 92% specificity (120).

Results can be obtained from IR-inactive molecules and can be quite complementary to IR spectroscopy. However, the methods for statistical analysis of spectra are similar (121). Utzinger *et al.* used peak ratios of cervical tissue spectra to distinguish normal from dysplastic areas (122). Diagnostically significant molecules were demonstrated to be DNA, phospholipids, and collagen (122). Analysis of characteristic peaks and fingerprint regions agrees with the biochemical picture provided by IR spectroscopy (86): basal cells show characteristic bands of nucleic acids, epithelial cells show high levels of glycogen in normal cells. Decreased levels of glycogen and increased levels of DNA and amino acids in epithelial layers are a feature of cancerous cells (117, 121, 122). A study on three cell lines (keratinocytes representing normal, HPV16 – infected keratinocytes, and

a cervical cancer cell line), was able to distinguish pairs of these lines with 70-100% sensitivity and 70-90% specificity, *i.e.*, also those cells regarded as normal but expressing a viral protein from normal ones (125). The larger variation occurred for live cells; the model designed for fixed cells had a sensitivity of 93-100%.

In contrast to IR spectroscopy, *in vivo* probe systems that can probe for the biochemical changes exist and have been used for RS (126). *In vivo* results differ from *ex vivo* measurements, but are still able to discriminate healthy or squamous dysplastic tissue (123). Databases from *ex vivo* tissue spectra can, however, not easily be used for designing *in vivo* discrimination models. Although spectra from formalin fixed discriminate well (127), they are not comparable with fresh tissue: In fact, RS is so sensitive to embedding resins that it proved useful for characterizing the efficacy of dewaxing agents (128). Hence, Raman pilot studies should be performed in an environment the actual system is intended for: fresh tissue for *in vivo* applications (129), or on fixed cells for cytological measurements (125). Keeping this consideration in mind, RS can be used for both *in vivo* and *ex vivo* measurements of biochemical constituents. RS requires reduced ambient light to minimize spectral contributions. Taking measurements in a dark room or using a covering cloth can address this issue (122). Recent *in vivo* probes require few seconds for measuring a (point-) spectrum and can be considered real-time (111, 124, 130). *In vivo* mapping seems unlikely due to these time requirements, and exactly locating the measurement on the cervix is necessary. Even although motion artifacts could potentially be addressed by image registration methods (131), the length of the procedure would still have to be kept on an acceptable level for patients.

The available tissue studies restrict themselves to classification between normal and high-grade or cancerous samples: Krishna *et al.* (117) investigated 150 biopsies, distinguishing normal vs. malignant tissue nearly 100% correct. Lyng *et al.* (123) subjected histological samples from 40 patients to classification. Raman signals from CIN samples lie between normal and cancerous tissue, and can be distinguished with nearly 100% sensitivity and specificity. CIN was not further staged and the clinically relevant distinction between CIN1 and CIN2/3 was not addressed in this paper. The latest *in vivo* study on data from 66 patients showed that RS has 89% sensitivity and 81% specificity to distinguish high-grade dysplasia from benign tissue, higher than the colposcopy for these cases (87% and 72%, respectively) (124). Here the classification performance against low-grade dysplasia was attempted; because of the small sample size (6 spectra) in that group the model cannot be considered strong. The introduction of new statistical algorithms for classifying spectral variations of disease groups might help improving the classification performance and is an active research field (132).

RS has further attractions: The technical feasibility of depth probing Raman signals has been displayed repeatedly (116, 133). The possibility of probing biochemical processes opens another field: While the other modalities give a diagnostic snapshot at the time of the measurement, RS could potentially provide a prognosis (134).

Discussion

Emerging technologies are leading to the possibility of a real-time, minimally invasive assessment of early cervical disease. Table II summarizes how the discussed technologies provide information on structure which correlates with histology and cytology, or on molecular composition that can be explained with knowledge of biochemical processes.

The differences in cervical epithelial cells that are surrogates for determining the malignant potential of a lesion are sound: size and shape of cells will alter the way light is reflected and scattered, different concentrations of cell components will result in different amounts of absorbance and transmission, and different amounts of fluorescent biochemical constituents would show semi-quantitative emission patterns.

Let us remind the reader that colposcopy itself is an assessment of the visual appearance of different types of epithelium, in order to obtain a combination of optical diagnostic markers: A typical examination refers to interaction of white light as seen under a microscope, and makes use of contrast agents for proteins (acetic acid) and glycogen (iodine). The visual impression contributes to the prognosis of CIN progression. However, the correct interpretation is

based on the expert opinion of the clinician. In addition, aceto-whitening and iodine staining are not 100% correlated with the cancer risk.

Yet although markers for an increased proliferative behavior are intuitive, it remains difficult to directly conclude an increased risk of cancer progression from them. Better understanding even a rather defined entity as cervical cancer is challenging, and even more would be developing an appropriate *a priori* model for cervical cancer. For a complex system like human tissue, data evaluation requires multivariate statistics, and data reduction techniques. Inevitably, a reference in the form of databases has to be established in large clinical trials.

Combining reflectance spectroscopy and fluorescence imaging has shown that non-destructive testing of epithelium is possible and such systems are close to commercialization. Based on available trial data, these systems seem to increase the sensitivity for CIN2/3 on the expense of a reduced specificity for CIN1. However, their advantage lies in the objective judgement, which does not require the immediate expertise of a trained colposcopist. In this context, when regarding a spectral pattern as suitable for discrimination, it is not the raw spectra, but the classification model based on them, which is used for discrimination. Therefore, the choice of statistical classification might well influence the performance.

Although systems utilising reflectance spectroscopy and fluorescence are near commercialization, they are not 100% definite in diagnosis (135, 136). They are consequently positioned as adjunct for colposcopy, in order to guide biopsies. For better classifying disease and underlying mecha-

Table II
Overview of modalities discussed in this review.

Modality	Confocal microscopy	Optical coherence tomography	Infrared spectroscopy	Raman spectroscopy
Probing for	Subcellular structure	Structure-related histological features	Molecular composition, biochemical patterns	Molecular composition, biochemical patterns
Contrast	Backscattering and refraction, fluorescent metabolites	Change of refractive index	Peak intensities: absorption as function of molecule bonds	Raman signals as function of molecule flexibility
Axial Resolution*	around 1 μm	~1-15 μm	-	-
Lateral Resolution*	around 1 μm	~1-15 μm , depending on beam guiding instrumentation	Application/ stage dependent	Application/ stage dependent
Penetration depth*	~300 μm	~1 mm	-	-
Field of view	Microscopic area	~5 mm – few cm	Microscopic area	In vivo: point measure Microscopic area
Main imaging mode	En face, volumes from stacks	B-scan, volumes from adjacent B-scans	Point measurements and imaging	Point measurements, scan modes for mapping
Investigated precancer/proliferation features	Sub-cellular structure, nuclear/cytoplasmic ratio	Histological structure, intact basement membrane	Characteristics of biochemical processes: Glycogen, methylation, protein phosphorylation, DNA hydration, glycogen/phosphate ratio, RNA/DNA ratio	Peak intensities and ratios of collagen, elastin, nucleic acids, phospholipids, glucose-1-phosphate

* for typically applied wavelengths in cervical tissue.

nisms, it might however be beneficial to directly measure biochemical constituents and their relative concentration, or subcellular and sub-surface structure. High resolution modalities, which are being actively developed, might be able to address these issues. For high resolution *in vivo* measurements, it is important to realize how large the required area would be: Studies show that the radial location of a lesion on the cervix is random (137), and evaluation of the 2002 screening protocol of the International Federation for Cervical Pathology and Colposcopy showed that classifying the extent of a lesion (minor or major) and the localization of a lesion in respect to the TZ – which is changing for each patient – are important for correct classification at the stage of colposcopy (13). Therefore, tools aiming to replace colposcopy need to image the entire cervix surface.

Justification for Continuing Research on Cervical Cancer Prevention in the Prospect of HPV Vaccination

With the availability of FDA-approved vaccines, which are aimed against the Human Papilloma Virus (HPV), there is now a prospect of having a new means of tackling this disease: A vaccination against cancer. These vaccines can be expected to reduce cervical cancer linked to HPV strains 16/18. These are the two most common genotypes of HPV in precancerous lesions and account for an estimated 70-75% of all cases (6, 138). However, the aetiology of cervical cancer cannot be reduced to infection by these two strains. The vaccine is prophylactic, which requires an individual not to have been exposed to these HPV strains prior to vaccination (139).

It remains to be observed whether the currently available HPV vaccines will eradicate cervical cancer. The “ESGO statement” of the European Society of Gynaecological Oncology reiterates that carefully conducted research and long time monitoring will remain necessary before HPV vaccination schemes may lead to a change in present cervical cancer prevention programmes (140). Neither HPV nor CIN are the actual disease that causes worry. It is the higher risk of progression to cancer that is of concern. Disease management, including screening, superficial assessment, and subsurface examinations as part of precancer prevention will remain necessary.

Acknowledgements

This work is funded by the Pump Prime Fund of Cranfield University, the Gloucestershire Hospitals NHS Foundation Trust, and receives funding from the Technology Strategy Board. Dr. Nicholas Stone holds a Career Scientist Fellowship, which is funded by the UK National Institute of Health Research.

References

1. Follen, M., Levenback, C. F., Iyer, R. B., Grigsby, P. W., Boss, E. A., Delpassand, E. S., Fornage, B. D., Fishman, E. K. Imaging in Cervical Cancer. American Cancer Society. *Cancer* 98, 2028-2038 (2003).
2. Department of Health. Cervical Screening – The Facts. NHS Cancer Screening Programmes. Bulletin: 272855/Cervical Screening, (2006).
3. Shafi, M. I., Welton, K. Colposcopy and cervical intraepithelial neoplasia. *Obstetrics, Gynaecology & Reproductive Medicine* 17, 173-180 (2007).
4. Gaiotto, M. A., Focchi, J., Ribalta, J. L., Stávale, J. N., Baracat, E. C., Lima, G. R., Guerreiro da Silva, I. D. Comparative study of MMP-2 (matrix metalloproteinase 2) immune expression in normal uterine cervix, intraepithelial neoplasias, and squamous cells cervical carcinoma. *Am J Obstet Gynecol* 190, 1278-1282 (2004).
5. Sahai, E. Mechanisms of cancer cell invasion. *Current Opinions in Genetics and Development* 15, 87-96 (2005).
6. Burd, E. M. Human Papillomavirus and Cervical Cancer. *Clinical Microbiology Reviews* 16, 1-17 (2003).
7. Behtash, N., Mehrdad, N. Cervical cancer: screening and prevention. *Asian Pac J Cancer Prev* 7, 683-686 (2006).
8. Lindeque, B. G. Management of cervical premalignant lesions. *Best Pract Res Clin Obstet Gynaecol* 19, 545-561 (2005).
9. Kyrgiou, M., Koliopoulos, G., Martin-Hirsch, P., Kehoe, S., Flannelly, G., Mitrou, S., Arbyn, M., Prendiville, W., Paraskevaidis, E. Management of minor cytological abnormalities: a systematic review and meta-analysis of the literature. *Cancer Treat Rev* 33, 514-520 (2007).
10. Cancer Research UK. Cancer Stats – Cervical Cancer – Cervical Screening Programme Results <http://info.cancerresearchuk.org/cancerstats/types/cervix/screening/cervicalscreeningresults/?a=5441> (accessed October 2007)
11. Elit, L. M. Pitfalls in the diagnosis of cervical intraepithelial neoplasia. *J Low Genit Tract Dis* 8, 181-187 (2004).
12. Carreon, J. D., Sherman, M. E., Guillén, D., Solomon, D., Herrero, R., Jerónimo, J., Wacholder, S., Rodríguez, A. C., Morales, J., Hutchinson, M., Burk, R. D., Schiffman, M. CIN2 is a much less reproducible and less valid diagnosis than CIN3: results from a histological review of population-based cervical samples. *Int J Gynecol Pathol* 26, 441-446 (2007).
13. Hammes, L. S., Naud, P., Passos, E. P., Matos, J., Brouwers, K., Rivoire, W., Syrjänen, K. J. Value of the International Federation for Cervical Pathology and Colposcopy (IFCPC) Terminology in predicting cervical disease. *J Low Genit Tract Dis* 11, 158-165 (2007).
14. Owens, C. L., Moats, D. R., Burroughs, F. H., Gustafson, K. S. Low-grade squamous intraepithelial lesion, cannot exclude high-grade squamous intraepithelial lesion is a distinct cytologic category: histologic outcomes and HPV prevalence. *Am J Clin Pathol* 128, 398-403 (2007).
15. Iatrakis, G., Kourounis, G., Georgopoulos, N., Karachotzitis, J. Treatment delay and pathology results in women with low-grade squamous intraepithelial lesions. A preliminary study. *Eur J Gynaecol Oncol* 25, 376-378 (2004).
16. Diakomanolis, E., Haidopoulos, D., Chatzipapas, I., Rodolakis, A., Stefanidis, K., Markaki, S. Negative cone biopsies. A reappraisal. *J Reprod Med* 48, 617-621 (2003). Erratum in: *J Reprod Med* 48, 833 (2003).
17. The Information Centre, Part of the Government Statistical Service. Cervical screening statistics 2006/7. ISBN: 978 1 84636 166 1. Bulletin: 2007/14/HSCIC (2007).
18. Martin-Hirsch, P., Rash, B., Martin, A., Standaert, B. Management of women with abnormal cervical cytology: treatment patterns and associated costs in England and Wales. *BJOG* 114, 408-415 (2007).
19. Basen-Engquist, K., Shinn, E. H., Warneke, C., de Moor, C., Le, T., Richards-Kortum, R., and Follen, M. Patient distress and satisfaction with optical spectroscopy in cervical dysplasia detection. *Am J Obstet Gynecol* 189, 1136-1142 (2003).

20. Orbell, S., Hagger, M., Brown, V., Tidy, J. Appraisal theory and emotional sequelae of first visit to colposcopy following an abnormal cervical screening result. *Br J Health Psychol* 9, 533-555 (2004).
21. Balasubramani, L., Orbell, S., Hagger, M., Brown, V., Tidy, J. Do women with high-grade cervical intraepithelial neoplasia prefer a see and treat option in colposcopy? *BJOG* 114, 39-45 (2007).
22. Quek, S. C., Mould, T., Canfell, K., Singer, A., Skladnev, V., Coppleston, M. The Polarprobe - Emerging Technology for Cervical Cancer Screening. *Ann Acad Med Singapore* 27, 717-721 (1998).
23. Livasy, C. A., Maygarden, S. J., Rajaratnam, C. T., Novotny, D. B. Predictors of recurrent dysplasia after a cervical loop electrocautery excision procedure for CIN-3: a study of margin, endocervical gland, and quadrant involvement. *Mod Pathol* 12, 233-238 (1999).
24. Rojat-Habib, M. C., Cravello, L., Bretelle, F., Roger, V., Liprandi, A., de Burtel, I., d'Ercole, C., Pellissier, J. F., Blanc, B. Value of endocervical margin examination of conization specimens. Prospective study conducted on 150 patients [Article in french]. *Gynecol Obstet Fertil* 28, 518-525 (2000).
25. Tyler, L. N., Andrews, N., Parrish, R. S., Hazlett, L. J., Korourian, S., Significance of margin and extent of dysplasia in loop electrosurgery excision procedure biopsies performed for high-grade squamous intraepithelial lesion in predicting persistent disease. *Arch Pathol Lab Med* 131, 622-624 (2007).
26. Cardoza-Favarato, G., Fadare, O. High-grade squamous intraepithelial lesion (CIN 2 and 3) excised with negative margins by loop electrosurgical excision procedure: the significance of CIN 1 at the margins of excision. *Hum Pathol* 38, 781-786 (2007).
27. Maluf, P. J., Adad, S. J., Murta, E. F. Outcome after conization for cervical intraepithelial neoplasia grade III: relation with surgical margins, extension to the crypts and mitoses. *Tumori* 90, 473-477 (2004).
28. Johnson, N., Khalili, M., Hirschowitz, L., Ralli, F., Porter, R. Predicting residual disease after excision of cervical dysplasia. *BJOG* 110, 952-955 (2003).
29. Fadare, O., Rodriguez, R. Squamous dysplasia of the uterine cervix: tissue sampling-related diagnostic considerations in 600 consecutive biopsies. *Int J Gynecol Pathol* 26, 469-474 (2007).
30. Kyrgiou, M., Tsoumpou, I., Vrekoussis, T., Martin-Hirsch, P., Arbyn, M., Prendiville, W., Mitrou, S., Koliopoulos, G., Dalkalitsis, N., Stamatopoulos, P., Paraskevaidis, E. The up-to-date evidence on colposcopy practice and treatment of cervical intraepithelial neoplasia: the Cochrane colposcopy & cervical cytopathology collaborative group (C5 group) approach. *Cancer Treat Rev* 32, 516-523 (2006).
31. Tabor, A., Berget, A. Cold-knife and laser conization for cervical intraepithelial neoplasia. *Obstet Gynecol* 76, 633-635 (1990).
32. Bodner, K., Bodner-Adler, B., Wierrani, F., Kubin, A., Szölds-Szölts, J., Spängler, B., Grünberger, W. Cold-knife conization versus photodynamic therapy with topical 5-aminolevulinic acid (5-ALA) in cervical intraepithelial neoplasia (CIN) II with associated human papillomavirus infection: a comparison of preliminary results. *Anticancer Res* 23, 1785-1788 (2003).
33. Yamaguchi, S., Tsuda, H., Takemori, M., Nakata, S., Nishimura, S., Kawamura, N., Hanioka, K., Inoue, T., Nishimura, R. Photodynamic therapy for cervical intraepithelial neoplasia. *Oncology* 69, 110-116 (2005).
34. Andikyan, V., Kronschnabl, M., Hillemanns, M., Wang, X., Stepp, H., Hillemanns, P. Fluorescence diagnosis with 5-ALA thermogel of cervical intraepithelial neoplasia [Article in German]. *Gynakol Geburtshilfliche Rundsch* 44, 31-37 (2004).
35. U.S. National Library of Medicine and the National Institutes of Health. PubMed Service. <http://www.ncbi.nlm.nih.gov/sites/entrez>
36. Thompson Reuters. ISI Web of Knowledge. <http://isiwebofknowledge.com/>
37. Drezek, R. A., Richards-Kortum, R., Brewer, M. A., Feld, M. S., Pitris, C., Ferenczy, A., Faupel, M. L., Follen, M. Optical imaging of the cervix. American Cancer Society. *Cancer* 98, 2015-2027 (2003).
38. Sokolov, K., Follen, M., Richards-Kortum, R. Optical spectroscopy for detection of neoplasia. *Current Opinion in Chemical Biology* 6, 651-658 (2002).
39. Sokolov, K., Aaron, J., Hsu, B., Nida, D., Gillenwater, A., Follen, M., MacAulay, C., Adler-Storthz, K., Korgel, B., Descour, M., Pasqualini, R., Arap, W., Lam, W., Richards-Kortum, R. Optical Systems for *In Vivo* Molecular Imaging of Cancer. *Technology in Cancer Research & Treatment* 2, 491-504 (2003).
40. Baak, J. P., Kruse, A. J., Robboy, S. J., Janssen, E. A., van Diermen, B., Skaland, I. Dynamic behavioural interpretation of cervical intraepithelial neoplasia with molecular biomarkers. *J Clin Pathol* 59, 1017-1028 (2006).
41. Mathur, S. P., Mathur, R. S., Creasman, W. T., Underwood, P. B., Kohler, M. Early non-invasive diagnosis of cervical cancer: beyond Pap smears and human papilloma virus (HPV) testing. *Cancer Biomark* 1, 183-191 (2005).
42. Brink, A. A., Snijders, P. J., Meijer, C. J. HPV detection methods. *Dis Markers* 23, 273-281 (2007).
43. Syrjänen, K. J. Immunohistochemistry in assessment of molecular pathogenesis of cervical carcinogenesis. *Eur J Gynaecol Oncol* 26, 5-19 (2005).
44. Takamatsu, T., Fujita, S. Microscopic tomography by laser scanning microscopy and its three-dimensional reconstruction. *J Microsc* 149, 167-174 (1988).
45. Carlsson, K. Three-dimensional specimen reconstruction by confocal microscopy and digital image processing. *Bull Assoc Anat (Nancy)* 75, 105-108 (1991).
46. Boon, M. E., Kok, L. P., Sutudja, G., Dutrieux, R. P. Confocal sectioning of thick, otherwise undiagnosable cell groupings in cervical smears. *Acta Cytol* 37, 40-48 (1993).
47. Kahn, E., Lizard, G., Frouin, F., Roignot, P., Chardonnet, Y., Di Paola, R. Factor analysis of confocal image sequences of human papillomavirus DNA revealed with fast red in cervical tissue sections stained with TOTO-iodide. *Anal Quant Cytol Histol* 22, 168-174 (2000).
48. Crist, K. A., Kim, K., Goldblatt, P. J., Boone, C. W., Kelloff, G. J., You, M. DNA Quantification in Cervical Intraepithelial Neoplasia Thick Tissue Sections by Confocal Laser Scanning Microscopy. *Journal Of Cellular Biochemistry -Supplement* 25, 49-56 (1996).
49. Pavlova, I., Sokolov, K., Drezek, R., Malpica, A., Follen, M., Richards-Kortum, R. Microanatomical and Biochemical Origins of Normal and Precancerous Cervical Autofluorescence Using Laser-scanning Fluorescence Confocal Microscopy. *Photochemistry and Photobiology* 77, 550-555 (2003).
50. Drezek, R. A., Collier, T., Brookner, C. K., Malpica, A., Lotan, R., Richards-Kortum, R., Follen, M. Laser scanning confocal microscopy of cervical tissue before and after application of acetic acid. *Am J Obstet Gynecol* 182, 1135-1139 (2000).
51. Zuluaga, A. F., Drezek, R., Collier, T., Lotan, R., Follen, M., Richards-Kortum, R. Contrast agents for confocal microscopy: how simple chemicals affect confocal images of normal and cancer cells in suspension. *Journal of Biomedical Optics* 7, 398-403 (2002).
52. Collier, T., Shen, P., de Pradier, B., Sung, K.-B., Richards-Kortum, R. Near Real Time Confocal Microscopy of Amelanotic Tissue: Dynamics of Aceto-Whitening Enable Nuclear Segmentation. *Optics Express* 6, 40-48 (2000).
53. Loewke, K. E., Camarillo, D. B., Jobst, C. A., Salisbury, J. K. Real-time image mosaicing for medical applications. *Stud Health Technol Inform* 125, 304-309 (2007).
54. Carlson, K. D., Pavlova, I., Collier, T., Descour, M., Follen, M., and Richards-Kortum, R. Confocal microscopy: Imaging cervical precancerous lesions. *Gynecol Oncol* 99, 84-88 (2005).
55. Collier, T., Follen, M., Malpica, A., and Richards-Kortum, R. Sources of scattering in cervical tissue- determination of the scattering coefficient by confocal microscopy. *Applied Optics* 44, 2072-2081 (2005).
56. Collier, T., Guillaud, M., Follen, M., Malpica, A., Richards-Kortum,

- R. Real Time Reflectance Confocal Microscopy: Comparison of Two Dimensional Images and Three Dimensional Image Stacks for Detection of Cervical Precancer. *Journal of Biomedical Optics* 12, 024021 (2007).
57. Sung, K.-B., Liang, C., Descour, M., Collier, T., Follen, M., Richards-Kortum, R. R. Fiber-Optic Confocal Reflectance Microscope With Miniature Objective for *In Vivo* Imaging of Human Tissues. *IEEE Transactions On Biomedical Engineering* 49, 1168-1172 (2002).
 58. Jean, F., Bourg-Heckly, G., and Viellerobe, B. Fibered confocal spectroscopy and multicolor imaging system for *in vivo* fluorescence analysis. *Opt Express* 15, 4008-4017 (2007).
 59. Carlson, K., Chidley, M., Sung, K. B., Descour, M., Gillenwater, A., Follen, M., Richards-Kortum, R. *In vivo* fiber-optic confocal reflectance microscope with an injection-molded plastic miniature objective lens. *Appl Opt* 44, 1792-1797 (2005).
 60. Luck, B., Carlson, K., Collier, T., Sung, K.-B. Confocal microscopy [detecting and diagnosing cancers]. *IEEE Potentials* 23, 14-17 (2004).
 61. Choi, H. J., Choi, I. H., Kim, T. Y., Cho, N. H., Choi, H. K. Three-dimensional visualization and quantitative analysis of cervical cell nuclei with confocal laser scanning microscopy. *Anal Quant Cytol Histol* 27, 174-180 (2005).
 62. Luck, B., Bovic, A., and Richards-Kortum, R. Segmenting cervical epithelial nuclei from confocal images Gaussian Markov random fields. in *Image Processing 2003, ICIP 2003, Vol. 2*, pp. 14-17 (2003).
 63. Rahman, M., Abd-El-Barr, M., Mack, V., Tkaczyk, T., Sokolov, K., Richards-Kortum, R., Descour, M. Optical imaging of cervical precancers with structured illumination: An integrated approach. *Gynecologic Oncology* 99, S112-S115 Suppl. (2005).
 64. Collier, T., Lacy, A., Richards-Kortum, R., Malpica, A., Follen, M. Near real-time confocal microscopy of amelanotic tissue: detection of dysplasia in *ex vivo* cervical tissue. *Acad Radiol* 9, 504-512 (2002).
 65. Huang, D., Swanson, E. A., Lin, C. P., Schuman, J. S., Stinson, W. G., Chang, W., Hee, M. R., Flotte, T., Gregory, K., Puliafito, C. A., Fujimoto, J. G. Optical coherence tomography. *Science* 254, 1178-1181 (1991).
 66. Fujimoto, J. G., Pitris, C., Boppart, S. A., Brezinski, M. E. Optical Coherence Tomography: An Emerging Technology for Biomedical Imaging and Optical Biopsy. *Neoplasia* 2, 9-25 (2000).
 67. Tomlins, P. H., Wang, R. K. Theory, developments and applications of optical coherence tomography. *J Phys D: Appl Phys* 38, 2519-2535 (2005).
 68. Yang, C. Molecular Contrast Optical Coherence Tomography: A Review. *Photochem Photobiol* 81, 215-237 (2005).
 69. Podoleanu, A. G. Optical coherence tomography. *British Journal of Radiology* 78, 976-988 (2005).
 70. Jenkins, M. W., Adler, D. C., Gargasha, M., Huber, R., Rothenberg, F., Belding, J., Watanabe, M., Wilson, D. L., Fujimoto, J. G., and Rollins, A. M. Ultrahigh-speed optical coherence tomography imaging and visualization of the embryonic avian heart using a buffered Fourier Domain Mode Locked laser. *Opt Express* 15, 6251-6267 (2007).
 71. Bazant-Hegemark, F., Stone, N. Towards automated classification of clinical optical coherence tomography data of dense tissues. *Lasers in Medical Science*, In press (2008).
 72. Sergeev, A., Gelikonov, V., Gelikonov, G., Feldchtein, F., Kuranov, R., Gladkova, N., Shakhova, N., Snopova, L., Shakhov, A., Kuznetzova, I., Denisenko, A., Pochinko, V., Chumakov, Y., and Streltsova, O. *In vivo* endoscopic OCT imaging of precancer and cancer states of human mucosa. *Opt. Express* 1, 432-440 (1997).
 73. Feldchtein, F. I., Gelikonov, G. V., Gelikonov, V. M., Kuranov, R. V., Sergeev, A. M., Gladkova, N. D., Shakhov, A. V., Shakhova, N. M., Snopova, L. B., Terent'eva, A. B., Zagainova, E. V., Chumakov, YuP., Kuznetzova, I. A. Endoscopic applications of optical coherence tomography. *Optics Express* 3, 257-270 (1998).
 74. Pitris, C., Goodman, A., Boppart, S. A., Libus, J. J., Fujimoto, J. G., Brezinski, M. E. High-resolution imaging of gynecologic neoplasms using optical coherence tomography. *Obstet Gynecol* 93, 135-139 (1999).
 75. Escobar, P. F., Belinson, J. L., White, A., Shakhova, N. M., Feldchtein, F. I., Karet, M. V., Gladkova, N. D. Diagnostic efficacy of optical coherence tomography in the management of preinvasive and invasive cancer of uterine cervix and vulva. *Int J Gynecol Cancer* 14, 470-474 (2004).
 76. Escobar, P. F., Rojas-Espaillet, L., Belinson, J. L. Optical diagnosis of cervical dysplasia. *Int J Gynaecol Obstet* 89, 63-64 (2005).
 77. Escobar, P. F., Rojas-Espaillet, L., Tisci, S., Enerson, C., Brainard, J., Smith, J., Tresser, N. J., Feldchtein, F. I., Rojas, L. B., Belinson, J. L. Optical coherence tomography as a diagnostic aid to visual inspection and colposcopy for preinvasive and invasive cancer of the uterine cervix. *Int J Gynecol Cancer* 16, 1815-1822 (2006).
 78. Zuluaga, A. F., Follen, M., Boiko, I., Malpica, A., Richards-Kortum, R. Optical coherence tomography: a pilot study of a new imaging technique for noninvasive examination of cervical tissue. *Am J Obstet Gynecol* 193, 83-88 (2005).
 79. Hsiung, P.-L., Pantanowitz, L., Aguirre, A. D., Chen, Y., Phatak, D., Ko, T. H., Bourquin, S., Schnitt, S. J., Raza, S., Connolly, J. L., Mashimo, H., Fujimoto, J. G. Ultrahigh-resolution and 3-dimensional optical coherence tomography *ex vivo* imaging of the large and small intestines. *Gastrointest Endosc* 62, 561-574 (2005).
 80. Lee, S. W., Yoo, J. Y., Kang, J. H., Kang, M. S., Jung, S. H., Chong, Y., Cha, D. S., Han, K. H., Kim, B. M. Optical diagnosis of cervical intraepithelial neoplasia (CIN) using polarization-sensitive optical coherence tomography. *Optics Express*, 2709-2719 (2008).
 81. Turchin, I. V., Sergeeva, E. A., Dolin, L. S., Kamensky, V. A., Shakhova, N. M., Richards-Kortum, R. Novel algorithm of processing optical coherence tomography images for differentiation of biological tissue pathologies. *J Biomed Opt* 10, 064024 (2005).
 82. Bazant-Hegemark, F., Stone, N., Read, M. D., McCarthy, K., Wang, R. K. Optical coherence tomography (OCT) imaging and computer aided diagnosis of human cervical tissue specimens. *Proc SPIE* 6627-6665 (2007).
 83. Bazant-Hegemark, F., Stone, N. Near real-time classification of optical coherence tomography data using principal components fed linear discriminant analysis. *J Biomed Opt* 13, 034002 (2008).
 84. Yun, S. H., Tearney, G. J., Vakoc, B. J., Shishkov, M., Oh, W. Y., Desjardins, A. E., Suter, M. J., Chan, R. C., Evans, J. A., Jang, I., Nishioka, N. S., de Boer, J. F., and Bouma, B. E. Comprehensive volumetric optical microscopy *in vivo*. *Nature Medicine* 12, 1429-1433 (2006).
 85. Barton, J. K., Guzman, F., Tumlinson, A. Dual modality instrument for simultaneous optical coherence tomography imaging and fluorescence spectroscopy. *J Biomed Opt* 9, 618-623 (2004).
 86. Wong, P. T. T., Wong, R. K., Caputo, T. A., Godwin, T. A., Rigas, B. Infrared spectroscopy of exfoliated human cervical cells: evidence of extensive structural changes during carcinogenesis. *Proc Natl Acad Sci USA* 88, 10988-10992 (1991).
 87. Walsh, M., German, M., Singh, M., Pollock, H., Hammiche, A., Kyrgiou, M., Stringfellow, H., Paraskevaidis, E., Martin-Hirsch, P., Martin, F. IR microspectroscopy: potential applications in cervical cancer screening. *Cancer Letters* 246, 1-11 (2007).
 88. Cohenford, M. A., Godwin, T. A., Cahn, F., Bhandare, P., Caputo, T. A., Rigas, B. Infrared spectroscopy of normal and abnormal cervical smears: evaluation by principal component analysis. *Gynecol Oncol* 66, 59-65 (1997).
 89. Fung Kee Fung, M., Senterman, M., Eid, P., Faught, W., Mikhael, N. Z., Wong, P. T. Comparison of Fourier-transform infrared spectroscopic screening of exfoliated cervical cells with standard Papanicolaou screening. *Gynecol Oncol* 66, 10-15 (1997).
 90. Chiriboga, L., Xie, P., Yee, H., Vigorita, V., Zarou, D., Zakim, D., Diem, M. Infrared spectroscopy of human tissue. I. Differentiation and maturation of epithelial cells in human cervix. *Biospectroscopy* 4, 47-53 (1998).

91. Wood, B. R., Quinn, M. A., Tait, B., Ashdown, M., Hislop, T., Romeo, M., McNaughton, D. FTIR microspectroscopic study of cell types and potential confounding variables in screening for cervical malignancies. *Biospectroscopy* 4, 75-91 (1998).
92. Neviliappan, S., Fang Kan, L., Tiang Lee Walter, T., Arulkumaran, S., Wong, P. T. Infrared spectral features of exfoliated cervical cells, cervical adenocarcinoma tissue, and an adenocarcinoma cell line (SiSo). *Gynecol Oncol* 85, 170-174 (2002).
93. Cohenford, M. A., Rigas, B. Cytologically normal cells from neoplastic cervical samples display extensive structural abnormalities on IR spectroscopy: implications for tumor biology. *Proc Natl Acad Sci USA* 95, 15327-15332 (1998).
94. Chang, J. I., Huang, Y. B., Wu, P. C., Chen, C. C., Huang, S. C., Tsai, Y. H. Characterization of human cervical precancerous tissue through the fourier transform infrared microscopy with mapping method. *Gynecol Oncol* 91, 577-583 (2003).
95. Podshyvalov, A., Sahu, R. K., Mark, S., Kantarovich, K., Guterman, H., Goldstein, J., Jagannathan, R., Argov, S., Mordechai, S. Distinction of cervical cancer biopsies by use of infrared microspectroscopy and probabilistic neural networks. *Appl Opt* 44, 3725-3734 (2005).
96. Mordechai, S., Sahu, R. K., Hammody, Z., Mark, S., Kantarovich, K., Guterman, H., Podshyvalov, A., Goldstein, J., Argov, S. Possible common biomarkers from FTIR microspectroscopy of cervical cancer and melanoma. *J Microsc* 215, 86-91 (2004).
97. Lowry, S. R. The analysis of exfoliated cervical cells by infrared microscopy. *Cell Mol Biol (Noisy-le-grand)* 44, 169-177 (1998).
98. Romeo, M., Burden, F., Quinn, M., Wood, B., McNaughton, D. Infrared microspectroscopy and artificial neural networks in the diagnosis of cervical cancer. *Cell Mol Biol (Noisy-le-grand)* 44, 179-187 (1998).
99. Romeo, M. J., Quinn, M. A., Burden, F. R., McNaughton, D. Influence of benign cellular changes in diagnosis of cervical cancer using IR microspectroscopy. *Biopolymers* 67, 362-366 (2002).
100. Wong, P. T. T., Senterman, M. K., Jackli, P., Wong, R. K., Salib, S., Campbell, C. E., Feigel, R., Faught, W., Fung, M. F. K. Detailed account of confounding factors in interpretation of FTIR spectra of exfoliated cervical cells. *Biopolymers* 67, 376-386 (2002).
101. Chiriboga, L., Yee, H., Diem, M., Wood, B. Comment on: *Gynecol Oncol*. 2002 Apr;85(1):170-4. Infrared spectral features of exfoliated cervical cells, cervical adenocarcinoma tissue, and an adenocarcinoma cell line (SiSo). By Neviliappan S., et al. *Gynecol Oncol* 91, 275-276; author reply 276-277 (2003).
102. Chiriboga, P., Xie, H., Yee, D., Zarou, D., Zakim, and Diem, M. Infrared spectroscopy of human tissue. IV. Detection of human cervical tissue via infrared microscopy. *Cell Mol Biol Lett* 44, 219-229 (1998).
103. Romeo, M. J., Wood, B. R., Quinn, M. A., McNaughton, D. Removal of blood components from cervical smears: implications for cancer diagnosis using FTIR spectroscopy. *Biopolymers* 72, 69-76 (2003).
104. Morris, B. J., Lee, C., Nightingale, B. N., Molodysky, E., Morris, L. J., Appio, R., Sternhell, S., Cardona, M., Mackerras, D., Irwig, L. M. Fourier transform infrared spectroscopy of dysplastic, papillomavirus-positive cervicovaginal lavage specimens. *Gynecol Oncol* 56, 245-249 (1995).
105. Walsh, M. J., Singh, M. N., Pollock, H. M., Cooper, L. J., German, M. J., Stringfellow, H. F., Fullwood, N. J., Paraskevaidis, E., Martin-Hirsch, P. L., Martin, F. L. ATR microspectroscopy with multivariate analysis segregates grades of exfoliative cervical cytology. *Biochem Biophys Res Commun* 352, 213-219 (2007).
106. Wood, B. R., Chiriboga, L., Yee, H., Quinn, M. A., McNaughton, D., Diem, M. Fourier transform infrared (FTIR) spectral mapping of the cervical transformation zone, and dysplastic squamous epithelium. *Gynecol Oncol* 93, 59-68 (2004).
107. Steller, W., Einkenkel, J., Horn, L. C., Braumann, U. D., Binder, H., Salzer, R., Krafft, C. Delimitation of squamous cell cervical carcinoma using infrared microspectroscopic imaging. *Anal Bioanal Chem* 384, 145-154 (2006).
108. Mark, S., Sahu, R. K., Kantarovich, K., Podshyvalov, A., Guterman, H., Goldstein, J., Jagannathan, R., Argov, S., Mordechai, S. Fourier transform infrared microspectroscopy as a quantitative diagnostic tool for assignment of premalignancy grading in cervical neoplasia. *J Biomed Opt* 9, 558-567 (2004).
109. Wood, B. R., Bambery, K. R., Evans, C. J., Quinn, M. A., McNaughton, D. A three-dimensional multivariate image processing technique for the analysis of FTIR spectroscopic images of multiple tissue sections. *BMC Medical Imaging* 6, (2006).
110. Stone, N., Kendall, C., Shepherd, N., Crow, P., Barr, H. Near-Infrared Raman Spectroscopy for the Classification of Epithelial Pre-Cancers and Cancers. *Journal of Raman Spectroscopy* 33, 564-573 (2002).
111. Utzinger, U., Richards-Kortum, R. R. Fiber optic probes for biomedical optical spectroscopy. *J Biomed Opt* 8, 121-147 (2003).
112. Mahadevan-Jansen, A., Follen-Mitchell, M., Ramanujam, N., Utzinger, U., Richards-Kortum, R. Development of a Fiber Optic Probe to Measure NIR Raman Spectra of Cervical Tissue *In Vivo*. *Photochemistry and Photobiology* 68, 427-431 (1998).
113. Yazdi, Y., Ramanujam, N., Lotan, R., Follen Mitchell, M., Hittelman, W., Richards-Kortum, R. Resonance Raman spectroscopy at 257nm excitation of normal and malignant cultured breast and cervical cells. *Appl Spectrosc* 53, 82-85 (1999).
114. Lieber, C. A., Mahadevan-Jansen, A. Automated method for subtraction of fluorescence from biological Raman spectra. *Appl Spectrosc* 57, 1363-1367 (2003).
115. Hutchings, J., Kendall, C. A., Stone, N. Rapid Raman microscopic biochemical imaging for potential histological screening. *Proc. SPIE*, 6853A-33 (2008).
116. Hart-Prieto, M. C., Matousek, P., Towrie, M., Parker, A. W., Wright, M., Ritchie, A. W., Stone, N. The use of picosecond Kerr-gated Raman spectroscopy to suppress signals from both surface and deep layers in bladder and prostate tissue. *Journal of Biomedical Optics* 10, 4 (2005).
117. Krishna, C. M., Prathima, N. B., Malini, R., Vadhiraaja, B. M., Bhatt, R. A., Fernandes, D. J., Kushtagi, P., Vidyasagar, M. S., Kartha, V. B. Raman spectroscopy studies for diagnosis of cancers in human uterine cervix. *Vibrational Spectroscopy* 41, 136-141 (2006).
118. Stone, N., Kendall, C., Smith, J., Crow, P., Barr, H. Raman spectroscopy for identification of epithelial cancers. *Faraday Discussions: Applications of Spectroscopy to Biomedical Problems* 126, 141-157 (2004).
119. Liu, C. H., Das, B. B., Sha Glassman, W. L., Tang, G. C., Yoo, K. M., Zhu, H. R., Akins, D. L., Lubicz, S. S., Cleary, J., Prudente, R., et al. Raman, fluorescence, and time-resolved light scattering as optical diagnostic techniques to separate diseased and normal biomedical media. *J Photochem Photobiol B* 16, 187-209 (1992).
120. Mahadevan-Jansen, A., Follen-Mitchell, M., Ramanujam, N., Malpica, A., Thomsen, S., Utzinger, U., Richards-Kortum, R. Near-Infrared Raman Spectroscopy for *In Vitro* Detection of Cervical Precancers. *Photochem Photobiol* 68, 123-132 (1998).
121. Mourant, J. R., Short, K. W., Carpenter, S., Kunapareddy, N., Coburn, L., Powers, T. M., Freyer, J. P. Biochemical differences in tumorigenic and nontumorigenic cells measured by Raman and infrared spectroscopy. *J Biomed Opt* 10, 031106 (2005).
122. Utzinger, U., Heintzelman, D. L., Mahadevan-Jansen, A., Malpica, A., Follen, M., Richards-Kortum, R. Near-Infrared Raman Spectroscopy for *In Vivo* Detection of Cervical Precancers. *Applied Spectroscopy* 55, 955-959 (2001).
123. Lyng, F. M., et al. Vibrational spectroscopy for cervical cancer pathology. From biochemical analysis to diagnostic tool. *Experimental and Molecular Pathology* 82, 121-129 (2007).
124. Robichaux-Viehoever, A., Kanter, E., Shappell, H., Billheimer, D., Jones, III, H., Mahadevan-Jansen, A. Characterization of Raman Spectra Measured *In Vivo* for the Detection of Cervical Dysplasia. *Applied Spectroscopy* 61, 986-993 (2007).

125. Jess, P. R. T., Smith, D. D. W., Mazilu, M., Dholakia, K., Riches, A. C., Herrington, C. S. Early detection of cervical neoplasia by Raman spectroscopy. *Int J Cancer* 121, 2723-2728 (2007).
126. Krafft, C., Sergo, V. Biomedical applications of Raman and infrared spectroscopy to diagnose tissues. *Spectroscopy* 20, 195-218 (2006)
127. Krishna, C. M., Sockalingum, G. D., Vadhira, B. M., Maheedhar, K., Rao, A. C., Rao, L., Venteo, L., Pluot, M., Fernandes, D. J., Vidyasagar, M. S., Kartha, V. B., Manfait, M. Vibrational spectroscopy studies of formalin-fixed cervix tissues. *Biopolymers* 85, 214-221 (2007).
128. Faoláin, E. O., Hunter, M. B., Byrne, J. M., Kelehan, P., Lambkin, H. A., Byrne, H. J., Lyng, F. M. Raman spectroscopic evaluation of efficacy of current paraffin wax section dewaxing agents. *J Histochem Cytochem* 53, 121-129 (2005).
129. Huang, Z., McWilliams, A., Lam, S., English, J., McLean, D. I., Lui, H., Zeng, H. Effect of formalin fixation on the near-infrared Raman spectroscopy of normal and cancerous human bronchial tissues. *Int J Oncol* 23, 649-655 (2003).
130. Motz, J. T., Gandhi, S. J., Scepanovic, O. R., Haka, A. S., Kramer, J. R., Dasari, R. R., Feld, M. S. Real-time Raman system for *in vivo* disease diagnosis. *J Biomed Opt* 10, 031113 (2005).
131. Wu, T. T., Qu, J. Y. Optical imaging for medical diagnosis based on active stereo vision and motion tracking. *Opt Express* 15, 10421-10426 (2007).
132. Majumder, S. K., Kanter, E., Robichaux Viehovever, A., Jones, III, H., Mahadevan-Jansen, A. Near-infrared Raman spectroscopy for *in-vivo* diagnosis of cervical dysplasia: a probability-based multi-class diagnostic algorithm. *Proc. SPIE* 6430, 64300Q (2007).
133. Baker, R., Matousek, P., Ronayne, K. L., Parker, A. W., Rogers, K., Stone, N. Depth profiling of calcifications in breast tissue using picosecond Kerr-gated Raman spectroscopy. *Analyst* 132, 48-53 (2007).
134. Kendall, C., Stone, N., Shepherd, N., Geboes, K., Warren, B., Bennett, R., Barr, H. Raman spectroscopy, a potential tool for the objective identification and classification of neoplasia in Barrett's oesophagus. *J Pathol* 200, 602-609 (2003)
135. Orfanoudaki, I. M., Themelis, G. C., Sifakis, S. K., Fragouli, D. H., Panayiotides, J. G., Vazgiouraki, E. M., Koumantakis, E. E. A clinical study of optical biopsy of the uterine cervix using a multispectral imaging system. *Gynecol Oncol* 96, 119-131 (2005).
136. Alvarez, R. D., Wright, T. C. Effective cervical neoplasia detection with a novel optical detection system: A randomized trial. *Gynecologic Oncology* 104, 281-289 (2007).
137. Lurie, S., Eliaz, M., Boaz, M., Levy, T., Golan, A., Sadan, O. Distribution of cervical intraepithelial neoplasia across the cervix is random. *Am J Obstet Gynecol* 196, 125 (2007).
138. Franco, E. L., Duarte-Franco, E., Ferenczy, A. Prospects for controlling cervical cancer at the turn of the century. *Salud pública de México* 45 (3 Suppl), S367-S375 (2003).
139. Monsonego, J. HPV infections and cervical cancer prevention. Priorities and new directions. Highlights of EUROGIN 2004 International Expert Meeting, Nice, France, October 21-23, 2004. *Gynecologic Oncology* (2004).
140. Vergote, I., van der Zee, A. G. J., Kesic, V., Sert, B., Robova, H., Rob, L., Reed, N., Luesley, D. M., Leblanc, E., Hagen, B., Gitsch, G., Du Bois, A., Di Vagno, G., Colombo, N., Beller, U., Ayhan, A., Jacobs, I. ESGO statement on cervical cancer vaccination. *Int J Gynecol Cancer* 17, 1183-1185 (2007).

Received: May 16, 2008; Revised: October 24, 2008;

Accepted: November 4, 2008

Towards automated classification of clinical optical coherence tomography data of dense tissues

Florian Bazant-Hegemark · Nicholas Stone

Received: 25 April 2008 / Accepted: 1 September 2008
© Springer-Verlag London Ltd 2008

Abstract The native contrast of optical coherence tomography (OCT) data in dense tissues can pose a challenge for clinical decision making. Automated data evaluation is one way of enhancing the clinical utility of measurements. Methods for extracting information from structural OCT data are appraised here. A-scan analysis allows characterization of layer thickness and scattering parameters, whereas image analysis renders itself to segmentation, texture and speckle analysis. All fully automated approaches combine pre-processing, feature registration, data reduction, and classification. Pre-processing requires de-noising, feature recognition, normalization and refining. In the current literature, image exclusion criteria, initial parameters, or manual input are common requirements. The interest of the presented methods lies in the prospect of objective, quick, and/or post-acquisition processing. There is a potential to improve clinical decision making based on automated processing of OCT data.

Keywords Biomedical image processing · Biomedical optical imaging · Image analysis · Image classification · Medical diagnosis

Introduction

Optical coherence tomography (OCT) is an imaging modality based on low coherence interferometry and is frequently compared to ultrasonography. Typical light sources operate in the near infrared range and allow

location of a reflection event with an accuracy of a few micrometers on an axial scan (A-scan). Several A-scans can be combined to a two-dimensional (2-D) image (B-scan) or volumetric sets [1]. Penetration depth and resolution of OCT fit between confocal microscopy and high frequency ultrasound. Specifically in the medical field, this non-destructive imaging is attractive [2]. Through rapid high resolution imaging, detailed structural information becomes available—in vast datasets.

Structural information is often intuitive to view, as shown in Fig. 1. Strictly speaking, however, an OCT image might not be relevant in itself. For clinical diagnosis, a new repertoire of image interpretation skills and artifact knowledge is necessary, which constitutes a learning threshold for the intended operator. There are, therefore, efforts being made to provide more than a structural image alone.

Challenges with optical coherence tomography

Contrast

In dense tissue, characteristic limitations of OCT are the penetration depth in the magnitude of 1 mm and the mode of contrast, created by the variations of the refractive index. This native contrast can be weak, especially when compared with that of other non-destructive high-resolution methods. For example, fluorescence imaging or Raman spectroscopy are able to pick up single molecules or alterations of protein structures [3, 4]. OCT has advantages over these techniques: it is able to maintain the high resolution over a large area; it is able to provide depth-resolved high resolution *in vivo*; and the requirements of OCT instrumentation subsequently mean that OCT does not rely on imaging conditions such as reduced ambient light.

Since it would be attractive to have the combination of a strong marker and wide-field, high-resolution, depth-resolved, rapid data acquisition, there are efforts being made to engineer multimodal systems [5, 6]. Equally

F. Bazant-Hegemark
Cranfield Health, Cranfield University at Silsoe,
Silsoe, Bedfordshire, UK

F. Bazant-Hegemark · N. Stone (✉)
Biophotonics Research Group, Gloucestershire Royal Hospital,
Great Western Road,
Gloucester GL1 3NN, UK
e-mail: n.stone@medical-research-centre.com

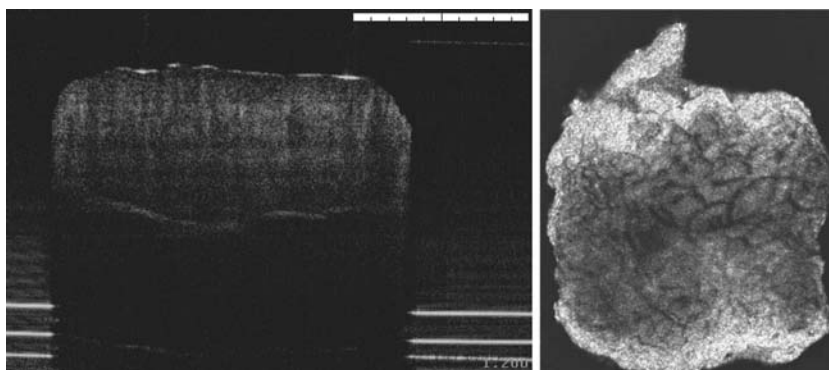


Fig. 1 *Left* B-scan of an esophageal biopsy specimen, of approximately 2 mm×2 mm size, obtained ex vivo by a 7.5 μm resolution (in air) frequency domain system operating at 1,300 nm (Michelson Diagnostics EX1301 OCT microscope, Michelson Diagnostics Ltd.). The faint tissue layers make it difficult for one to decide whether clear layers are amiss due to weak contrast or because of the clinical

importantly, attention is being brought to improving clinical contrast in OCT data: contrast agents have the potential to provide a clear signal for clinical decision making and have become quite a large research field [7, 8]. In addition to the structural characteristics, OCT data can provide functional information. Functional OCT imaging includes Doppler OCT, polarization-sensitive OCT, second-harmonic OCT, or spectroscopic OCT. These carry the potential for providing physiological information about flow properties [9, 10], birefringence [11, 12], neural activity [13, 14], or water content and oxygenation states [15, 16].

Automated data evaluation

A further approach to enhance clinical utility is to analyze OCT data by computational methods. This area has the potential to provide unsupervised, reproducible and objective measures and to provide both quick and/or post-acquisition classification. When engaging ourselves in this field, we found that research groups tend to use different approaches, and the emphasis is often on image acquisition rather than data analysis. As OCT is a novel technique, there is still a tendency for the OCT image to be treated as a result rather than as a step towards a clinician decision following detailed analysis. Subsequently, in some papers, the authors even describe their analyses in the results section. We wanted to investigate whether there is a common pattern or ideal algorithm that could be suggested for further research.

Rationale and limitations

The topic of automating OCT data evaluation has to be seen in context with other contrast-enhancing approaches. Algorithms can fail, and a clear marker, be it from contrast enhancing or functional data, will facilitate processing. In

presentation of the disease. *Right* En face sub-surface section, viewed from the top, of a volumetric data set composed of 500 such B-scans. The improved representation, as opposed to 2-D images, is evident. The raw data require several gigabytes of memory. The clinical requirement for this technique would be to contribute to diagnosis and determine whether the patient should receive cancer treatment

this manuscript, we want to concentrate on automated algorithms for retrieving information from conventional ‘intensity’ data. We therefore refer the reader to the above-mentioned literature for more detail on functional OCT and contrast enhancing. Even when limiting ourselves to structural data of dense tissues, we certainly cannot claim to provide an exhaustive summary of every attempt so far. However, with this manuscript, we would like to provide a starting point from where further analysis attempts can be systematically explored and combined with other image-enhancing methods for OCT.

Selection

Table 1 gives an overview of the approaches discussed in this manuscript. In general, automation requires recognition of relevant features. The identification of such a feature has to be mastered by an automation algorithm, which requires calculation time and has the potential to introduce errors. The grouping in Table 1 is somewhat arbitrary; e.g., most of the texture-based approaches are subsequently classified. However, Table 1 shows, for instance, that A-scan analysis renders itself to the measurement of layer thickness or for the derivation of optical tissue parameters, while B-scans permit speckle and texture analysis.

Bulk optical properties

Quantitative measurements from OCT data are a first step for establishing parameters and features for automation.

Quantitative measures

Examples of quantitative measurements from OCT data are work on the scattering properties and refractive index of

Table 1 Articles employing feature recognition or automated steps for analysis of dense tissue OCT data (*ROI* region of interest)

Parameter studied	First author and year	Specimen type	Sample size for analysis	Objective in cited study	Remark
Optical properties	Esenaliev 2001 [19], Kuranov 2006 [24] Levitz 2004 [25]	phantom, animal tissue arteries	–, 8 pigs 151 regions, 14 images	monitor blood glucose levels differentiate normal vs atherosclerotic	automation not mentioned but seems feasible ROI and initial parameters required
	van der Meer 2005 [20] Turchin 2005 [23]	arteries phantom, epithelium (cervix)	20 images, 13 samples 19 images, 1 patient	describe plaques and components classify normal vs pre-malignant	ROI marked manually ROI and initial parameters required
	Jeon 2006 [21] Ramrath 2007 [22]	rat brain tissue rat brain tissue	19 ROIs of 3 images –	guide deep brain stimulation guide deep brain stimulation	ROI marked manually feature correlates with distinct spikes, requires initial training
	Ifimia 2005 [48], Goldberg 2008, [49] Gossage 2003, 2006 [27, 30] MacNeill 2004 [31]	breast tissue Brodatz images, animal tissue, phantoms blood vessels	158 scans, 58 patients 9 test images, 40 tissue images 160 images, 49 patients	guide fine needle aspiration biopsy proof of principle (classify specimen type) macrophage population as predictor of plaque rupture	ROI marked manually feature correlates with distinct spikes, requires initial training feasibility study on 3 biopsy samples, low coherence interferometry equidistance probe—specimen only for homogeneous bulk solutions unsuitable images excluded manually
	Qi 2006 [32] Chen 2008 [33]	upper GI tract upper GI tract	106 images, 13 patients 100 images, 3 patients	classify pre-malignancy provide measure for system resolution	unsuitable images excluded manually ROI marked manually
	Lingley-Papadopoulos 2008 [34] Gambichler 2007 [37]	bladder epithelium	182 images, 21 patients 114 images	pre-cancer staging thickness (general dermatological interest)	ROI marked manually, unsuitable images excluded manually
	Rogowska 2002, 2003 [40, 35] Weissman 2004 [36] Korde 2007 [38]	rabbit cartilage forearm, lower leg epithelium (forearm and upper inner arm)	– 12 patients 764 images, 112 patients including 448 forearm images 5 volunteers	thickness as marker for degradation demarcate benign/cancerous lesions detect actinic keratosis	– – unsuitable images excluded manually. Two methods
	Hori 2006 [42]	infundibula of cheek, forehead, forearm	–	density of infundibula and role as acne risk contributors	surface recognition for flat interfaces
	Bonnema 2008 [44]	cell cultures on blood vessel mimic (BVM)	12 BVM in total, longitudinal study	monitor stent endothelialization	exploits characteristic reflection and shadowing of metallic material
	Zysk 2006 [50] Jørgenson 2008 [47] Cheng 2006, [45]	breast tissue skin fake/normal fingerprints	4015 A-scans, 4 patients 78 lesions, 34 patients 10–20 scans per finger, 8 fingers plus dummies	classify normal–malignant–adipose classify non-melanoma skin cancers detect forgeries	bubbly appearance of adipocytes features measured manually no mention about layer selection
Classification	Bazant-Hegemark 2008 [51] Bazant-Hegemark 2008 [53]	vegetable samples, porcine tissue port wine stains	330 B-scans (11 groups) 80 B- scans (4 groups) 7 images	proof of principle (classify similar groups amongst other groups) prognosis of treatment response	requires surface recognition and normalizing to surface requires initial training

skin [17] and tissue [18], monitoring of blood glucose levels [19], differentiation of atherosclerotic arteries [20] and brain tissues [21, 22], or the determination of optical parameters of cervical tissue [23].

Blood glucose

The monitoring of blood glucose levels is a challenging topic. Esenaliev et al. report an inverse association between the slope of an averaged A-scan and the blood glucose concentration [19]. As the sample properties are being related to OCT data, it is possible to determine and classify them automatically. In the original paper, the way of determining the slope has not been described in detail. In a more recent publication a strict imaging protocol ensures tight contact of a sapphire glass window as part of the probe tip with the tissue surface. This allows one to set the maximum peak of an A-scan as the approximate surface point, which is a prerequisite for the slope determination [24].

Atherosclerotic lesions

Levitz et al. presented an algorithm for determining the scattering coefficient, μ_s , and the anisotropy factor, g_{eff} , from regions of interest [25]. For the measurement, the average A-scan of a region of interest (ROI) area was calculated and further smoothed by a median filter. In order to allow averaging, the biological specimens in this study were placed under a glass surface. The ROI had to be selected manually for this algorithm. This algorithm was tested for differentiation between normal and atherosclerotic arteries. There were 151 ROIs from 14 images that could be classified into four clinically relevant disease states. An algorithm presented by van der Meer et al. in a proof of principle study on 20 lesions in 13 samples showed that the light attenuation coefficient μ_t differs for lipid-rich regions, media, calcifications, and thrombi in ROIs [20]. There is no mention of how the ROIs were selected, and presumably this was done manually. An interesting aspect is that the approach allows localized determination of μ_t and visualization as overlay over the B-scan, thereby providing differentiation that would not be obvious from the intensity image.

Brain tissue

Jeon et al. used the maximum reflectance intensity and attenuation rate of the mean of five A-scans to determine the different properties of different brain tissue types of rat brain. Distinguishing white and gray matter might contribute to guiding deep brain stimulation probes in treatment scenarios [21]. The study was performed on 19 manually selected ROIs of three images in total. However, four tissue types (hippocampus, primary somatosensory cortex barrel field,

external and internal capsule region, and optic tract area) were shown to have different optical properties. A different approach was suggested by Ramrath et al.: by a three step procedure, areas within an A-scan that contain white matter are identified [22]. The essential step to recognizing a white brain matter peak amongst speckle noise is to transform the A-scan using a nonlinear-energy operator and a shift-and-multiply operator. Both are discrete counterparts of more complex operators and provide candidates for spikes, i.e., unlikely intensity peaks, amongst noisy data [26]. The identified spikes are validated by matching the attenuation coefficient μ_t within a window after this spike to a reference value for white brain matter. The reference coefficient needs to be provided, e.g., determined manually, but there is no requirement for a pre-defined surface or an ROI. Both approaches were demonstrated on images from small animal tissue. The difference in size from that of a human organ might pose a challenge.

The above-mentioned publications have to be seen as examples for a wealth of research communications treating the determination of optical properties using OCT. The crucial point of automation seems to be how to determine the ROIs, or to have prior knowledge about a pronounced feature. Databases of such properties may well serve as references for automated tissue characterization.

Epithelial neoplasia

Turchin et al. assessed the scattering behavior of phantoms and cervical tissue with progressing malignancy using various parameters that included total scattering coefficient, variance of a small-angle scattering phase function, and the probability of backscattering [23].

They noted that for the classification of epithelial images, layer disruption is not well associated with progression of pre-cancerous lesions. Continuous layers can be seen in healthy and dysplastic tissue alike, and they can be absent in leukoplakia, metaplasia, and neoplasia. They proposed an algorithm which reconstructs scattering properties of the sample. Three parameters, the total scattering coefficient, μ_s , the variance of a small-angle scattering phase function, γ^2 , and the probability of backscattering, p_b , together describe the signal as a function of depth. For a single-layer model, these parameters allow classification in the absence of layers. The calculation for many layers then becomes complex, but, for two-layer samples, the computational effort appears to be manageable. A restriction is that there is an assumption that layers are parallel to the surface. In an approximation, epithelial tissue can be regarded as two parallel layers. Turchin et al. created single-layer and two-layer models. Proof of principle was displayed on tissue phantoms consisting of polystyrene beads (0.14 μm , 1 μm , and 4.75 μm) suspended in water or

glycerol. In addition to the tissue phantoms, the paper mentions that medical data from two clinical centers were used to evaluate the algorithm. However, only the results for 19 images from one patient are shown.

Their paper did not specify how images were selected for either the single-layer or two-layer model, or how well the algorithm works for multiple patients or on samples where the precancerous lesions were unknown prior to collection of the OCT image. Such a proof of principle is, however, an important step in the realization of a clinical classification tool.

Texture and speckle analysis

The information from OCT images does not exclusively consist of evident structure. Speckle noise carries characteristics which can be exploited for classification and even semi-quantitative analysis [27, 28]. This allows determination of properties of homogeneous bulk solutions of particles smaller than the resolution in the absence of intuitive structural information. Table 2 gives an overview of methods which classify OCT data on the basis of B-scan texture.

Speckle patterns

In early work, Hillman et al. attempted to establish a correlation between speckle and the concentration of scatterers (polystyrene microspheres) within the probe volume [28]. The reference for the axial depth of a signal was the bottom of the sample–glass slide interface.

Work by Gossage et al. was based on the fact that collagen does not show structural features on conventional

OCT because of speckle [27]. Speckle noise, however, does carry information from sub-resolution features. The two representative measures for an image were (1) the spatial gray-level dependency matrix (SGLDM) and (2) a derivative of the complex 2-D Fourier transform of a selected image region. The SGLDM, in short, is a measure of the likelihood of a pixel of a defined gray-level to be in the vicinity of another pixel of a defined gray-level. This then can be expressed by various parameters, e.g., energy, entropy, correlation, local homogeneity, or inertia.

‘Brodatz’ images are photographs of textures, and they are commonly used in texture analysis [29]. Nine Brodatz test images of specified texture and ten images each from four types of animal tissues (skin, fat, normal lung, abnormal lung) were randomly split into training and classification sets. A Bayesian classification model selected the three best out of 24 total parameters. They were then classified according to their Mahalanobis distances. Mean classification rates ranged from 37.6% to 94.8% for a three-group tissue model, from 64.0% to 98.5% for two-group tissue-models, and from 97% to 100% for the test images. In a more recent study, Gossage et al. showed the ability to classify phantoms consisting of particles of sub-resolution size with the same method [30].

Macrophage density

Macrophages play a role in inflammation and tissue degradation and are investigated for their role as contributors to atherosclerotic plaque rupture. Macrophage density could be an indicator for vulnerable plaques. Macrophages have a size of 20–50 μm and produce a distinct scattering signal in OCT images. MacNeill et al. automatically measured the macro-

Table 2 Overview of approaches for classification on the basis of texture (*ANOVA* analysis of variance)

Parameter	Gossage et al. [27]	MacNeill et al. [31]	Qi et al. [32]	Lingley-Papadopoulos et al. [34]
Objective	Classify 4 animal tissue types, tissue phantoms and texture images	Macrophage density as measure for plaque vulnerability	Classify pre-malignant esophageal lesions	Classify pre-cancer of the bladder
Valid images	40	160 of 225	106 of 405	182
Pre-processing	<ul style="list-style-type: none"> • Logarithm • Normalization • Histogram equalization 	<ul style="list-style-type: none"> • Median filter • Bimodal segmentation 	<ul style="list-style-type: none"> • Manual cap removal • Binary threshold • Center-symmetric autocorrelation 	<ul style="list-style-type: none"> • Bimodal histogram to determine noise background • Binary threshold
Feature(s)	Spatial gray-level dependency matrix, 2-D Fourier transform	Normalized standard deviation	Textural features (intensity differences of center pixel from its 8 neighbor pixels)	74 texture features reduced to 18 independent features
Classification	Mahalanobis distance	Statistical, ANOVA	“Computer aided diagnosis”	Discriminant function and decision tree

phage density from OCT images [31]. The study consisted of 225 images from 49 patients. Sixty-five of the images were of insufficient quality for analysis and, hence, were manually excluded, leaving 160 images for analysis. The automated segmentation was performed by separation of a bimodal histogram. The normalized standard deviation of the signal intensity was calculated from this segmented area and found to correlate with macrophage density, thereby serving as a potential predictor for unstable plaques. Whilst the automated algorithm works on selected images, an operator would still have to select suitable images. As the authors themselves state, this could be a source of bias.

Epithelial lesions and pre-cancers

Qi et al. used center-symmetric autocorrelation on endoscopic OCT images to recognize automatically the clinically relevant features of Barrett's esophagus, which is a pre-cancerous condition of rising importance in the Western world [32]. Their data consisted of 405 images from 13 patients; however, only 106 of these images were used for the study. The presented algorithm required manual removal of the probe cap from the image. The image was then globally thresholded with a binary mask to remove noise areas. This pre-processed image served as the substrate for the extraction of six statistical textural descriptors for the distribution of pixels, and pixels of similar value on an image. These six parameters, as well as two principal components of them, were analyzed for their correlation with disease status using a "computer aided diagnosis" which was not further specified. The best results were obtained from the local intensity variation (of one pixel and the surrounding eight pixels) and the between-pair intensity variation. The scores from the principal components analysis (PCA) gave poorer results than did the original six parameters.

A study by Chen et al. used 100 images, of normal tissue and of Barrett's esophagus, imaged with both standard resolution and ultra-high resolution (UHR) OCT, using the discriminatory power of textural features between these disease groups, in order to establish whether there is a statistical measure to quantify the difference between normal

and UHR OCT images [33]. The textural features were calculated from ROIs whose selection was not further specified. Features were reduced to two principal components, which were used for the determination of two linear discriminants. The scores showed that UHR images provide a better basis to discriminate normal from dysplastic esophageal epithelium. These data were not cross-validated and, therefore, provide a distance measure, not a prediction on the performance in clinical staging.

Lingley-Papadopoulos et al. used texture features to assess a progressing lack in structure which may be indicative of bladder cancer [34]. In their study, 18 independent texture features, identified from a pool of 74 features, were used to discriminate 182 images of pre-cancerous stages of the bladder, and achieved a sensitivity of 92% and a specificity of 62%. These studies have in common that the textural features are classified using a discriminatory function. More attention to classification is given later in this manuscript.

Segmentation

The vast majority of the literature cited so far relies on manual selection of an ROI or some knowledge about a reference point from which to calculate optical tissue coefficients. Work towards automated segmentation and thickness measurements has been undertaken for cartilage thickness [35], epithelial thickness [36, 37], and sub-surface layer assessment [38], thereby providing a clinically relevant measure rather than a set of images. Table 3 shows a comparison of their approaches.

Layer thickness

Gambichler et al. reported two ways for determining epithelial thickness [37]: (1) the averaging of several A-scans and taking the first peak–valley distance as thickness measure. This requires a second peak to be present, and the authors reported that 40% of their A-scans had to be excluded from analysis for this reason. This approach is automated but has

Table 3 Comparison of approaches in the automated recognition of dense tissue layers

Pre-processing	Gambichler et al. [37]	Rogowska et al. [35]	Weissman et al. [36]	Korde et al. [38]
Feature enhancement	Averaging	Rotating kernel transformation	Factor analysis de-noising Tilt/slant image	Median filter
Feature recognition	Peak–valley distance, or area	Sobel edge detection	Shapelet convolution (several shapes)	Sobel edge detection, skeletonization
Correction	–	Edge linking cost function	Least median of squares	Linear interpolation, rule for multiple candidates, exclusion of top 10 (artifact-rich) pixels

poor correlation with histopathology; hence, it was deemed to be unacceptable for clinical application by the authors; (2) the manual encircling of the epithelium and the determination of the average thickness geometrically by division of the area by its length. This approach is not automated, but, provided that the epithelium can be marked by a thresholding approach similar to the macrophage approach, it would carry the potential for automation.

The work of Rogowska et al. can be regarded as pioneering in dense tissue segmentation from OCT images [35]. They have introduced and explored the rotating kernel transformation (RKT) filter for speckle treatment [39]. The RKT filter emphasizes straight lines within the intensity image and facilitates characterization of the surface and lower boundary of a cartilage layer [40]. It is, however, computationally expensive; edge linking methods can correct for errors. The edge linking used in this approach is semi-automated as it requires start and end nodes to be set manually.

The approach of Weissman et al. aims to recognize the dermis–epidermis junction by convolution with various shapelets (functions based on a Gaussian curve) [36]. Since the contrast of epidermal OCT data was not suitable for this surface recognition, transformed tilt and slant image representations were used instead. Several length scales of the Gaussian shape ensured good convolution results for varying shapes. This pool of convolution results then allows exclusion of outliers, and the optimal candidate for the position of the junction is determined by a least median of squares function. Since several shapes can be adapted for convolution, this approach is not restricted to tissue interfaces and two-dimensional scans.

While both approaches are fully automated, it seems interesting that they were not applied in the more recent works of Gambichler, or compared with each other. This indicates that, although the methods seem to work well in the way they are presented by the authors, they are not straightforward to implement.

Korde et al. have worked towards differentiating skin lesions from actinic keratosis, a precursor of squamous cell carcinoma [38]. Their algorithm utilizes parameters obtained from edge detection and employs a surface recognition algorithm. The slope of the mean A-scan is taken as representative of the ‘average skin attenuation.’ The very surface pixels are excluded from this fit, as they generally contain strong reflection features. A second structural feature was obtained by using vertical Sobel edge detection, which emphasizes horizontal lines. Short edge lines were likely to stem from noise and were therefore excluded. The number of remaining pixels was defined as ‘horizontal edge detection’; this could be characterized as a ratio of cohesive layer boundaries per image area. Whereas the ‘average skin attenuation’ did not show a significant correlation with skin lesions, the ‘horizontal edge detection’ had a sensitivity and

specificity of 73% and 65%, respectively, for recognizing actinic keratosis.

Ophthalmic layers

In ophthalmology, the many interfaces between air, corneal structure, vitreous humor, and retinal layers provide strong OCT contrast. The eye is an ideal organ for OCT imaging, owing to its optical properties and layered structures which are thin enough to allow full penetration and are yet wide enough for full resolving. Being able to observe such layers has revolutionized ophthalmology. By providing a means of automating layer recognition and providing a means of objective classification it is possible to enhance the clinical benefit of these features in the process of clinical decision making. Clinically important features are corneal thickness, retinal damage, and damage to the fovea or optic nerve, and, hence, efforts have been made to characterize data automatically.

The success of OCT in ophthalmology might be best illustrated by the progress in automated analysis, which has no equivalent in dense tissue measurements: As an example, we would like to cite the work of Fernández et al., who have demonstrated that seven layers of the retina can be readily characterized from ophthalmic images and that the retinal layers can be registered from the anterior and posterior sides [41]. Data processing follows an extensive protocol. A threshold cut-off value is used for noise reduction, followed by complex diffusion filter for speckle suppression. From this image the structure–coherence matrix of the image is obtained. The internal limiting membrane is determined by the first inner peak, and the retinal pigment epithelium is determined by the maximum outer peak. Further peaks are correlated with the remaining layers. Outliers are corrected by linear interpolation. This algorithm worked reliably for data from 72 healthy subjects. It can fail in abnormalities where initial assumptions on layer composition are not met. However, it allows the automated generation of 2-D thickness maps of separate retinal layers and, therefore, also longitudinal and observational studies. The authors point out that this study has to be seen as an example for segmentation algorithms for retinal layers, and the analysis and comparison of them would probably merit a separate review.

Segmentation of dense tissue

Hori et al. automated the registration of the infundibula in skin, which are being investigated for their role in the development of acne [42]. They performed the surface recognition by assuming the maximum peak as the surface. This was corrected by a median filter and a so-called Olympic exclusion (which excludes a set amount of minimum and

maximum values regardless of their being outliers or not). The epidermal junction is defined as the first intensity minimum between the first two maxima of an A-scan.

Both values, surface and epidermal junction, were smoothed with a polynomial function to provide a thickness map of the epidermis. Whilst in comparison with the more sophisticated shapelet approach this seems to yield rough results, Hori et al. claim that the difference between their uncorrected points and the fitted curve was within the magnitude of their system resolution [42]. Such a finding opens the question of whether a (computationally expensive) correction would be necessary in the first place.

After establishment of the epidermal thickness, a band (volume) below the epidermis was defined. This band was then treated as an en face image; a 'shadowgram' of this image was characterized by thresholding. From the resulting binary image, infundibula were then characterized as circular structures and extracted with the Danielsson distance mapping algorithm [43]. Distance maps, or distance transforms, are representations of Euclidean distances in a different parametric ordinate system. This facilitates the finding of set geometries, in this case circles on a 2-D image.

The algorithm is able to deliver epidermal thickness parameters of the imaged volume as well as distribution of infundibula. The surface of such a volume covered $4\text{ mm} \times 4\text{ mm}$; however, no information about the speed of the fully automated algorithm was given. Simplifying the determination of the epidermal thickness in contrast to the shapelet analysis by Weissman et al. [36] seems to work well on the volumetric data.

Bonnema et al. used a protocol to determine stent endothelialization on blood vessel mimics [44]. The clinical interest is to monitor the coverage by endothelial cells under different growth conditions. An algorithm thus has to distinguish between a superficial stent, an overgrown (sub-surface) stent, and tissue without a stent mesh. A strut cross-section provides a round shape with a diameter of a few pixels on a B-scan. Characteristic reflection and shadow artifacts from the strut can be exploited, although the protocol remains an extensive set of rules. Briefly, a combination of thresholding and median filtering allows one to determine sharp (strut) peaks and surface peaks amongst artifacts from debris and reflection. This fully automated approach shows a way of providing clinically relevant information without the need for assessing separate B-scans.

Classification

We have already mentioned quantitative measures which seem fit for automated classification. An autocorrelation score can be suitable for expressing the likelihood that an

expected shape will correlate with an OCT image. When a relation is complex, statistical tools can become essential for automated classification. There are many reduction and clustering algorithms, such as least median of squares, PCA, genetic algorithms, or linear discriminant analysis (LDA).

It is common practice to express the likelihood that a data point will belong to a group as a Mahalanobis distance. This is a score representing the distance between groups, taking into account the distribution of data points of one group. These techniques can be combined, or applied in a decision tree. For example, the structural features obtained by Lingley-Papadopoulos et al. were differentiated in a decision tree, where, in each step, a two-group classification model was designed [34].

Autocorrelation scores of an A-scan with itself have been shown to characterize spoof fingerprints [45]. Fingerprint dummies are created from materials like wax or silicone rubber with a plasticine mold. These materials are homogeneous and, therefore, they have rapidly decreasing correlation scores. For fingerprints, de-correlation happens as well, less strongly, and this difference is clear enough to distinguish artificial material from human tissue. There is no mention of how exactly the area of an A-scan is selected for autocorrelation, and, hence, we assume that this was done manually. However, the surface peaks of both spoof fingerprint and tissue can be expected to be pronounced and distinguishable, which should make it possible to automate layer recognition.

Weissman et al. used PCA to compare the agreement of their surface recognition algorithm with the line drawn by human operators [36]. Chen et al. used a combination of PCA and LDA to compare the performance of two systems with different axial resolution [33]. These studies, rather than assessing clinical parameters such as sensitivity and specificity, used classification as an objective performance measure. Qi et al. used PCA for reducing six statistical measures for texture features to two principal components, PC1 and PC2 [32]. As mentioned before, this approach did not provide a benefit over the use of the six-dimensional data.

Glaucoma

Again, we refer to ophthalmology for what is possible when feature parameters are readily available. In order to differentiate normal eyes from glaucomatous eyes, Huang and Chen applied several classification methods [46]: LDA, Mahalanobis distance (MD), artificial neural networks (ANN), and combined PCA/LDA, PCA/MD, and PCA/ANN. The parameters for classification were extracted from a set of 25 morphological parameters (retinal thickness and size). The study involved scans

from 89 participants. Inclusion requirements for analysis were that 100% of the A-scans had to be of good quality and above a specified signal-to-noise ratio (SNR), and one of the features (cup/disc ratio) had to be big enough for automated detection.

This is noteworthy with respect to dense tissue assessment: Even though this classification approach was developed for images which usually allow the extraction of a wealth of morphological parameters, some images did not meet these criteria. In dense tissue OCT imaging, outside the ophthalmologic field, it is even less likely to encounter specimens which show a layered structure to this extent. As shown above, the characteristics of dense tissue layers pose a much greater challenge for reliable demarcation. This might be one reason why, in dense tissue imaging, textural features are candidates for classification.

Skin cancer

Jørgenson et al. used 14 image features which were fed into a support vector machine model to classify actinic keratosis from basal cell carcinomas [47]. Although the approach managed to classify 60 of 78 images correctly, there was no exact specification on how features were detected; e.g., ‘layer thickness’ is “measured by ruler on the screen”, or a degree of ‘graininess’ is categorized. The evaluators therefore needed training on feature recognition. However, the characteristics seemed to have a degree of correlation with disease. It may be assumed that at least some characteristics can be described by methods mentioned above. This would have the potential for an objective and fully automated classification.

Breast tissue

Breast tissue contains human adipocytes that measure between 50 μm and 150 μm . These seem to lend themselves to frequency based analysis: Iftimia et al. described a system intended to guide fine needle aspirations for breast biopsies via A-scans (‘reflectivity profiles’) [48]. Three groups of tissues, (adipose, fibroglandular, and cancerous) have different profiles, and the A-scans therefore have different slopes and standard deviations. In this feasibility study three tissue specimens were measured. A study by Goldberg et al. was performed on 158 of 260 scans from 58 patients [49]. In this study, the determination of the slope was automated. Noise was reduced by cutting off below the average noise level, determined by the signal from the first 200 μm . The surface was then defined as 100 μm below the first zero crossing of the 1st order derivative, in order to avoid contributions from specular reflection. Slope, standard deviation and spatial frequency

content served as parameters to be separated by a multivariate Gaussian classification model. The performance for a validation set consisting of 86 samples was 91.9%.

Zysk and Boppart reported three methods to classify A-scans of breast cancer biopsies into three relevant groups (adipose, cancerous, stroma): (1) taking the Fourier domain component of an A-scan and comparing it with the mean of a training set; (2) using periodicity analysis, which looks at the distance between high-intensity sub-surface peaks and the surface peak, and (3) a combination of both [50]. The periodicity analysis seems to benefit from the above-mentioned bubbly structural pattern in OCT images of adipose breast tissue, resulting in A-scans with several distinct spikes. Cancerous tissue has a less characteristic structure, and the A-scans are merely a loosely correlated decrease in intensity with increasing penetration depth. The authors calculated the frequency component for each A-scan and attempted to distinguish three types of tissues. This classification model was trained with data from three patients; data from a fourth patient were used for prediction. There were 4,015 A-scans available in total (1,666 adipose, 1,408 cancerous, and 941 stroma).

Adjacent results were clustered to determine the presence of A-scans which had been classified as cancerous within a B-scan. An interesting approach mentioned in this paper is the classification of sliding windows on an A-scan, thereby allowing classification of different sub-surface areas.

Predict treatment response

The publications mentioned so far have used a variety of characteristics or markers for classification. For the correct classification, some knowledge about a feature as a disease marker is necessary. Our group has demonstrated an approach that can discriminate between groups of surface normalized A-scans by LDA after reduction by PCA [51]. Most studies glance over the processing speed, and this is the only paper to mention a rate (60,000 A-scans per minute). If not sufficient as a stand-alone classification, such an algorithm seems suitable to run in addition to other classification routines. The approach has been applied to medical tissue samples; however, for medical images the algorithm requires manual exclusion of unsuitable scans, and classification seems to be heavily influenced by the variation within a healthy population [52]. Such a classification method may, however, have the potential to predict the response to laser treatment of port-wine stains, as suggested by a pilot study on seven images (7,000 A-scans), in near real-time [53].

Discussion

Common patterns and suggestions

There is a common pattern of five steps involved in extracting information from OCT data: obtaining the measurements, pre-processing, feature registration, data reduction, and classification of data.

Pre-processing is the essential part of automation in feature description and effectively increases robustness for further analysis: a common pattern is de-noising, feature recognition, normalization and refining. There are, principally, two approaches for evaluating OCT data: evaluating features of the A-scan, and applying image analysis tools to B-scans. Knowledge about the specimen surface is a common requirement. This can be achieved by the measurement protocol, e.g., by pressing a surface with a glass slide, or computationally. A frequently encountered method seems to be the crude cut-off threshold that is determined by the noise level above the specimen.

The ideal choice of pre-processing and analysis method depends on the sample type, the scan mode, and the actual research question asked. However, it is possible for one to draw a variety of conclusions from the automation algorithms considered in this review. All fully automated approaches rely on a combination of several steps. A-scan based analysis usually profits from the averaging of several A-scans in order to increase the SNR. Outlier A-scans can be corrected for by the use of a sliding median window or Olympic exclusion (removing a set of extreme values). Some publications disregard the first ten pixels or $\sim 100\ \mu\text{m}$ of a surface due to reflection artifacts [38,49]. This seems appropriate when the sub-surface data, rather than the exact surface location, are relevant. Textural features seem to show a potential for distinguishing abnormalities that are related to a vaguely progressing lack of structure, where sifting B-scans would become a straining task. Cross-correlation techniques need to employ several shapes to cover biological variation [36,45]. Several publications exclude unsuitable data manually. For proof of concept this is acceptable, but this carries the potential of introducing bias and should be based on an objective parameter, e.g., overall intensity or dynamic range of an image. Statistical methods therefore have a large potential for helping automated classification of OCT data.

Outlook

Automation is a way of helping an operator with image interpretation. Unfortunately, algorithms cannot be expected to achieve 100% accuracy. This should not be regarded as a major disadvantage: Even if data evaluation is only able to provide clear results from data with sufficient

diagnostically relevant contrast, resources can be channeled to concentrate on the more difficult cases. Furthermore, automation has the potential to provide an objective measure. A classification performance lower than 100% can still outperform human judgment, where, for some conditions, agreement can be considerably lower. Studies assessing the consensus of expert histopathologists, which represent the gold standard in tissue grading, have found kappa values (which illustrate the chance-corrected agreement) as low as 0.36 between eight experts for cervical biopsies [54], or 0.49 between three experts when classifying esophageal specimens [55].

Acquisition speed and resolution, which presented early challenges in OCT development, have experienced remarkable improvements. Limitations caused by data transfer rate, storage, and computing power are ceasing to be an issue: volumetric movies (4-D OCT) can become standard. This is a double-edged sword: with the possibility of obtaining vast data sets in near real-time, it becomes inherently impracticable for a human operator to evaluate these numbers of data, and it is virtually impossible to do so at the rate of data acquisition. OCT allows real-time data acquisition, and it would be a pity not to exploit the speed benefit of computational approaches to provide real-time diagnoses. In private conversations, we gained the impression that at least some OCT experts seem to have the opinion that the lack of real-time diagnosis is not a restriction. We would like to emphasize that—even if this opinion is not commonplace—we strongly feel that information should be available as quickly as possible. For instance, topical diffusion of contrast agents takes time, thereby ruining the advantage of having a real-time imaging modality. There are pitfalls for automation; some studies have found that assessment by a trained operator fairs better than the unsupervised approach [37, 38]. Algorithms do, however, have the potential to run in parallel with the scanning procedure and to provide an objective support for clinical decision making in real-time. This might even make the displaying of the OCT *image* redundant, much in the way spectroscopic techniques may show maps rather than raw spectral data. One field of interest today is therefore data evaluation in near real-time (Fig. 2).

After all, this manuscript would not have been possible without all the efforts towards automated classification of OCT data, which may be the next chapter in the success story of OCT.



Fig. 2 Scheme of landmark developments for OCT. *TD* time domain OCT, *FD* frequency domain OCT, *3D* volumetric OCT. The present challenge is automated data evaluation in near real-time

Acknowledgments This work was funded by the Pump Prime Fund of Cranfield University, the Gloucestershire Hospitals NHS Foundation Trust, and receives funding from the Technology Strategy Board [Grant/Project title OMICRON]. The authors would like to thank Professor Hugh Barr and Ms. Joanne Hutchings for kindly providing the esophageal sample for Fig. 1 (ethics approval by Gloucestershire Local Research Ethics Committee), and Dr. Catherine Kendall and Mr. Martin Isabelle for proofreading this manuscript. Dr. Nicholas Stone holds a Career Scientist Fellowship, which is funded by the UK National Institute of Health Research.

References

- Huang D, Swanson EA, Lin CP, Schuman JS, Stinson WG, Chang W et al (1991) Optical coherence tomography. *Science* 254:1178–1181 doi:10.1126/science.1957169
- Tomlins PH, Wang RK (2005) Theory, developments and applications of optical coherence tomography. *J Phys D Appl Phys* 38:2519–2535 doi:10.1088/0022-3727/38/15/002
- Ellis DI, Goodacre R (2006) Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and Raman spectroscopy. *Analyst (Lond)* 131:875–885 doi:10.1039/b602376m
- Stelzer EHK (1998) Contrast, resolution, pixelation, dynamic range and signal-to-noise ratio: fundamental limits to resolution in fluorescence light microscopy. *J Microsc* 189:15–24 doi:10.1046/j.1365-2818.1998.00290.x
- Patil CA, Bosschaart N, Keller MD, van Leeuwen TG, Mahadevan-Jansen A (2007) Combined Raman spectroscopy and optical coherence tomography device for tissue characterization. *Opt Lett* 33:1135–1137 doi:10.1364/OL.33.001135
- Kanter E, Walker R, Marion S, Hoyer P, Barton JK (2005) Optical coherence tomography imaging and fluorescence spectroscopy of a novel rat model of ovarian cancer. *Prog Biomed Opt Imaging* 6:58610P.1-58610P.8
- Yang C (2005) Molecular contrast optical coherence tomography: a review. *Photochem Photobiol* 81:215–237 doi:10.1562/2004-08-06-IR-266.1
- Xu C, Ye J, Marks DL, Boppart SA (2004) Near-infrared dyes as contrast-enhancing agents for spectroscopic optical coherence tomography. *Opt Lett* 29:1647–1649 doi:10.1364/OL.29.001647
- Xi C, Marks DL, Parikh DS, Raskin L, Boppart SA (2004) Structural and functional imaging of 3D microfluidic mixers using optical coherence tomography. *Proc Natl Acad Sci U S A* 101:7516–7521 doi:10.1073/pnas.0402433101
- Bonesi M, Churnakov DY, Ritchie LJ, Meglinski IV (2006) Turbulence monitoring with Doppler optical coherence tomography. *Laser Phys Lett* 4:304–307 doi:10.1002/lapl.200610098
- Park B, Pierce M, Cense B, de Boer J (2003) Real-time multifunctional optical coherence tomography. *Opt Express* 11:782–793
- Su J, Tomov IV, Jiang Y, Chen Z (2007) High-resolution frequency-domain second-harmonic optical coherence tomography. *Appl Opt* 46:1770–1775 doi:10.1364/AO.46.001770
- Lazebnik M, Marks DL, Potgieter K, Gillette R, Boppart SA (2003) Functional optical coherence tomography for detecting neural activity through scattering changes. *Opt Lett* 28:1218–1220 doi:10.1364/OL.28.001218
- Maheswari RU, Takaoka H, Kadono H, Homma R, Tanifuji M (2003) Novel functional imaging technique from brain surface with optical coherence tomography enabling visualization of depth resolved functional structure in vivo. *J Neurosci Methods* 124:83–92 doi:10.1016/S0165-0270(02)00370-9
- Morgner U, Drexler W, Kärtner FX, Li XD, Pitris C, Ippen EP et al (2000) Spectroscopic optical coherence tomography. *Opt Lett* 25:111–113 doi:10.1364/OL.25.000111
- Faber DJ, Mik EG, Aalders MC, van Leeuwen TG (2005) Toward assessment of blood oxygen saturation by spectroscopic optical coherence tomography. *Opt Lett* 30:1015–1017 doi:10.1364/OL.30.001015
- Knüttel A, Boehlau-Godau M (2000) Spatially confined and temporally resolved refractive index and scattering evaluation in human skin performed with optical coherence tomography. *J Biomed Opt* 5:83–92 doi:10.1117/1.429972
- Kholodnykh AI, Petrova IY, Larin KV, Motamedi M, Esenaliev RO (2003) Precision of measurement of tissue optical properties with optical coherence tomography. *Appl Opt* 42:3027–3037 doi:10.1364/AO.42.003027
- Esenaliev RO, Larin KV, Larina IV, Motamedi M (2001) Noninvasive monitoring of glucose concentration with optical coherence tomography. *Opt Lett* 26:992–994 doi:10.1364/OL.26.000992
- van der Meer FJ, Faber DJ, Sassoon DMB, Aalders MC, Pasterkamp G, van Leeuwen TG (2005) Localized measurement of optical attenuation coefficients of atherosclerotic plaque constituents by quantitative optical coherence tomography. *IEEE Trans Med Imaging* 24:1369–1376 doi:10.1109/TMI.2005.854297
- Jeon SW, Shure MA, Baker KB, Huang D, Rollins AM, Chahlavi A et al (2006) A feasibility study of optical coherence tomography for guiding deep brain probes. *J Neurosci Methods* 154:96–101 doi:10.1016/j.jneumeth.2005.12.008
- Ramrath L, Hofmann UG, Huettmann G, Moser A, Schweikard A (2007) Towards automated OCT-based identification of white brain matter. In: *Bildverarbeitung für die Medizin—Algorithmen—Systeme—Anwendungen*. Springer, Berlin Heidelberg, pp 414–418. ISBN 978-3-540-71091-2
- Turchin IV, Sergeeva EA, Dolin LS, Kamensky VA, Shakhova NM, Richards-Kortum R (2005) Novel algorithm of processing optical coherence tomography images for differentiation of biological tissue pathologies. *J Biomed Opt* 10:064024 doi:10.1117/1.2137670
- Kuranov RV, Sapozhnikova VV, Prough DS, Ciceaite I, Esenaliev RO (2006) In vivo study of glucose-induced changes in skin properties assessed with optical coherence tomography. *Phys Med Biol* 51:3885–3900 doi:10.1088/0031-9155/51/16/001
- Levitz D, Thrane L, Frosz MH, Andersen PE, Andersen CB, Valanciunaite J et al (2004) Determination of optical scattering properties of highly-scattering media in optical coherence tomography images. *Opt Express* 12:249–259 doi:10.1364/OPEX.12.000249
- Kaiser JF (1990) On a simple algorithm to calculate the ‘energy’ of a signal. *Proc IEEE Int Conf Acoust Speech Signal Process (ICASSP’90)* 1:381–384
- Gossage KW, Tkaczyk TS, Rodriguez JJ, Barton JK (2003) Texture analysis of optical coherence tomography images: feasibility for tissue classification. *J Biomed Opt* 8:570–575 doi:10.1117/1.1577575
- Hillman TR, Adie SG, Seemann V, Armstrong JJ, Jacques SL, Sampson DD (2006) Correlation of static speckle with sample properties in optical coherence tomography. *Opt Lett* 31:190–192 doi:10.1364/OL.31.000190
- Brodatz P Textures. (Dover Publications Inc., ISBN: 0-486-40699-7, 2000)
- Gossage KW, Smith CM, Kanter EM, Hariri LP, Stone AL, Rodriguez JJ et al (2006) Texture analysis of speckle in optical coherence tomography images of tissue phantoms. *Phys Med Biol* 51:1563–1575 doi:10.1088/0031-9155/51/6/014
- MacNeill BD, Jang I-K, Bouma BE, Ifimia N, Takano M, Yabushita H et al (2004) Focal and multi-focal plaque macrophage distributions in patients with acute and stable presentations of coronary artery disease. *J Am Coll Cardiol* 44:972–979 doi:10.1016/j.jacc.2004.05.066

32. Qi X, Sivak MV Jr, Isenberg G, Willis JE, Rollins AM (2006) Computer-aided diagnosis of dysplasia in Barrett's esophagus using endoscopic optical coherence tomography. *J Biomed Opt* 11:044010 doi:[10.1117/1.2337314](https://doi.org/10.1117/1.2337314)
33. Chen Y, Aguirre AD, Hsiung P-L, Huang S-W, Mashimo H, Schmitt JM et al (2008) Effects of axial resolution improvement on optical coherence tomography (OCT) imaging of gastrointestinal tissues. *Opt Express* 16:2469–2485 doi:[10.1364/OE.16.002469](https://doi.org/10.1364/OE.16.002469)
34. Lingley-Papadopoulos CA, Loew MH, Manyak MJ, Zara JM (2008) Computer recognition of cancer in the urinary bladder using optical coherence tomography and texture analysis. *J Biomed Opt* 13:024003 doi:[10.1117/1.2904987](https://doi.org/10.1117/1.2904987)
35. Rogowska J, Bryant CM, Brezinski ME (2003) Cartilage thickness measurements from optical coherence tomography. *J Opt Soc Am A* 20:357–367 doi:[10.1364/JOSAA.20.000357](https://doi.org/10.1364/JOSAA.20.000357)
36. Weissman J, Hancewicz T, Kaplan P (2004) Optical coherence tomography of skin for measurement of epidermal thickness by shapelet-based image analysis. *Opt Express* 12:5760–5769 doi:[10.1364/OPEX.12.005760](https://doi.org/10.1364/OPEX.12.005760)
37. Gambichler T, Moussa G, Regeniter P, Kasseck C, Hofmann MR, Bechara FG et al (2007) Validation of optical coherence tomography in vivo using cryostat histology. *Phys Med Biol* 52:85
38. Korde VR, Bonnema GT, Xu W, Krishnamurthy C, Ranger-Moore J, Saboda K et al (2007) Using optical coherence tomography to evaluate skin sun damage and precancer. *Lasers Surg Med* 39:687–695 doi:[10.1002/lsm.20573](https://doi.org/10.1002/lsm.20573)
39. Lee Y-K, Rhodes WT (1990) Nonlinear image processing by a rotating kernel transformation. *Opt Lett* 15:1383–1385
40. Rogowska J, Brezinski ME (2002) Image processing techniques for noise removal, enhancement and segmentation of cartilage OCT images. *Phys Med Biol* 47:641–655 doi:[10.1088/0031-9155/47/4/307](https://doi.org/10.1088/0031-9155/47/4/307)
41. Fernández DC, Salinas HM, Puliafito CA (2005) Automated detection of retinal layer structures on optical coherence tomography images. *Opt Express* 13:10200–10216 doi:[10.1364/OPEX.13.010200](https://doi.org/10.1364/OPEX.13.010200)
42. Hori Y, Yasuno Y, Sakai S, Matsumoto M, Sugawara T, Madjarova V et al (2006) Automatic characterization and segmentation of human skin using three-dimensional optical coherence tomography. *Opt Express* 14:1862–1877 doi:[10.1364/OE.14.001862](https://doi.org/10.1364/OE.14.001862)
43. Danielsson PE (1980) Euclidean distance mapping. *Comput Graph Image Process* 14:227–248 doi:[10.1016/0146-664X\(80\)90054-4](https://doi.org/10.1016/0146-664X(80)90054-4)
44. Bonnema G, Cardinal K, Williams S, Barton J (2008) An automatic algorithm for detecting stent endothelialization from volumetric optical coherence tomography datasets. *Phys Med Biol* 53:3083–3098 doi:[10.1088/0031-9155/53/12/001](https://doi.org/10.1088/0031-9155/53/12/001)
45. Cheng Y, Larin KV (2006) Artificial fingerprint recognition by using optical coherence tomography with autocorrelation analysis. *Appl Opt* 45:9238–9245 doi:[10.1364/AO.45.009238](https://doi.org/10.1364/AO.45.009238)
46. Huang ML, Chen HY (2005) Development and comparison of automated classifiers for glaucoma diagnosis using Stratus optical coherence tomography. *Invest Ophthalmol Vis Sci* 46:4121–4129 doi:[10.1167/iovs.05-0069](https://doi.org/10.1167/iovs.05-0069)
47. Jørgenson TM, Tycho A, Mogensen M, Bjerring P, Jemec GBE (2008) Machine-learning classification of non-melanoma skin cancers from image features obtained by optical coherence tomography. *Skin Res Technol* 14:364–369 doi:[10.1111/j.1600-0846.2008.00304.x](https://doi.org/10.1111/j.1600-0846.2008.00304.x)
48. Iftimia NV, Bouma BE, Pitman MB, Goldberg B, Bressner J, Tearney GJ (2005) A portable, low coherence interferometry based instrument for fine needle aspiration biopsy guidance. *Rev Sci Instrum* 76:064301 doi:[10.1063/1.1921509](https://doi.org/10.1063/1.1921509)
49. Goldberg BD, Iftimia NV, Bressner JE, Pitman MB, Halpern E, Bouma BE et al (2008) Automated algorithm for differentiation of human breast tissue using low coherence interferometry for fine needle aspiration biopsy guidance. *J Biomed Opt* 13:014014 doi:[10.1117/1.2837433](https://doi.org/10.1117/1.2837433)
50. Zysk AM, Boppart SA (2006) Computational methods for analysis of human breast tumor tissue in optical coherence tomography images. *J Biomed Opt* 11:054015 doi:[10.1117/1.2358964](https://doi.org/10.1117/1.2358964)
51. Bazant-Hegemark F, Stone N (2008) Near real-time classification of optical coherence tomography data using principal components fed linear discriminant analysis. *J Biomed Opt* 13:034002 doi:[10.1117/1.2931079](https://doi.org/10.1117/1.2931079)
52. Bazant-Hegemark F, Stone N, Read MD, McCarthy K, Wang RK (2007) Optical coherence tomography (OCT) imaging and computer aided diagnosis of human cervical tissue specimens. *Proc SPIE* 6627:66270F doi:[10.1117/12.728366](https://doi.org/10.1117/12.728366)
53. Bazant-Hegemark F, Meglinski I, Kandamany N, Monk B, Stone N, (2008) Optical coherence tomography: a potential tool for unsupervised prediction of treatment response for port-wine stains. *Photodiagn Photodyn Ther* (in press) doi:[10.1016/j.pdpdt.2008.09.001](https://doi.org/10.1016/j.pdpdt.2008.09.001)
54. Ismail SM, Colclough AB, Dinnen JS, Eakins D, Evans DMD, Gradwell E et al (1989) Observer variation in histopathological diagnosis and grading of cervical intraepithelial neoplasia. *BMJ* 298:707–710
55. Kendall C, Stone N, Shepherd N, Geboes K, Warren B, Bennett R et al (2003) Raman spectroscopy, a potential tool for the objective identification and classification of neoplasia in Barrett's oesophagus. *J Pathol* 200:602–609 doi:[10.1002/path.1376](https://doi.org/10.1002/path.1376)

Near real-time classification of optical coherence tomography data using principal components fed linear discriminant analysis

Florian Bazant-Hegemark
Nicholas Stone

Cranfield University at Silsoe
Cranfield Health
Bedfordshire MK45 4DT
United Kingdom

and
Gloucestershire Royal Hospital
Biophotonics Research Group
Great Western Road
Gloucester GL1 3NN
United Kingdom

E-mail: n.stone@medical-research-centre.com

Abstract. An optical coherence tomography (OCT) prediction algorithm is designed and tested on a data set of sample images (taken from vegetables and porcine tissues) to demonstrate proof of concept. Preprocessing and classification of data are fully automated, at a rate of 60,000 A-scans/min on a standard computer and can be considered to deliver in near real-time. A data set consisting of nine groups was classified correctly in 82% of cases after cross-validation. Sets of fewer groups reach higher rates. The algorithm is able to distinguish groups with strong visual similarity among several groups of varying resemblance. Surface recognition and normalizing to the surface are essential for this approach. The mean divided by the standard deviation is a suitable descriptor for reducing a set of surface normalized A-scans. The method enables grouping of separate A-scans and is therefore straightforward to apply on 3-D data. OCT data can reliably be classified using principal component analysis combined with linear discriminant analysis. It remains to be shown whether this algorithm fails in the clinical setting, where interpatient variation can be greater than the deviations that are investigated as a disease marker.
© 2008 Society of Photo-Optical Instrumentation Engineers. [DOI: 10.1117/1.2931079]

Keywords: classification model; optical coherence tomography; automated discrimination; medical imaging; preprocessing; image processing.

Paper 07241RR received Jul. 5, 2007; revised manuscript received Jan. 8, 2008; accepted for publication Jan. 15, 2008; published online May 22, 2008.

1 Introduction

1.1 Optical Coherence Tomography

Optical coherence tomography (OCT) is a relatively novel imaging modality first described by Huang et al.,¹ which is becoming established as a technique for noninvasive, real-time subsurface imaging at resolutions of 2 to 10 μm . Non-destructive probing of this type is of interest specifically in medical imaging^{2,3} and also in other fields such as art conservation,^{4,5} quality assurance, or homeland security (e.g., finger prints). Compared with other emerging imaging modalities, OCT provides high resolution and imaging speed for detailed structural information—in vast data sets. We investigate automated classification, which is quicker than visually assessing images and likely to be more reproducible.

1.1.1 Principal component analysis and linear discriminant analysis

Principal component analysis (PCA) and linear discriminant analysis (LDA) are mathematical methods that, on a statistical basis, enable data reduction and clustering into groups.⁶ There are papers on applying PCA for OCT related analysis. Weissman et al.⁷ used PCA for comparing the agreement of their surface recognition algorithm with the line drawn by human operators. Qi et al.⁸ used PCA for preprocessing. They were reducing six statistical measures for texture features to two

principal components. However, using this approach they reported no benefit over using the 6-D data.

Huang and Chen⁹ also applied PCA and LDA for classification, but not on the OCT image itself. Rather, they were using a set of 25 morphological parameters to characterize ophthalmologic pathologies. These parameters were thickness and size measures that could be extracted from the OCT image. In nonophthalmic OCT imaging, such defined layers are usually not observed.

Zysk and Boppart¹⁰ reported three methods for classifying three relevant groups (adipose, cancerous, stroma) in A-scans of breast cancer biopsies. The first was taking the Fourier transform (FT) of an A-scan and using its difference from a reference (the mean of a training set) for the analysis. The second was periodicity analysis, which looks at the distance between high-intensity subsurface peaks to the surface peak. The third was a combination of both. The periodicity analysis seems to benefit from the “bubbly” structural pattern in OCT images of adipose breast tissue, which, however, is not present in every type of tissue. Part of this method, taking the FT of an A-scan, we compare with our method. (We did not implement the periodicity technique.) This enabled us to test the Fourier domain technique on sets containing more than three groups.

1083-3668/2008/13(3)/034002/8/\$25.00 © 2008 SPIE

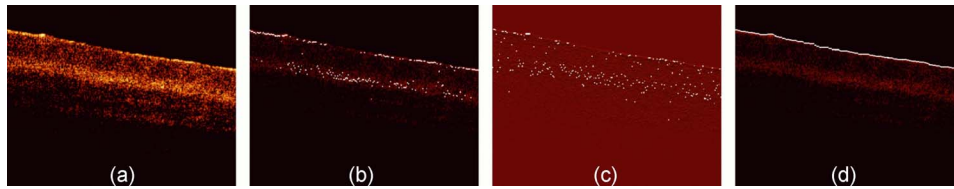


Fig. 1 (a) Denoised B-scan of a biological sample (porcine esophagus), (b) white dots mark the point of highest intensity within one A-scan (overlay over dimmed intensity image), (c) white dots mark maximum rate of change in intensity within one A-scan, and (d) surface pixels as recognised from a binary threshold mask. (b) and (c) show that the maximum intensity and derivative do not necessarily correlate with the surface.

1.1.2 Surface recognition

For OCT there is no established method for unsupervised surface recognition. This is, however, not straightforward as the point with the highest intensity within one A-scan is not necessarily correlated with the sample surface. This is also true for taking the first derivative of the intensity; the largest change in intensity is not necessarily at the surface. Were this the case, surface recognition could be done quickly and elegantly using a weighing of the magnitude and the derivative of the signal intensity over the single A-scan. Specifically for biological samples, this is prone to errors (cf. Fig. 1).

There is a small amount of literature about surface recognition in OCT and research groups seem to have methods that seem fit for the purpose of their publication, but are not always useful to implement for a different specimen type. We found a variety of reports, e.g., shapelet-based boundary recognition;⁷ rotating kernel transformation, which emphasizes straight lines within an image;¹¹ erosion/dilation techniques on a binary threshold image;⁸ or polynomial fitting through median-blurred intensity peaks.¹²

We chose binary thresholding because it is accurate and fast. A visualization of the full process is shown in Fig. 2.

1.2 Research Questions

In this paper, we investigate whether OCT data can be classified in real time on the basis of A-scans without any input from a human operator. We investigate the classification performance on a small demonstrator group, and plan to assess the impact of surface normalization. We further investigate whether representative measures can be established for larger data sets, which strain computer resources. Finally, we investigate the performance on animal tissue.

2 Methods

2.1 Measurements

2.1.1 OCT device

The OCT device used in this study was a time-domain bench-top system, which was similar in design to the system described by Yang et al.¹³ The light source is a superluminescent diode (SLD, Superlum) with a central wavelength of 1310 nm that provided an axial resolution of $\sim 15 \mu\text{m}$ in air and a $10\text{-}\mu\text{m}$ lateral resolution. The rapid delay line is realized using a double-pass mirror and a reference mirror mounted on a galvanometer (Cambridge Instruments). The data acquisition was performed using LabVIEW.

2.1.2 Samples

As a demonstrator of proof of concept, fruit and vegetables served as readily available samples with varying structure (Fig. 3). Ten scans were taken of each vegetable; three vegetables of one kind form a group. In addition to this, fresh porcine specimens were obtained from a local butcher. Twenty different samples were prepared from each investigated organ, although all specimens originate from one animal. The size and shape of the samples enabled them to be placed in petri dishes for the measurements. Where necessary, samples were kept stable with needles.

2.2 Preprocessing

2.2.1 Image

All images used in this study measure 300×500 pixels, covering a 3-mm width and an approximately 5-mm depth. The log of the intensity data (time domain signal) was stored as a matrix. Values were handled in double precision format. For the automated noise reduction, a ribbon of "air," intensity in-

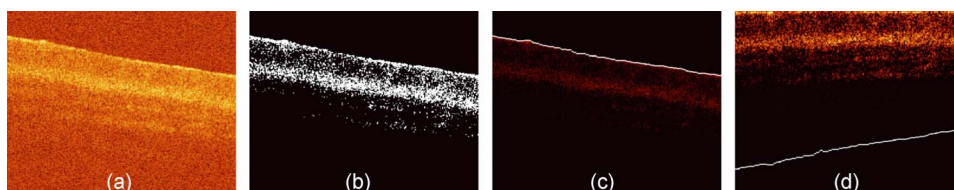


Fig. 2 (a) Original noisy OCT image (false color log of intensity envelope data), representative for many *ex vivo* scans, where the sample surface, although somewhat straight, is not exactly parallel to the scanning direction; and (b) binary image using a 0.90 quantile threshold; (c) surface recognized from (b); and (d) image after wrapping.

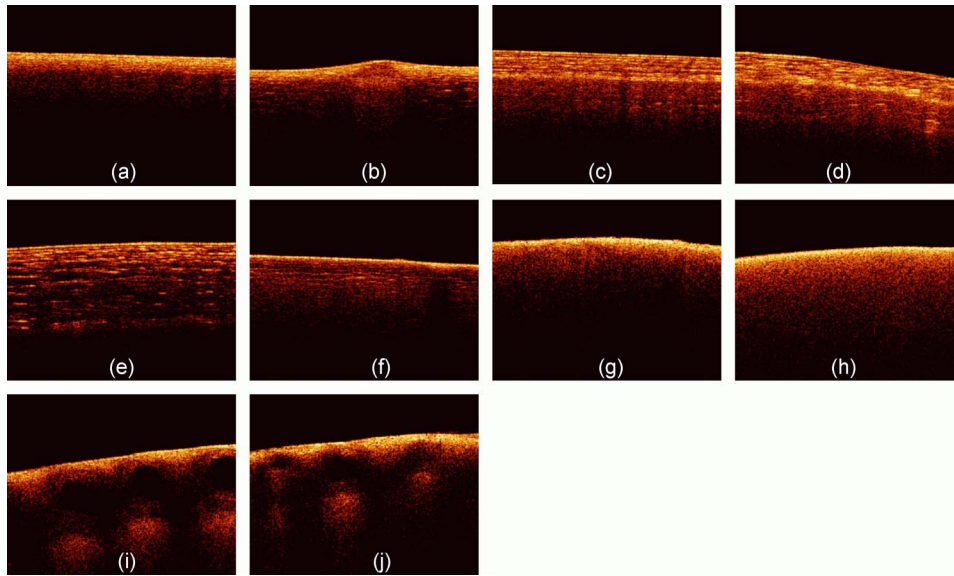


Fig. 3 Representative OCT images of vegetables (after denoising, to be fed into surface recognition): (a) celery, longitudinal; (b) Celery, axial; (c) leek, longitudinal; (d) leek, axial; (e) onion; (f) potato; (g) mushroom stem; (h) mushroom top; (i) lemon; and (j) lime. All B-scans in this paper have a width of 3 mm. Note the somewhat similar appearance of (c) and (d), (g) and (h), and (i) and (j).

formation above the surface, was regarded as background noise level.¹⁴ The mean of this plus $7 \times$ the standard deviation of this background signal was used as the intensity threshold. This relatively high threshold does crudely cut off low-level signals from deeper areas. We found it, however, to be more important for our subsequent surface recognition to have a noise-free image, with the surface not being obscured by artefacts or layers.

The scanning protocol therefore required an area of air above the surface. We aimed to cover approximately 1/5 to 1/3 of the image height, as this was convenient for manual handling. The surface of the sample was positioned to ensure that the full penetration depth could be depicted on the image. Specimens were placed so that the beam angle would not cause excessive reflections. While these limitations for the distance of the probe to the sample were chosen with respect to easy handling, no measurement was taken from the same site, as that would spoil the validity of the subsequent correlation.

After thresholding, the intensity values of one B-scan were normalized to the maximum intensity peak, by scaling to the maximum peak being unity.

2.2.2 Surface recognition and normalizing

Detecting the surface based on a binary threshold mask of the image, as shown in Fig. 2, is an approach we found to be robust for the purpose of our preprocessing. After finding no substantial influence on the result for threshold levels between 0.85 and 0.975, we arbitrarily defined a 0.95 quantile threshold for creating a binary image. The first pixel of this mask was taken as the surface. No smoothing or correction of eventual outliers (e.g., reflection artefacts) was performed. Normalizing to the surface was achieved by shifting each A-scan according to the surface vector.

Of these images, we calculated several representative measures of the depth profile of the intensity variation within one

image: the mean B-scan, median B-scan, standard deviation, or the mean divided by the standard deviation¹⁵ (M/SD). Furthermore, we looked into combinations of these, e.g., M/SD and mean as two representative measures for one B-scan. We also used the FT of an A-scan to be able to roughly compare our results with the approach of Zysk and Boppart.¹⁰

2.3 Data Analysis

Data handling, preprocessing, PCA, and LDA were performed using MATLAB V6.1.0.450, Release 12.1 (The MathWorks, Inc.), extended by the “PLS_Toolbox” (Eigenvector Research, Inc.) and our in-house tools (“Classification_Toolbox”). Emphasis was put on speed, and therefore execution times were assessed using the tic-toc functions and the “profile” function in MATLAB. Bottlenecks, e.g., the routine to normalize A-scans to the surface, were vectorized.

We did not perform mean centering of the data as we found this to have no effect on the quality of results (data not shown). Using PCA, the data was reduced to the first 25 principal components (PCs), which were then sorted in order of their statistical significance, characterized by their respective F values determined by an analysis of variance¹⁵ (ANOVA) ($p < 0.001$). [The number of PCs (25) was chosen arbitrarily and is not fundamental to the method. The number of PCs does affect the results (data not shown), but for the purpose of this paper, whereby different approaches are studied, a constant of 25 PCs is a fair representation.]

These 25 PCs were then used in their entirety for LDA, which is inherently able to suppress loadings of lower significance for the model. The training model was tested using leave-one-out cross-validation. For a small group (30 scans), one measure was left out: one A-scan when classifying separate A-scans, or one representative measure when classifying the representative measures (leave-on-scan-out, LOSO). While for the LOSO cross-validation, each prediction data set is “unknown” to the algorithm, the training model can still

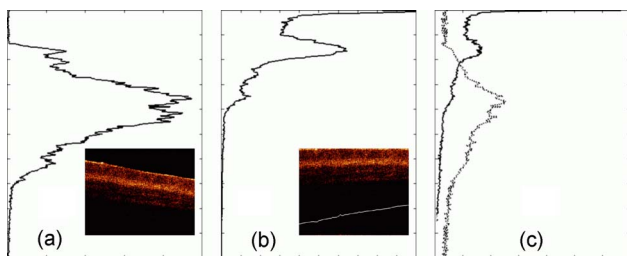


Fig. 4 (a) Mean A-scan of original data, (b) mean A-scan of surface normalized image, and (c) M/SD along row of A-scans: (—) surface normalized and (- - -) original.

include data from the very same sample. A better cross-validation approach is therefore to test a classification model with a prediction set whose samples are “unknown” in their entirety. This is achieved by splitting the data into a training group and a prediction group (leave-one-group-out, LOGO). For the large sample groups (330 scans) we created a training model from 220 randomly chosen scans, and evaluated the prediction performance for the remaining 110 scans, which are unknown to the test model. These 110 scans were from different samples than the 220 training scans. In this way, cross-validation could be performed on three different combinations. The mean of the results for the three cross-validations for one set of samples we accept as classification performance. For comparing pairs of similar vegetables and the animal samples, a prediction model was created using a randomly assigned half of the B-scans of each group and cross-validated using the other half of the B-scans. Also here, test and prediction groups were from different samples, and the mean of the two results was taken as overall performance.

3 Results

The images appear typical of common OCT image quality. Representatives of fruit and vegetable structures are shown in Fig. 3. Distinguishing such images, when they are presented next to each other, is fairly easy. However, when working

through a large quantity, some of these could start to look similar to an insufficiently trained or overworked eye.

The results section is divided into two groups. Part 1 shows the effects of preprocessing and demonstrates the rationale for choosing our parameters. Part 2 shows the actual classification results.

3.1 Part 1: Preprocessing

3.1.1 Surface recognition and normalizing

Vectorization reduced the processing time roughly by a factor of 4 to 5. The visual results of the unsupervised surface normalizing were effectively shown in Fig. 2(d). Figure 4 shows how normalizing to the surface affects the mean and the M/SD of a B-scan.

3.1.2 Vegetables—separate A-scans

In a first approach, we used 10 images of three vegetables, resulting in 30 images or 9000 A-scans. Performing PCA on the surface normalized A-scans [Fig. 5(A)] or nonsurface normalized FT curves [Fig. 5(B)] show a tendency to group, but also considerable overlap. Note that the plots in Fig. 5 show only the two most significant (ANOVA) PCs for classification against each other. The subsequent LDA separates on the basis of a pool of several PCs (25 in the examples of this paper).

Table 1 shows the performance of classification models for this case. For the amplitude of the frequency content, surface normalization does not improve the classification. For classifying time domain data (spatial information), surface normalization yields a higher rate.

3.1.3 Data reduction prior to PCA

The classification rates in Table 1 are not disappointing, but take a long time to calculate. Hence we reduced the B-scan to a representative measure, as shown in Fig. 6. From the mean of these images, the separation is clearer than in Fig. 5. For this approach, surface normalization has a stronger impact

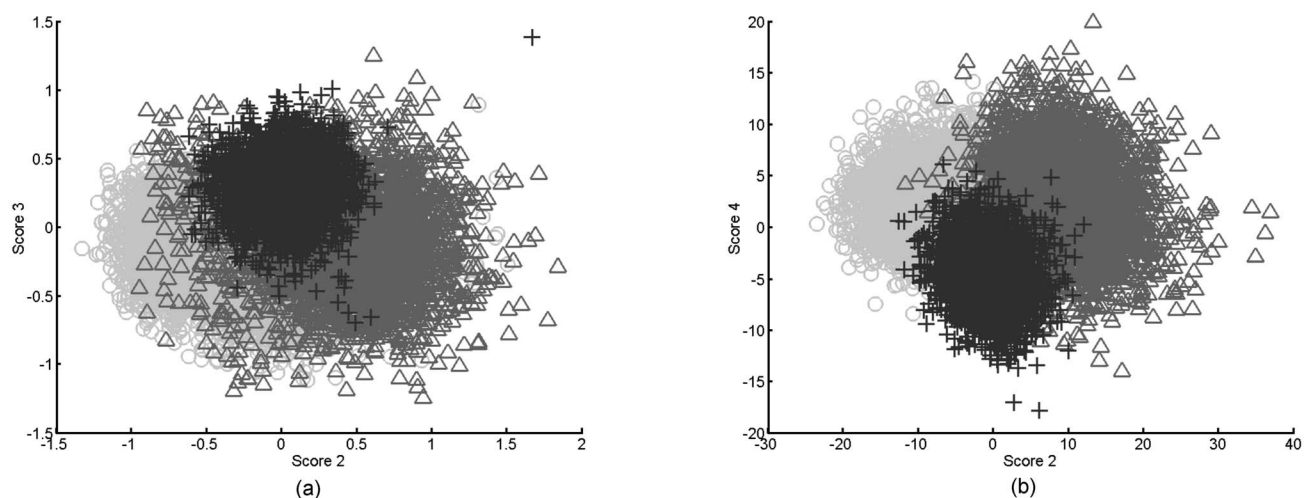


Fig. 5 PCA of three vegetables, 10 images each. The plots show the PCs of 9000 A-scans with the two highest scores (determined by ANOVA) out of 25: (a) surface normalized and (b) FT of A-scans (no surface normalization). Legend: ○=celery; △=onion; +=mushroom.

Table 1 Comparison for classification of three A-scan representations: the FT, the amplitude of the FT, and time domain data.

Preprocessing	FT	Amplitude of FT	Depth profile (A-scan)
Without surface normalization	65.4	96.1	65.4
Surface normalization	92.8	95.8	92.8

The data set consisted of three groups of 10 images each (9000 A-scans). Scores are percentages correctly classified by the algorithm, after LOSO cross-validation.

than before: classification was correct in 93.3% for the surface-normalized data [Fig. 6(A)], but only 33.3% for not-surface-normalized data [Fig. 6(B)].

3.1.4 Data reduction—large dataset

A large data set consisting of 330 images, categorized into nine groups, was used to assess various representative measures. These 99,000 A-scans could not be fed into analysis on a Pentium type processor with 1 Gbyte of RAM, so it was essential to extract a representative measure. Table 2 shows the results for data reduction. For time domain data, models created using the mean normalized by the standard deviation (M/SD) had the best results, as shown in Table 2. For models from the FT of an A-scan, the standard deviation (SD) has the highest classification score. After LOGO cross-validation, the model constructed from the mean values has a higher score.

3.2 Part 2: Classification

From the results so far, combinations with high scores were chosen for further comparison: surface normalization with M/SD, and the mean and the SD of the amplitude of the FT. We abbreviate these as SN-MSD, AFT-M, and AFT-SD, respectively.

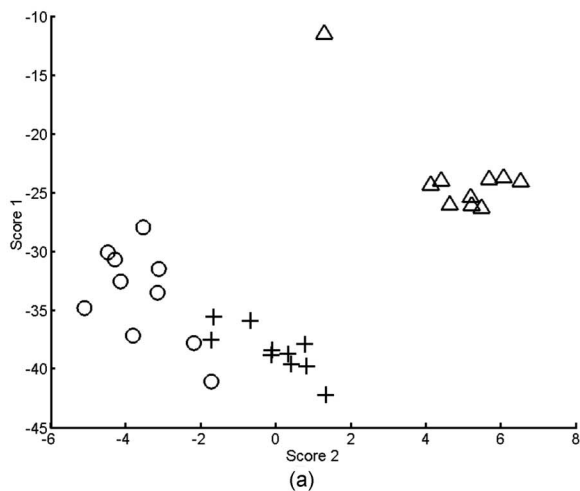


Table 2 Training performance (TP) and cross-validation (CV) of a data set consisting of nine groups; seven groups of 30 images each and two groups of 60 images each (330 representative measures).

Representative Measure	A-Scan ^a			FT of A-Scan ^b		
	TP ^c	LOSO ^d	LOGO ^e	TP ^c	LOSO ^d	LOGO ^e
Median	86.7	80.0	81.6	88.8	83.3	82.7
Mean	88.2	84.2	81.6	90.0	86.4	86.7
SD	86.4	77.6	72.9	91.2	85.5	85.1
M/SD	92.1	86.1	82.0	90.0	85.5	81.8

Scores are percentages correctly classified by the algorithm, before and after cross-validation. The highest scores are bold.

^aSurface-normalized time domain intensity data, ^bamplitude of FT (nonsurface-normalized) A-scan, ^ctraining performance, ^dleave-one-scan-out cross-validation, ^eleave-one-group-out cross-validation.

3.2.1 Vegetables

A different set consisting of 330 images, categorized into seven groups, was classified correctly (after LOGO cross-validation) as follows: SN-MSD, 91.8%; AFT-M, 93.9%; and AFT-SD, 79.1%. For comparison, the non-surface-normalized approach reached 32.0% in this case.

3.2.2 Similar image types

In the next step, we assessed the classification of similar images, i.e., measurements of one type of vegetable with different orientation, or similar images (lemon versus lime). Representative B-scans are shown in Figs. 3(G)–3(J). Classification results for 30 images each are shown in Table 3.

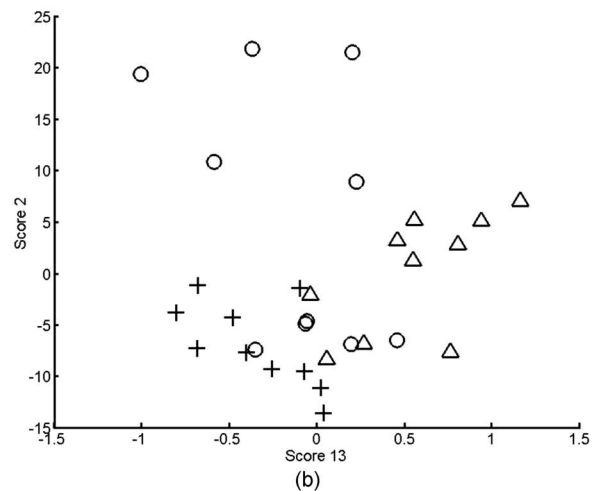


Fig. 6 PCA of three vegetables, 10 images each. The data points show the two highest scoring PCs (ANOVA) of the 30 representative measures: (a) surface normalized and (b) no surface normalisation. ◦=celery; △=onion; +=mushroom.

Table 3 Classification rates after LOGO cross-validation; training models were created for pairs of similar images (2×30 images for a pair).

Type of Similar Images	Representative Measure (%)		
	SN-MSD	AFT-M	AFT-SD
Celery (two orientations)	56.7	68.3	86.7
Leek (two orientations)	61.7	63.3	80.0
Mushroom (two orientations)	96.7	90.0	90.0
Lemon versus Lime	83.3	95.0	95.0

3.2.3 Similar images in a large group

If groups are similar, further added groups do decrease the power of the separation. To show this, we grouped two vegetables to yield four groups. Separation is affected when these two groups must be distinguished not only from each other, but also among other data, as shown in Table 4.

3.2.4 Preliminary tissue pilot study

Extending on the vegetable findings, we used porcine samples for a preliminary tissue pilot study. We measured 20 sites each of four different tissue types (tongue, esophagus, tracheae, and bronchioles). Representative images are shown in Fig. 7. The results for PC-fed LDA classification are encouraging. After LOGO cross-validation, the four, albeit clearly distinct tissue types, are classified 90.0% (SN-MSD), 82.5% (AFT-M), and 70.0% (AFT-SD) correctly.

To see whether more subtle changes can be picked up, we looked at samples from one tissue type (esophageal) measuring them in a fresh state, after degradation for 5 days, and after application of 5% acetic acid, a contrast agent that is used in clinical settings. These three groups can be separated well between each other in 91.7% (SN-M/SD), 93.3% (AFT-M), and 76.7% (AFT-SD) of the cases (LOGO cross-validated).

4 Discussion

OCT provides structural images that are intuitive to look at and this might even be an important aside for this technology to become accepted by many users. Evaluation of such structural data seems straightforward; however, it remains a challenge in detailed areas. This is evident for an imaging technology that enables one to look closely at areas that previously have been impossible to observe.

Table 4 Classification rates after LOGO cross-validation; training models were created for sets of pairs of similar images (2×30 images for a pair).

Type of Similar Images	Representative Measure (%)		
	SN-MSD	AFT-M	AFT-SD
Two mushroom orientations, lemon versus lime	90.8	91.7	86.7
All four pairs (eight groups) from Table 3	67.1	74.2	82.1

Acquisition speed enables generation of high-resolution volumetric real-time data; this is difficult to handle and display meaningfully and will remain complex to interpret in the future. Searching through a vast voxel space can hardly be imagined to be an enjoyable task for a human operator.

4.1 Preprocessing

4.1.1 Time domain data and frequency content

Classification requires preprocessing. Both methods, the amplitude of the FT of an A-scan and the surface normalized A-scan, give good results. The FT we used to create a model differed in four points from that of Zysk and Boppart.¹⁰ First, our denoising and normalizing, as already described, was different. Second, we were using the FT of the A-scan for classification, and not the difference from a reference. Third, Zysk and Boppart report that they did perform surface normalizing (“truncating”) prior to FT, whereas we found that to be beneficial only when feeding the real and the imaginary part of the FT. However, we achieved better results when using the amplitude of the FT directly of the non-surface-normalized A-scan. Finally, we took a representative measure for a set of A-scans and classified the set, rather than classifying single A-scans.

It then probably depends on the sample structure, e.g., the presence of distinctive periodic patterns, whether the frequency information or time domain data enable a better classification. Time domain data must be surface normalized. This requires a robust and quick surface recognition and means an additional computational step.

We are aware that the data for this paper are suitable for binary thresholding as surface recognition, and that the converse is not necessarily always the case. Such scenarios could be imagined where the surface of interest is covered by a

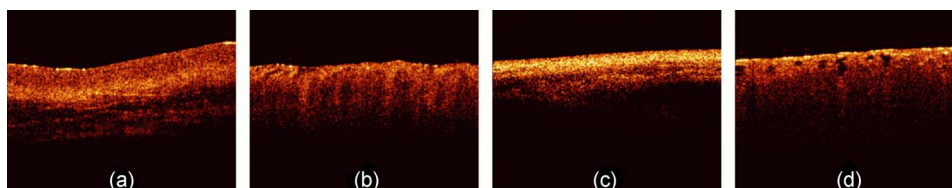


Fig. 7 Representative OCT images of porcine samples: (a) squamous lining of the esophagus, (b) tongue, (c) tracheae, and (d) bronchioles. Each image covers 3 mm in width.

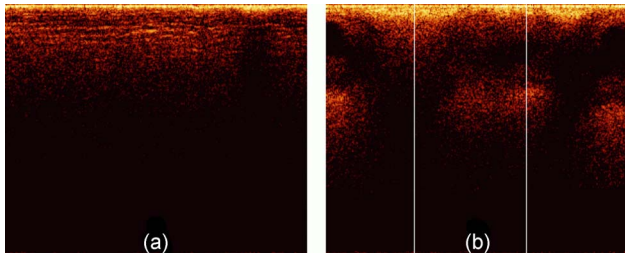


Fig. 8 (a) Surface-normalized B-scans, where the depth profile is homogeneous across the scan width and can be well described by a mean A-scan, and (b) structures that result in different A-scans, e.g., A-scans at the two white lines will be different. In this situation, a standardized mean seems to better represent the B-scan.

layer of mucus, a sheath, a balloon, or something similar. Research into more robust and reliable surface detection algorithms would therefore be beneficial.

4.1.2 Representative measure

Reducing the data to a representative A-scan is necessary; our Pentium type processor with 1 Gbyte of memory was not able to handle data sets of tens of thousands of A-scans. More powerful computers are of course available, but it is conceivable that OCT data will stretch these resources to the limit for the near future. We found that the M/SD of a surface-normalized B-scan or the mean of the frequency content yield good results. For instance, at separating two similar celery orientations, the SN-MSD approach practically fails (56.7%), whereas the AFT-SD reaches over 80%. For classifying porcine tissue, the SN-MSD approach had the best result. In the case of variations within a B-scan, such as in Fig. 8, standardizing using the SD does improve results for the subsequent classification.

In terms of preprocessing, we also investigated mean centering, and using more than one representative measure (data not shown). We did not perform any correction of the surface vector in this study, although that is beneficial (data not shown). Further, the performance of LDA can be improved by feeding an optimum number of PCs. However, this optimum depends on the number and kind of groups in the classification model. The different sample groups in this study do have different optimum numbers of PCs. For instance, for classifying four tissue types, the classification performance, using 25 PCs, was 90.0% (SN-MSD) and 82.5% (AFT-M). For these groups, the optimum for SN-MSD would be reached by feeding 11 PCs (97.5% LOGO cross-validated), and for AFT-M 7 PCs (91.3%). For the purpose of this demonstration, we decided to keep a fixed number of PCs. It might well be worth investigating a combination of two or more measures, and automatically refining extracted measures. However, this is beyond the scope of this paper and is probably only worthwhile when refining the technique for a dedicated purpose. Summarizing, it is possible to improve results by fine-tuning and combining parameters.

4.1.3 Speed

One criterion for processing is speed. OCT acquires data in real time and hence the subsequent analysis should not impact on this characteristic. Preprocessing does slightly delay re-

sults: on a 2.80-GHz Pentium 4, classification of 330 B-scans takes approximately 85 s, of which 73 s are required for surface normalization and extracting a representative measure. That corresponds to roughly 60,000 A-scans/min and makes us confident that the claim of providing a near real-time modality remains justified. It would be quite remarkable for a human operator to classify 330 images in $1\frac{1}{2}$ min. However, we were only using a run-length-encoded programming language for proof of principle; compiled programs run much faster.

4.2 Classification

Robust cross-validation methods are important for testing LDA. It is a powerful approach, and for a small enough set of samples some separation into groups can always be expected. For instance, in the example of 30 images, the model created by LDA was able to classify 100% correctly even without surface normalization. After cross-validation, the surface-normalized model was classified in 93.3 versus 33.3% for the not-surface-normalized model.

4.2.1 Relevance for medical data

The fact that the algorithm reaches over 90% accuracy in certain cases is reassuring when looking at obviously different images, but for our data set the displayed approach is also able to differentiate images reasonably well that are difficult to tell apart by the eye. Further, the algorithm is still able to classify similar images among a pool of other structural data.

This is important as we plan to apply this algorithm to medical data. There, we expect a larger variation among healthy individuals, and only eventually the small, but significant alterations for early, nonsymptomatic disease. As an example, we note epithelial cancers, carcinomas, where the epithelial thickness of the healthy population can vary considerably due to various factors. However, the subtle alterations of proliferative epithelial cells close to the basal layer are probably the feature that can be picked up by OCT scans. It still remains to be assessed whether the algorithm would fail in such scenarios.

4.2.2 Limitations and benefit

In this study the imaging protocol is strict in the sense that it requires the penetration depth to be fully exploited and to leave an area of air above the surface. However, most imaging must adhere to a protocol. As long as the operator is made aware of the requirements, we believe this limitation not to be substantial. After all, we have acquired thousands of images in this way.

The advantage we see in the presented algorithm is that the few parameters utilized here are robust enough for the analysis to be automated and in near real-time. It does not require assigning areas of interest. It does not rely on specific structures such as previously reported methods, but can still be used in combination with these algorithms. Hence, this algorithm is straightforward to implement for existing OCT data.

Note that, although we were taking representative measures of B-scans, this approach will work on volumetric data. For example, taking a representative measure of a volume

consisting of 20×20 A-scans on a system with $10\text{-}\mu\text{m}$ lateral resolution will enable classification of 25 “tiles”/mm², which is still superb for targeting clinical treatment.

4.2.3 Further work

Previous studies^{16,17} have shown that expert histopathologists do not always fully agree on classification of tissue specimens, and an algorithm with lower than 100% correct classification can still outperform subjective human assessment. In clinical work, OCT data would have to be compared to histopathologic results. Even though there are significant discrepancies found between expert pathologists on particular pathology groups,^{16,17} histopathology is defined as the gold standard. Therefore, we must base our analysis on the assumption that the sample pathology defined by histopathology is correct. In this study, we did not investigate the performance of the algorithm against a human operator, and hence assumed a 100% correct classification as a reference baseline. In the future, we intend to compare classification results with those achieved by histopathologists on the same OCT images from biological specimens.

The next step is testing this algorithm on medical OCT images. We expect it to distinguish structurally different pathologies correctly. In ongoing work, we are investigating how well it classifies disease groups with quite similar appearance, such as mild and moderate dysplasia in early cancer development.

Acknowledgments

This work is funded by the Pump Prime Fund of Cranfield University and Gloucestershire Hospitals NHS Foundation Trust. The work presented in this paper would not have been possible without the initial support of Prof. Ruikang K Wang, Dr. Shing Cheung, and Dr. Laurie Ritchie.

References

1. D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, and J. G. Fujimoto, “Optical coherence tomography,” *Science* **254**, 1178–1181 (1991).
2. P. H. Tomlins and R. K. Wang, “Theory, developments and applications of optical coherence tomography,” *J. Phys. D* **38**, 2519–253 (2005).
3. F. I. Feldchtein, G. V. Gelikonov, V. M. Gelikonov, R. V. Kuranov, A. M. Sergeev, N. D. Gladkova, A. V. Shakhov, N. M. Shakhova, L. B. Snopova, A. B. Terent’eva, E. V. Zagainova, Y. P. Chumakov, and I. A. Kuznetzova, “Endoscopic applications of optical coherence tomography,” *Opt. Express* **3**(6), 257–270 (1998).
4. P. Targowski, B. Rouba, M. Wojtkowski, and A. Kowalczyk, “The application of optical coherence tomography to non-destructive examination of museum objects,” *Stud. Conserv.* **49**, 107–114 (2004).
5. H. Liang, M. Cid, R. Cucu, G. Dobre, A. Podoleanu, J. Pedro, and D. Saunders, “En-face optical coherence tomography—a novel application of non-invasive imaging to art conservation,” *Opt. Express* **13**, 6133–6144 (2005).
6. S. Wold, K. Esbensen, and P. Geladi, “Principal component analysis,” *Chemom. Intell. Lab. Syst.* **2**, 37–52 (1987).
7. J. Weissman, T. Hancewicz, and P. Kaplan, “Optical coherence tomography of skin for measurement of epidermal thickness by shapelet-based image analysis,” *Opt. Express* **12**, 5760–5769 (2004).
8. X. Qi, M. V. Sivak, Jr., G. Isenberg, J. E. Willis, and A. M. Rollins, “Computer-aided diagnosis of dysplasia in Barrett’s esophagus using endoscopic optical coherence tomography,” *J. Biomed. Opt.* **11**(4), 044010 (2006).
9. M. L. Huang and H. Y. Chen, “Development and comparison of automated classifiers for glaucoma diagnosis using Stratus optical coherence tomography,” *Invest. Ophthalmol. Visual Sci.* **46**(11), 4121–4129 (2005).
10. A. M. Zysk and S. A. Boppart, “Computational methods for analysis of human breast tumor tissue in optical coherence tomography images,” *J. Biomed. Opt.* **11**(5), 054015 (2006).
11. J. Rogowska, C. M. Bryant, and M. E. Brezinski, “Cartilage thickness measurements from optical coherence tomography,” *J. Opt. Soc. Am. A* **20**, 357–367 (2003).
12. Y. Hori, Y. Yasuno, S. Sakai, M. Matsumoto, T. Sugawara, V. Madjarova, M. Yamanari, S. Makita, T. Yasui, T. Araki, M. Itoh, and T. Yatagai, “Automatic characterization and segmentation of human skin using three-dimensional optical coherence tomography,” *Opt. Express* **14**, 1862–1877 (2006).
13. Y. Yang, A. Dubois, X.-P. Qin, J. Li, A. El Haj, and R. K. Wang, “Investigation of optical coherence tomography as an imaging modality in tissue engineering,” *Phys. Med. Biol.* **51**, 1649–1659 (2006).
14. D. C. Fernández, H. M. Salinas, and C. A. Puliafito, “Automated detection of retinal layer structures on optical coherence tomography images,” *Opt. Express* **13**(25), 10200–10216 (2005).
15. R. G. Brereton, “Experimental design and Pattern recognition,” Chaps. 1 and 4 in *Chemometrics—Data Analysis for the Laboratory and Chemical Plant*, pp. 15–53, 191–224, John Wiley & Sons Ltd. (2003).
16. C. Kendall, N. Stone, N. Shepherd, K. Geboes, B. Warren, R. Bennett, and H. Barr, “Raman spectroscopy, a potential tool for the objective identification and classification of neoplasia in Barrett’s esophagus,” *J. Laser Appl.* **200**, 602–609 (2003).
17. N. A. Shepherd, “Barrett’s esophagus: its pathology and neoplastic complications,” *Esophagus* **1**(1), 17–29 (2003).

Optical Coherence Tomography (OCT) Imaging and Computer Aided Diagnosis of Human Cervical Tissue Specimens

F. Bazant-Hegemark^{*a,b}, N. Stone^{a,b}, M.D. Read^c, K. McCarthy^d, R.K. Wang^e

^aCranfield Health, Cranfield University at Silsoe, Bedfordshire MK45 4DT, UK

^bBiophotonics Research Group, Gloucestershire Royal Hosp., Great Western Rd, Gloucester GL1 3NN, UK

^cWomen's Health Directorate, Gloucestershire Royal Hospital, Great Western Rd, Gloucester GL1 3NN, UK

^dDept. of Histopathology, Cheltenham General Hosp., Sandford Road, Cheltenham GL53 7AN, UK

^eDept. of Biomedical Engineering, Oregon Health & Science University, Beaverton, OR 97006, US

ABSTRACT

The keyword for management of cervical cancer is prevention. The present program within the UK, the 'National Health Service (NHS) cervical screening programme' (NHSCSP), is based on cytology. Although the program has reduced the incidence of cervical cancer, this program requires patient follow ups and relies on diagnostic biopsying. There is potential for reducing costs and workload within the NHS, and relieving anxiety of patients. In this study, Optical Coherence Tomography (OCT) was investigated for its capability to improve this situation. Our time domain bench top system used a superluminescent diode (Superlum), centre wave length $\sim 1.3 \mu\text{m}$, resolution (air) $\sim 15 \mu\text{m}$. Tissue samples were obtained according to the ethics approval by Gloucestershire LREC, Nr. 05/Q2005/123. 1387 images of 199 participants have been compared with histopathology results and categorized accordingly. Our OCT images do not reach the clarity and resolution of histopathology. Further, establishing and recognizing features of diagnostic significance seems difficult. Automated classification would allow one to take decision-making to move from the subjective appraisal of a physician to an objective assessment. Hence we investigated a classification algorithm for its ability in recognizing pre-cancerous stages from OCT images. The initial results show promise.

Keywords: OCT, epithelial malignancy, cancer prevention, computer aided diagnosis, pre-cancer classification

1. INTRODUCTION

Cancer management in the 21st century has shifted from providing treatment or palliative care towards prevention. Epithelial malignancies contribute to the vast majority of cancers. Justification for preventive surgery is often difficult to obtain.

The management of cervical cancer in the western world can be regarded as a model for cancer management. It has been in place for several decades, and has generated knowledge about precancerous lesions. Screening and early diagnosis have reduced the mortality from this cancer. However, decision making is still widely based on diagnosing tissue biopsies. Histopathology of these tissue samples is the gold standard, yet it is time consuming and invasive.

Various emerging modalities have been assessed for their feasibility or benefit in alleviating this problem. Amongst these are many optical techniques utilized for their non-destructive probing of tissue morphology or biochemistry. Optical Coherence Tomography (OCT) has been proposed as a method for investigating early epithelial cancers in the late 1990s (1, 2).

1.1 Cervical cancer

The most prevalent form of cervical cancer is squamous cell carcinoma and it is preceded by a long phase of preclinical disease. This involves dysplasia which does not necessarily lead to cancer, but with increasing severity the likelihood of developing cancer rises. These early changes do not cause symptoms which would be noticeable by the patient. The pre-cancerous development, cervical intra-epithelial neoplasia (CIN), is classified into 3 groups: CIN1 (mild), CIN2 (moderate), and CIN3 (severe). Essentially the decision for surgical removal of a pre-cancerous lesion is based on a

* f.bazant@medical-research-centre.com, phone +44 845 422 5474

CIN2 or worse staging; CIN1 does not usually justify treatment. However, CIN1 and CIN2 are similar and therefore are the most difficult for the clinician to distinguish.

Cervical cancer differs from the majority of cancers in two main ways: Firstly, its etiology is defined and linked to the patient's vulnerability to pathological strains of the human papilloma virus. Secondly, because of the defined pre-cancerous stages, development and implementation of cancer prevention programs has been possible.

1.2 Non-invasive probing

Still, there is much research on cervical cancer: Screening programs are not perfect and can be improved in terms of specificity and sensitivity. Successful screening programs require complex infrastructure, and a simpler approach would hold the potential for being used in low resource settings. Further, because of the decades of experience with the cervical screening, the performance of such programs can serve as a reference for comparison with new methods – this is not the case for other organs. So even if a new technology is not sufficient to replace present cervical cancer management, it might still of use in a different body site where no such management is in place.

Emerging technologies which have been investigated for improving management of cervical cancer include Fluorescence Spectroscopy, Light Scattering, Confocal Microscopy, Infrared Spectroscopy, Raman Spectroscopy, and other, non optical probes (3). In OCT, there are a few studies not only suggesting its application for cervical pre-cancer assessment, but also demonstrating its feasibility. OCT can be beneficial as an adjunct in cervical pre-cancer assessment (4-7).

1.3 Investigation in this study

This study has established a cervical image database to enable a novel image analysis technique to be evaluated. This technique uses Principal Components Analysis (PCA) and Linear Discriminant Analysis (LDA). Both are statistical tools which have not been applied widely for evaluation of OCT data yet. In previous work we were able to display the ability of this approach on a set of non-clinical images. We believe this approach to be suitable for medical imaging data. Specifically in OCT imaging, data evaluation is not always straight forward. Therefore algorithms which are able to improve our understanding of biological OCT images are of interest, such as e.g. presented in research by Turchin et al. (8).

There is still no commercial OCT system established for cancer assessment – in contrast to ophthalmic OCT systems which have been widely introduced. The amount of studies applying OCT in pre-cancer assessment, over the past 10 years, illustrates that the challenge of data evaluation cannot be addressed by improving image quality by technical advances alone: In order to solve this, classification algorithms could be of benefit.

2. MATERIALS & METHODS

2.1 The study protocol

Human tissue samples were collected following the 'Favourable opinion' (approval) of the Gloucestershire Local Research Ethics Committee, Study Nr. 05/Q2005/123. Samples were obtained from participants who needed a biopsy for diagnosis or treatment. These were acquired routinely for histopathological examination, and therefore there was no need for additional biopsies to be taken for this study. The samples were categorized into three groups: hysterectomies, large loop excisions of the transformation zone (LLETZ) (both are treatments) and punch biopsies (diagnostic).

Samples from LLETZ, or "loop", procedures are also referred to as 'biopsies'. To clarify the difference, we refer to diagnostic samples as "punch biopsies", to treatment samples as "LLETZ", and we use the term "sample" rather than "biopsy" when addressing both groups.

2.2 The OCT system

The OCT device used was a bench top time-domain system similar in design to the system described by Yang et al. (9). The light source used is a super luminescent diode (SLD, Superlum) with a central wavelength of 1310 nm and has an axial resolution of ~15 μm in air and 10 μm lateral resolution. The rapid delay line uses a double pass mirror and a reference mirror mounted on a galvanometer (Cambridge Instruments). The components were connected using fiber rather than bulk optics to display feasibility for *in vivo* application. Data acquisition was performed using in-house written Labview programs.

2.3 Measurements

For routine diagnosis, samples were put in formalin-containing pots immediately after surgical removal, and processed by the histopathology department. However, in our study, samples were placed in physiological saline before imaging was performed. After imaging, samples were placed in formalin and sent to histopathology, as not to cause a delay for the diagnostic processing of the sample.

Sample orientation was difficult for small punch biopsies, whereas larger samples (LLETZ and from hysterectomies) were easier to handle: Samples, which did not provide useful images (because of orientation difficulties, twisting or curling) were not included in the automated classification. The imaging sites were tagged using an orientation needle thereby roughly marking the imaging positions. From larger samples it was possible to obtain images from healthy areas (near the excision margins) and images of progressed CIN.

2.4 Data evaluation

For the purpose of this data evaluation, we were using the shape of an A-scan and descriptors for groups of A-scans after normalizing them to the surface shape. These relatively complex descriptors were then reduced by PCA. The strongest components were then fed into LDA for building classification models.

On basis of this approach we have previously performed a study (submitted for publication). We were able to distinguish different types of animal tissue (oesophagus, tongue, trachea, and broncheoli from a pig) with a classification rate up to 100 %. Figure 1 shows four representative images of such tissue types which can be classified correctly in a fully automated fashion, while Figure 2 shows this classification as a bar chart plot. A detailed description of this classification will be introduced in a future publication.

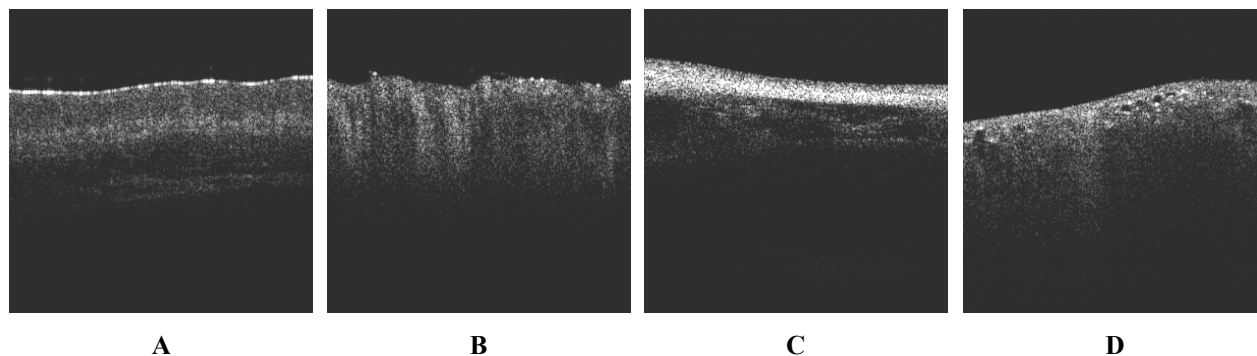


Figure 1: Four representative OCT images of different animal tissue types. Each image has a width of 3 mm and a resolution of 10 μm . Sites are: A) oesophagus, B) tongue, C) tracheae, D) broncheoli.

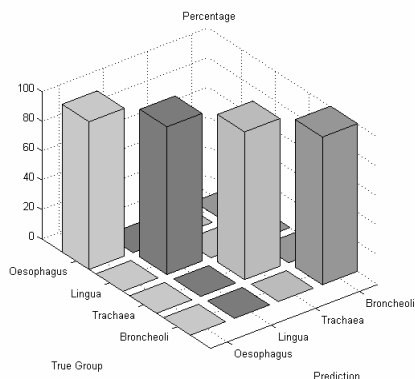


Figure 2: The plot shows how the separation of 80 images as shown in Figure 1 allowed 100 % correct classification.

Figure 1 and Figure 2 show that automated classification shows promise and hence we used this approach for classifying OCT imaging data from cervical pre-cancerous samples.

Data handling, pre-processing, PCA and LDA were performed using Matlab V6.1.0.450, Release 12.1, extended by the 'PLS_Toolbox' [Eigenvector] and our in house tools (the 'Classification_Toolbox').

3. RESULTS

3.1 Recruitment

Over the course of 56 clinic sessions, 223 women consented to be included in the study. Not all of the patients recruited needed a biopsy and therefore the number of study participants was lower. Samples were collected from 199 patients; in some patients multiple biopsies were taken. 245 tissue specimens were obtained, 149 punch biopsies, 92 LLETZ, and 4 specimens from hysterectomies.

The mean age (\pm standard deviation) of participants was 31.8 ± 8.3 . Clinically relevant information including their ethnic background, parity, and hormonal status was collected in order to be able to investigate potential trends. This information may be used in future subgroup analysis of the data.

3.2 Images

1387 images were obtained from 245 samples; the breakdown of pathology identified was as followed: 20 healthy ('no sign of neoplasia'), 87 CIN1, 65 CIN2, and 73 CIN3. Figure 3 demonstrates images obtained of the individual pathologies as listed above. The images measure 500×300 pixels for 3 mm width and approximately 5 mm depth. Orientation of tiny biopsies was difficult and this was found to affect image quality.

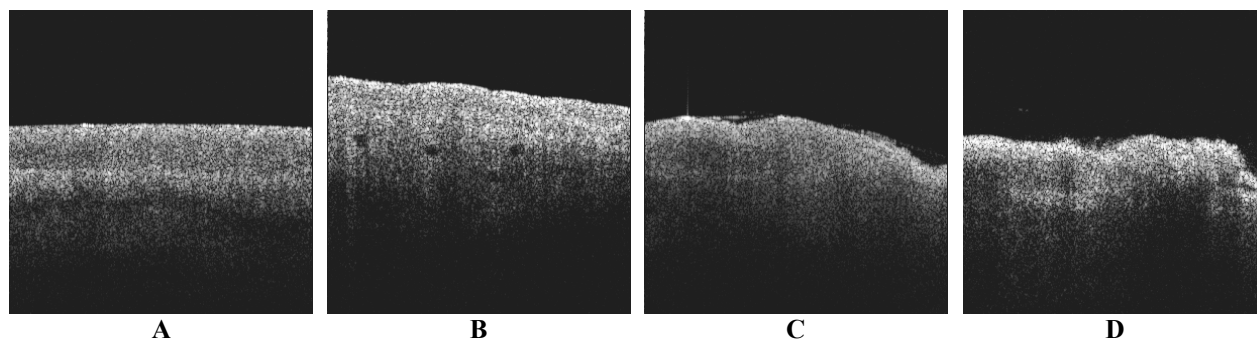


Figure 3: Representative OCT images of cervical specimens. Each image has a width of 3 mm and a resolution of $10 \mu\text{m}$. The staging according to expert histopathology is: A) healthy ('no sign of neoplasia'), B) CIN1, C) CIN2, D) CIN3. From these images it is not always straight forward to establish or recognize features which are relevant of disease progress.

3.3 Data evaluation

The preliminary result of the data analysis for a training set of 2 groups (Healthy/CIN1 vs. CIN2/CIN3) is illustrated in Table 1. 64 good quality images were used to construct the training set, 29 of healthy/CIN1, and 35 of CIN2/CIN3. The training performance of this model was 82.8 % accurate.

Table 1: Result of the training performance (82.8 %) of a set of 64 images (29 healthy/CIN1 vs. 35 CIN2/CIN3).

Training performance of preliminary model (64 images)			
		Prediction	
		Healthy/CIN1	CIN2/CIN3
True group	Healthy/CIN1	24	5
	CIN2/CIN3	6	29

The largest data set to date includes 623 images (226 Healthy/CIN1 vs. 397 CIN2/CIN3), using the same pre-processing parameters as the 2-group-model used above. The training performance of this model (illustrated in Table 2) was 61.0 %.

Table 2: Result of the training performance (61.0 %) of a set of 623 images (226 healthy/CIN1 vs. 397 CIN2/CIN3).

Training performance of preliminary model (623 images)			
		Prediction	
		Healthy/CIN1	CIN2/CIN3
True group	Healthy/CIN1	130	96
	CIN2/CIN3	147	250

4. DISCUSSION

4.1 Recruitment

Because this study did not interfere with the management a patient receives normally, we made the assumption that patients would usually be willing to participate. This was generally the case for patients with treatment. Patients who attended a clinic for a diagnostic biopsy were in some cases more reluctant. We assume that although the provided information sheet clearly emphasized that no extra biopsies would be taken, this point might not have always come across. Nevertheless our study size seems comparable with earlier studies in this field (6, 7).

4.2 Sample quality

When measuring the samples, we had problems as the penetration depth of our system exceeded the size of the punch biopsies, as shown in Figure 4. This does affect the data evaluation as it is based on the shape of a full A-scan. For small samples, manual handling was difficult and made proper imaging impossible. The images which are representative for an *in vivo* situation are therefore mainly from LLETZ procedures and hysterectomies. As a consequence, we had to exclude some of the images from our automated analysis based on the criterion whether an artifact would have occurred *in vivo*.

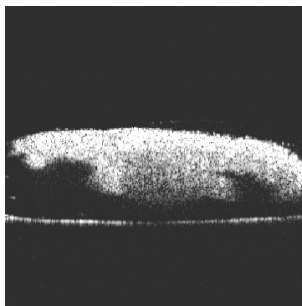


Figure 4: Representative OCT image of a punch biopsy (Image width 3 mm, resolution $\sim 10 \mu\text{m}$). Although layers can be seen, it is not useful for proper assessment of the technique. The line below the sample is from the support hold. This is a situation which would not occur *in vivo*.

From the study design, non-cancerous samples would have been obtained from hysterectomies (which were performed for other indication than cervical cancer). We have actually obtained a further 16 non-cancerous samples from diagnostic punch biopsies: this shows the potential of a non-invasive technology: With today's means, these punch biopsies were justified and necessary for ruling out the presence of cancer. In the future, these are exactly the cases which we would like to make redundant.

4.3 Data evaluation

The results from the non-clinical demonstrator set and the results from the included cervical images show that the algorithm can classify tissue automatically. However, this relies on the appropriate pre-processing. Pre-processing aims to emphasize features which can then be easily extracted and fed into the classification. Such features can be layer thickness, grade of layer disruption, or intensity patterns of A-scans, just to name a few.

Ideally, but not necessarily, these features correlate clearly with the progression of pre-cancerous lesions. Establishing such features in pre-cancerous OCT images of the cervix so far remains challenging. There are many structures such as glands, cysts, or characteristic wart surfaces, which do not correlate with CIN grades. For a fully automated algorithm, such features need to be picked up properly and normalized in order to not disturb the analysis. Notably, the ability to extract features depends on the original quality of the image. Today, OCT systems provide higher resolution and higher acquisition speed. Both might be crucial for automated registration of image features and artifacts.

At the moment we are working on establishing such features which can serve as a set of rules for classifying medical data. We found that certain features, such as thickness of the epithelial, can vary considerably between patients. We have data about age, ethnic background, and hormonal status of the participants but have not yet been able to establish a correlation between such features or whether stratifying the data is necessary for the algorithm to give better results.

Specifically for epithelial layer thickness, and also for other features, the inter-patient variation can be larger than the intra-patient variation. For such cases an ideal study would have to be designed to observe individuals for a longer time.

4.4 Communication to clinician

For the clinician it is of interest to get clear answers. Even if new imaging modalities deliver astonishing pictures, the usability is flawed if the interpreter has to deal with a new repertoire of image features and artifacts. OCT is a real-time imaging modality and it is likely that the amounts of data will be too complex to be visually evaluated by a clinician. It will be of benefit if computer algorithms are able to classify data.

From our findings we believe that for automated evaluation, imaging of the whole cervix *in vivo* will be ideal. OCT would be applied in early stages of pre-cancerous development, when a patient still has a major healthy area which can be used as intra-patient reference. Classification models would therefore not have to be as robust to inter-patient variation.

Of course a 100 % reliable and robust classification will be desirable. If this aim cannot be achieved, OCT still can be used to rule out clear cases, thereby reducing the amount of cases which need the attention of a specialist.

REFERENCES

1. F.I. Feldchtein, G.V. Gelikonov, V.M. Gelikonov, R.V. Kuranov, A.M. Sergeev, N.D. Gladkova, A.V. Shakhov, N.M. Shakhova, L.B. Snopova, A.B. Terent'eva, E.V. Zagainova, YuP. Chumakov, I.A. Kuznetzova, "Endoscopic applications of optical coherence tomography", *Optics Express* **3**(6):257-270 (1998).
2. C. Pitris, A. Goodman, S.A. Boppart, J.J. Libus, J.G. Fujimoto, M.E. Brezinski, "High-resolution imaging of gynecologic neoplasms using optical coherence tomography", *Obstet Gynecol.* **93**(1):135-9(1999).
3. R.A. Drezek, R. Richards-Kortum, M.A. Brewer, M.S. Feld, C. Pitris, A. Ferenczy, M.L. Faupel, M. Follen, "Optical imaging of the cervix", *Cancer* **98**(9 Suppl):2015-27 (2003).
4. P.F. Escobar, J.L. Belinson, A. White, N.M. Shakhova, F.I. Feldchtein, M.V. Kareta, N.D. Gladkova, "Diagnostic efficacy of optical coherence tomography in the management of preinvasive and invasive cancer of uterine cervix and vulva", *Int J Gynecol Cancer* **14**(3):470-4 (2004).
5. P.F. Escobar, L. Rojas-Espaillet, J.L. Belinson, "Optical diagnosis of cervical dysplasia", *Int J Gynaecol Obstet.* **89**(1):63-4 (2005).
6. A.F. Zuluaga, M. Follen, I. Boiko, A. Malpica, R. Richards-Kortum, "Optical coherence tomography: a pilot study of a new imaging technique for noninvasive examination of cervical tissue", *Am J Obstet Gynecol.* **193**(1):83-8 (2005).
7. P.F. Escobar, L. Rojas-Espaillet, S. Tisci, C. Enerson, J. Brainard, J. Smith, N.J. Tresser, F.I. Feldchtein, L.B. Rojas, J.L. Belinson, "Optical coherence tomography as a diagnostic aid to visual inspection and colposcopy for preinvasive and invasive cancer of the uterine cervix", *Int J Gynecol Cancer* **16**(5):1815-22 (2006).
8. I.V. Turchin, E.A. Sergeeva, L.S. Dolin, V.A. Kamensky, N.M. Shakhova, R. Richards-Kortum, "Novel algorithm of processing optical coherence tomography images for differentiation of biological tissue pathologies", *J Biomed Opt* **10**(6):064024 (2005).
9. Y. Yang, A. Dubois, X-P. Qin, J. Li, A. El Haj, R.K. Wang, "Investigation of optical coherence tomography as an imaging modality in tissue engineering". *Phys. Med. Biol.* **51**:1649-1659 (2006).



Optical coherence tomography: A potential tool for unsupervised prediction of treatment response for Port-Wine Stains

F. Bazant-Hegemark^{a,b}, I. Meglinski^{a,c}, N. Kandamany^c,
B. Monk^c, N. Stone PhD, MBA^{a,b,*}

^a Cranfield Health, Cranfield University at Silsoe, Bedfordshire MK45 4DT, UK

^b Biophotonics Research Group, Gloucestershire Royal Hospital, Great Western Road, Gloucester GL1 3NN, UK

^c Laser Treatment Centre, Bedford Hospital NHS Trust, Kempston Road, Bedford MK42 9DJ, UK

Available online 19 October 2008

KEYWORDS

Discriminant analysis;
Optical coherence
tomography;
Port-Wine Stain

Summary

Background: Treatment of Port-Wine Stains (PWS) suffers from the absence of a reliable real-time tool for monitoring a clinical endpoint. Response to treatment varies substantially according to blood vessel geometry. Even though optical coherence tomography (OCT) has been identified as a modality with potential to suit this need, it has not been introduced as a standard clinical monitoring tool. One reason could be that – although OCT acquires data in real-time – gigabyte data transfer, processing and communication to a clinician may impede the implementation as a clinical tool.

Objectives: We investigate whether an automated algorithm can address this problem.

Methods: Based on our understanding of pulsed dye laser treatment, we present the implementation of an unsupervised, real-time classification algorithm which uses principal components data reduction and linear discriminant analysis. We evaluate the algorithm using 96 synthesized test images and 7 clinical images.

Results: The synthesized images are classified correctly in 99.8%. The clinical images are classified correctly in 71.4%.

Conclusions: Principal components-fed linear discriminant analysis (PC-fed LDA) may be a valuable method to classify clinical images. Larger sampling numbers are required for a better training model. These results justify undertaking a study involving more patients and show that disease can be described as a function of available treatment options.

© 2008 Elsevier B.V. All rights reserved.

Introduction

Port-Wine Stains

Capillary haemangiomas (Port-Wine Stains, PWS) are congenital malformations of dermal capillaries. In the UK, PWS

* Corresponding author at: Biophotonics Research Group, Gloucestershire Royal Hospital, Great Western Road, Gloucester GL1 3NN, UK.

E-mail address: n.stone@medical-research-centre.com (N. Stone).

are prevalent in 3 of 1000 inhabitants [1]. Particularly in facial areas, lesions may have an impact on self-esteem of patients. The development of the pulsed dye laser marked a significant advance in treatment; its mechanism is believed to involve coagulation of dermal capillaries via the absorption of laser energy by oxyhaemoglobin. However, blood vessels have varying diameter, wall thickness, flow rates, depth, and carry blood with different oxygenation states. For a successful response to treatment, a lesion needs to be within the penetration depth of the laser wavelength. The most common PDL lasers operate at 577 and 585 nm [2] and do not penetrate more than 1 mm into skin [3]. Blood vessels must be destroyed entirely to prevent regeneration and revascularization. Capillaries need to hold enough oxyhaemoglobin to absorb energy; a vessel thickness $<50\ \mu\text{m}$ does not provide sufficient blood volume for a good response [2]. With standard PDL lasers, good response is expected for blood vessels in a depth of up to 1 mm and with a diameter of 55–150 μm , depending on body site and skin type [4]. Thick skin layers and strong pigmentation reduce the flux that reaches the capillary. Response to treatment can be improved by altering pulse length, wavelength, energy, vessel dilation (thermal or drug-induced), or by cooling surrounding tissue to allow longer pulses [2,5,6]. These treatment options carry however an increased risk of complications.

Important practical problems are the lack of objectivity in evaluating clinical response, and the difficulty of predicting which patients will benefit from treatment. On average, skin lightens 12% per treatment [4]. Complete lightning can be achieved in less than 25% of the lesions, and 70% of PWS lighten at least 50% [7]. Purpura, crusting and bleeding are common side effects which can last between 2 and 14 days. After about 6 weeks, a more reliable, but yet again visual, assessment of treatment outcome is possible. Protocols usually require 6–10 sessions in 6–12 weeks intervals. For a vast majority, “excellent” results in terms of patient satisfaction can be achieved; however this is only established after repeated sessions with no progress in outcome. In up to 20%, no improvement or darkening of the affected lesion is observed [7–9]. For these patients, unsuccessful PDL sessions present therefore a rather ineffective way of determining treatment response.

Optical coherence tomography

Optical coherence tomography (OCT) is a non-destructive imaging modality. It uses low coherent laser light to localize backscattering events within a specimen (similar to ultrasonography). For systems employing typical infrared sources for human dense tissue assessment, the characteristic penetration depth is around a millimetre and the resolution is around 10 μm [10]. Conventional OCT probes for variations of the refractive index and renders structural information. It is therefore widely investigated for replacing or guiding destructive tissue sampling (biopsies) [11]. Today, fast systems can acquire about a million A-scans per second, in gigabytes of data, which would correspond to about a cm^2 per second. Penetration depth, usually a limit in dense tissue, is not so much a problem for this application: a PWS that

cannot be displayed with OCT will be unsuitable for PDL of shorter wavelength.

Studies towards characterizing blood vessel formation and flow using OCT have been performed on small animals [12,13] and in ophthalmic imaging [14], and in vivo human skin [15], and for monitoring treatment response in vivo [16]. However, despite the initial success and potential which has been identified by these studies, it seems that OCT is not yet available as a clinical modality for standard treatment assessment. Most studies investigate blood flow by exploiting the Doppler effect. We believe that computing accurate velocity models is time intensive: for instance, the angle between the beam and the vessel is integral to these calculations; a velocity vector orientated 90° to the probing beam will show no Doppler signal [17].

In such a scenario, the usually subtle difference between ‘real-time’ and ‘near real-time’ may become crucial. Although imaging is straightforward, the high scanning speed can only be fully utilised if information can be retrieved at similar speed [18]. Investigating clinical OCT images may require training, artefact knowledge, and data visualization skills. Vessel mapping and displaying require time and computational effort, let alone programming skills. Even although volumetric rendering can look astonishing, the actual clinical value lies in providing a reliable support for the choice of treatment.

Such can be given by unsupervised classification algorithms, which consequently are an active and emerging research field for OCT data [18]. Specifically for keratinized layers of skin, or cartilage, procedures to automatically retrieve layer thickness have been investigated [19–21]. All these algorithms are generally slow because they have certain requirements to image quality and clinical markers, and frequently employ iterative steps. A quick method is principal components-fed linear discriminant analysis (PC-fed LDA). It has been displayed to perform well for classifying test images and porcine tissue [22], but has not been applied to OCT data of PWS. This might be challenging due to the variation within a healthy population, but should be investigated because OCT images of PWS seem to show blood vessel diameter and depth with clear contrast. Such images are promising candidates for automated classification.

Research questions

In this communication we investigate how well PC-fed LDA, a classification algorithm developed for OCT data, can be applied to clinical images of PWS, in order to provide a quick and objective support for diagnosing and guiding treatment.

Methods

Due to the low number of clinical images available to us, we designed a set of synthesized images which was used to test the algorithm with a cross-validation approach using combinatory rules. The clinical images, which were not enough for training an algorithm properly, were tested using leave-one-out cross-validation.

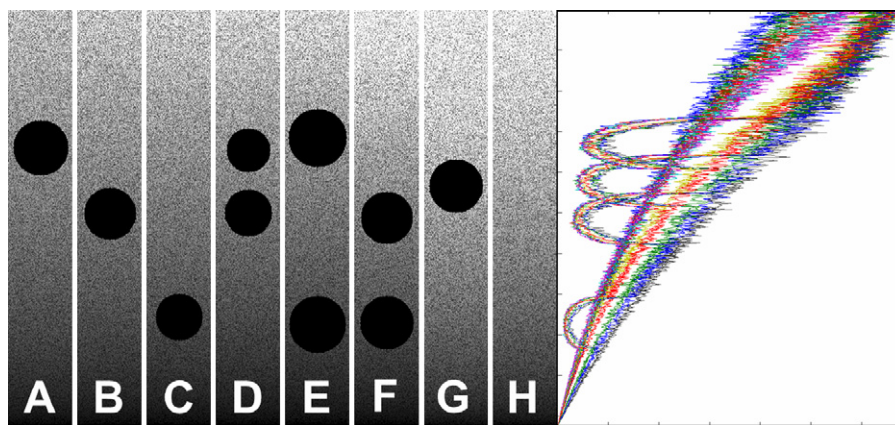


Figure 1 Synthesized test images for assessing the fundamental possibility of classifying PWS as a function of treatment parameters. Different skin types are represented as different grey scale gradients (8 of 12 gradients are shown). The images here are shown as fed into analysis, with random depth and size variations, and with artificially created speckle noise. The classification model is tested on its ability to classify into ‘Treatment option 1’, ‘Treatment option 2’, or ‘Untreatable’ (1, 2, or X). The expected outcome is listed in Table 1. Right graph: mean A-scans of these 96 synthesized images.

Synthesized images

The test images are designed to represent OCT images of PWS with three depth locations. The location was planned as a function of the available treatment option: ‘Treatment option 1’, ‘Treatment option 2’, or ‘Untreatable’, as shown in Fig. 1. These 256 bit-grey level images were created in the Matlab software package (The Mathworks Inc.) and measure 80×512 pixel and have logarithmic intensity gradients representing different signal-to-noise-gradients in dense tissue. For each gradient type, there are eight possibilities of PWS: three variations of depth and combinations as shown in Fig. 1. Assuming that 512 pixels are equivalent to 1 mm penetration depth, a disk of 64 pixels would represent a diameter of $125 \mu\text{m}$. In analogy to the standard processing, the intensity values were normalized to fit to unity values (i.e., ranging from 0 to 1).

The three depth options are presented by positioning the centre of this disk at pixel 164 ($320 \mu\text{m}$), 256 ($500 \mu\text{m}$), and 384 ($750 \mu\text{m}$). These artificial images served as starting point, and the following steps had the purpose to approximate a more realistic situation: a fourth vessel depth at 210 ($410 \mu\text{m}$), which the algorithm should classify as lesion that requires treatment with both PDL wavelengths, was introduced. Artificial speckle noise was added to the images using the Matlab function ‘imnoise’. The diameter of the vessels had a randomly introduced variation of within ± 10 pixel ($\sim 20 \mu\text{m}$), and the depth position varied

within ± 20 pixel ($\sim 40 \mu\text{m}$). For reproducibility, the function for producing random numbers was seeded before creating a set of images using the Matlab function ‘rand(‘state’, 0)’’.

Eight variations of depth, representing vessels, and 12 background gradients, representing tissue type – i.e., 96 images in total – were used for analysis. Table 1 shows the design of classification groups for treatment suggestion and expected response. Every model was trained from six tissue gradients and cross-validated with the remaining six gradients. The mean of the resulting 924 permutations (6 from 12), as shown in Fig. 2, was taken as classification result. At this stage, we were satisfied with a result to be either correct or not, and did not calculate specificities or sensitivities.

Measurements

OCT device

The OCT device used in this study was a time domain bench top system which was similar in design to the system described previously [23]. The light source is a super luminescent diode (SLD, Superlum) with a central wavelength of 1310 nm and provided an axial resolution of $\sim 15 \mu\text{m}$ in air and a $10 \mu\text{m}$ lateral resolution. The rapid delay line is realized using a double pass mirror and a reference mirror mounted on a galvanometer (Cambridge Instruments). The data acquisition was performed using Labview. The actual

Table 1 Desired classification performance in designed model

Classification design	Vessel combination							
	A	B	C	D	E	F	G	H
Vessel depth(s) (μm)	320	500	750	320, 500	320, 750	500, 750	410	–
Treatment option(s)	1	2	None	1 and 2	1	2	1 and 2	None
Treatment response	Full	Full	None	Full	Partial	Partial	Full	–

Vessel combinations are shown in Fig. 1A.

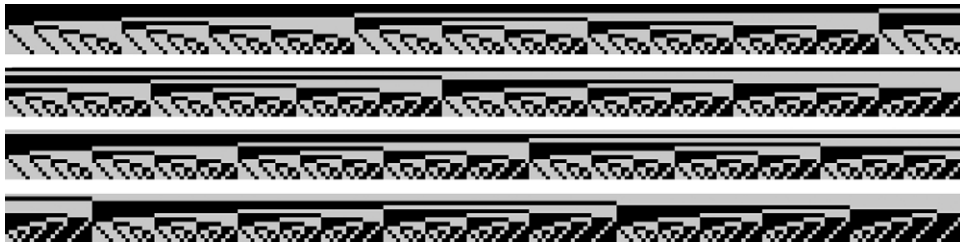


Figure 2 Assigning data to test set or prediction set for cross-validation: representation of 924 permutations for a selection of 6 out of 12, shown in 4 rows of 231 columns. Each column has 12 pixels (6 grey and 6 black). None of the 924 columns equal each other.

acquisition of 333 A-scans per millimetre meant an over-sampling of the beam resolution.

Tissue samples

Images were obtained in agreement with the guidelines by the local research ethics committee of the Bedford Hospitals NHS Trust. Seven patients were enrolled in this preliminary study. The measured sites were chosen due to their cosmetic relevance: forehead, upper-, mid- and lower-cheek, and forearm. The measurements were performed on site by clinicians of the Bedford Hospitals NHS foundation trust. The protocol did not involve any contrast agents or other preparation. At this stage, the samples were confirmed visually by the experience of the clinician. No biopsies were taken for exact matching of the OCT images.

Pre-processing

The top row of an OCT image does usually not coincide with the tissue surface. Because PC-fed LDA classifies depth-related features, the surface of a specimen needs to be known. Synthesized images did not require this pre-processing, as their surface equalled the top image row. All measured images however needed to be pre-processed so that the depth pixel of each A-scan would represent the same distance.

The clinical images in this study measure 512×1000 pixels and cover a width of 3 mm (Fig. 3). Every image depicts the full technical penetration depth of the sample, and there are no excessive reflections. The log of the intensity data (time domain signal) was stored as a matrix. Values were handled in double precision format. The upper part of the image which depicts air was regarded as background noise level. The mean of this plus $7 \times$ the standard deviation of this background signal was used as the intensity threshold. After thresholding, intensity values of one B-scan

were normalized to the maximum intensity peak. A binary thresholding was used to determine the sample surface, as shown in Figs. 4A and B. Some artefacts can cause the surface recognition to fail, e.g. minor reflections and hair which is present as small spots above the surface (cf. Figs. 3 and 4). These were reduced by a sliding median filter. Although the settings for this processing were determined manually, the actual processing was performed fully automated. For the sake of this full automation, we accepted slight misrecognition of the surface as shown in Fig. 4B. After surface normalising, data of low signal-to-noise ratio was removed by cropping data below 300 pixels (Fig. 4D).

From the seven available images, areas (A-scans) containing blood vessels were identified and marked manually using in house written Matlab code as shown in Fig. 4C. Although this is observer biased and based on the assumption that blood vessels show up as darker areas, we chose this approach for a proof of principle. Of the 7000 A-scans, 2864 were selected as vessel-free, the remaining were classified as containing vessels. No A-scan was excluded from analysis. A representative descriptor (the mean of the selected, surface normalised A-scans) was taken of all A-scans of one sample, thereby providing 14 descriptors (7 each for vessel and non-vessel images). We want to emphasise that this manual A-scan selection was purely for training purposes, and that no manual step is required for the classification algorithm.

Data analysis

The algorithm was implemented as described previously [22], in short: data handling, pre-processing, and PC-fed LDA were performed using Matlab V6.1.0.450, Release 12.1 (The MathWorks Inc.), extended by the 'PLS_Toolbox' (Eigenvector Research Inc.) and our in house tools ('Classification_Toolbox').

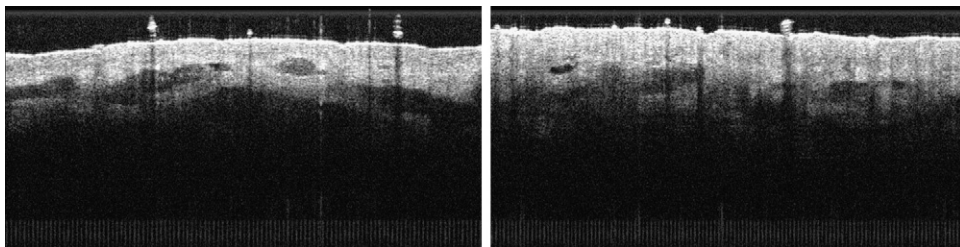


Figure 3 Example OCT images of PWS of two different patients, of the lower-cheek (left image) and mid-cheek (right image).

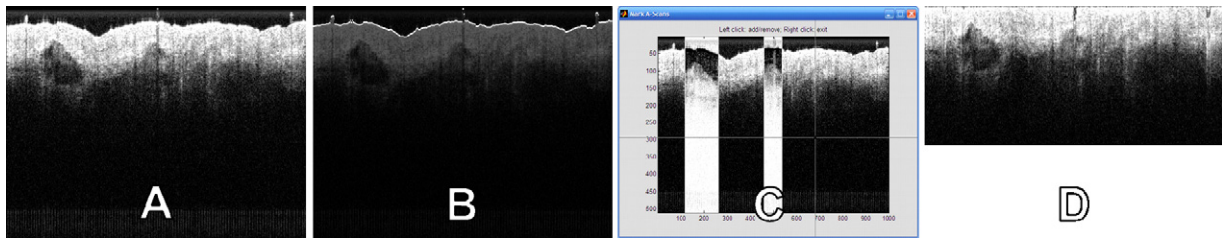


Figure 4 Pre-processing for clinical images. (A) Original scan. (B) The recognized surface as overlay over the image (the image is darker for visualization purposes only). (C) Screenshot of the program for manually selecting vessel A-scans. (D) Surface-normalised and cropped image.

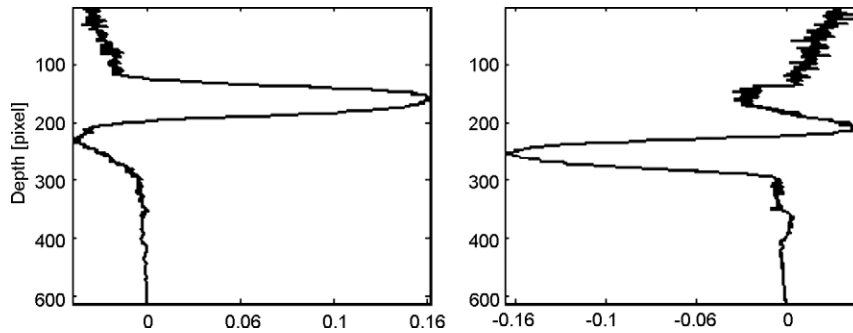


Figure 5 The two strongest loadings for LDA-model of the synthesized images. Note that extrema correlate with vessels depths (Fig. 1).

Using principal components analysis, the data was reduced to the first 10 principal components (PCs) [24,25]. This amount of PCs is an arbitrary number and likely to include PCs of low significance to this discrimination. The strongest loadings of these components were determined by characterising the respective F values using ANOVA. An example for the loadings is shown in Figs. 5 and 6. The PCs were used for linear discriminant analysis, which is inherently able to suppress loadings of lower significance for the model. The cross-validation for the model from synthesized images was performed as explained above using 924 permutations. The clinical images, which provided only 14 descriptors, were too few for training a strong classification model, and was tested using leave-one-out cross-validation.

tations which are possible when taking random 6 out of 12 tissue gradients) for training a classification model.

The training model based on the manual selection of vessel/no vessel, although for a small sample size of clinical images, had a performance of 100%. The leave-one-sample-out cross-validation was successful in 10 out of 14 images, and therefore had a performance of 71.4%. Although sensitivity and specificity seem not appropriate for this very limited study, we can define a success rate for vessel data (6 out of 7) and vessel-free data (4 out of 7). The authors want to emphasise that the last two figures can only give a very rough idea of the performance.

Results

The synthesized images were classified correctly in 99.8%. This is the mean of 924 cross-validations (from all permu-

Discussion

Procedure and limitations

There are three limitations which can be immediately identified from these specimens: because of this low

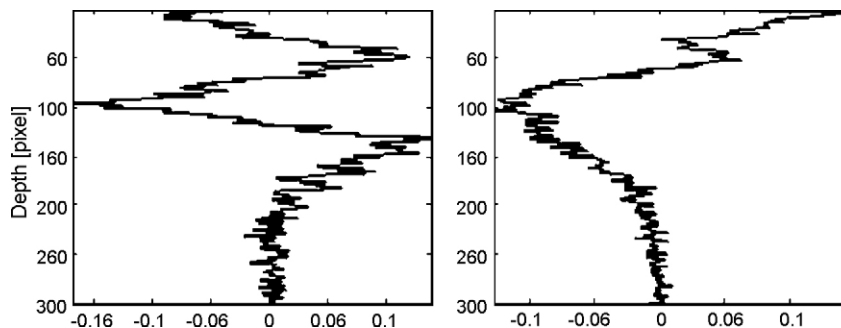


Figure 6 The two strongest loadings for LDA-model of tissue data.

sample number, the results of this classification can only be regarded as preliminary. However, within clinical OCT, where larger trials require approval by ethics boards and are timely, it is common to demonstrate the general applicability of an algorithm on only a few samples [16,26–28].

Secondly, we assume blood vessels to show up with clear contrast. Inspecting images from some research groups, this might not always be the case [16]. In such cases, the velocity profiles obtained by Doppler OCT allow a clearer contrast, and therefore also deeper penetration depths. For the presented algorithm, we expect no fundamental difference in distinguishing the clinical marker, be it an absorption maximum or a velocity profile, as long as such a marker can be clearly correlated with the vessel position.

Further, and more crucially, we were not provided with data on treatment outcome. With the present data, we are able to classify normal tissue from blood vessels, measured from PWS lesions which were visibly identified by eye. Exact correlation with histopathology is a major challenge [21]. Problems in a real clinical situation which we were not able to address in this paper are the impact of skin tone, composition of layers covering the PWS, or dynamic effects like vessel dilation. These issues can be addressed by using a large enough population for training a classification model.

The presented assumptions for vessel geometry and depth might not exactly describe the real situation. We believe that for this initial study these assumptions are justified, as the model can be tweaked to more realistic parameters. After all, our intention is to provide a prognosis on treatment response.

Our results were obtained in 'near' real-time. The cross-validation time for 924 permutations on a 2.8 GHz Pentium 4, 1 GByte RAM PC was 93 s (measured using the Matlab functions "tic" & "toc"), each process of creating a test model and classifying took therefore less than 0.1 s. Real, clinical data sets can be expected to be much larger. However, the time for developing and cross-validating a classification model will then not contribute to the processing time. Roughly extrapolating from the 7680 (80 × 96) A-scans used here, we estimate that a square centimetre imaged with 10 μm resolution, i.e., 10⁶ A-scans, will be classified in less than 13 s.

The model trained by clinical images reaches 71% correct classification. Intra- and inter-patient variation will inevitably result in a certain degree of misclassification. The data was collected from a small population and from different sites of the body; and a model based on a larger pool of data can be expected to perform better. While we cannot rule out an occasional misclassification, we would expect it to stand out of the surrounding classification. The general impression would be a reliable support for the treatment decision.

Summary and conclusions

For a further study design, lesions should be imaged with OCT repeatedly during the intervals between treatment sessions, in order to observe the response after PDL. We envisage a protocol which provides a means to co-register scanned sites in order to allow a reliable timeline, and the design of a training model based on the actual treatment outcome of study participants.

A classification algorithm for OCT opens the vast potential not only for characterizing a site by imaging, but also to determine whether a site is suitable for enhanced treatment options. E.g., optical clearing agents increase penetration depth [29]. Although the clearing effect depends on the wavelength, agents may positively affect treatment parameters like focusing or flow dynamics [30]. OCT would be able to observe, in real-time, when an optimal condition from a treatment preparation has been achieved. This information could also play a role when planning further treatment options for PWS management, such as photodynamic therapy [31–33].

We have displayed an algorithm robust enough to cope with slight variations in depth location and vessel diameter. This algorithm works fully automated and is fast on a standard computer. As such it could play a role in categorizing blood vessel depth according to treatment options.

Acknowledgements

This work is funded by the Pump Prime Fund of Cranfield University and Gloucestershire Hospitals NHS Foundation Trust.

The work presented in this article would not have been possible without the initial support of Prof. Ruikang K Wang, Dr. Shing Cheung, and Dr. Laurie Ritchie.

Dr. Nicholas Stone holds a Career Scientist Fellowship which is funded by the UK National Institute of Health Research.

References

- [1] British Association of Dermatologists, BAD Patient Information Leaflet [last update 2004 October, cited 2007 December 29]. Available from: <http://www.bad.org.uk/patients/leaflets/birthmarks.asp>; 2004.
- [2] Kelly KM, Choi B, McFarlane S, et al. Description and analysis of treatments for port-wine stain birthmarks. *Arch Facial Plast Surg* 2005;7:287–94.
- [3] Bashkatov AN, Genina EA, Kochubey VI, Tuchin VV. Optical properties of human skin, subcutaneous and mucous tissues in the wavelength range from 400 to 2000 nm. *J Phys D: Appl Phys* 2005;38:2543–55.
- [4] Smit JM, Bauland CG, Wijnberg DS, Spauwen PHM. Pulsed dye laser treatment, a review of indications and outcome based on published trials. *Br J Plast Surg* 2005;58:981–7.
- [5] Chiu CH, Chan HHL, Ho WS, et al. Prospective study of pulsed dye laser in conjunction with cryogen spray cooling for treatment of port wine stains in Chinese patients. *Dermatol Surg* 2003;29:909–15.
- [6] Lanigan SW, Taibjee SM. Recent advances in laser treatment of port-wine stains. *Brit J Dermatol* 2004;151:527–33.
- [7] Jasim ZF, Handley JM. Treatment of pulsed dye laser-resistant port wine stain birthmarks. *J Am Acad Dermatol* 2007;57:677–82.
- [8] Soueid A, Waters R. Re-emergence of port wine stains following treatment with flashlamp-pumped dye laser 585 nm. *Ann Plas Surg* 2006;57:260–3.
- [9] Huikeshoven M, Koster PHL, de Borgie CAJM, et al. Redarkening of port-wine stains 10 years after pulsed-dye-laser treatment. *New Engl J Med* 2007;356:1235–40.
- [10] Huang D, Swanson EA, Lin CP, et al. Optical coherence tomography. *Science* 1991;254:1178–81.
- [11] Tomlins PH, Wang RKK. Theory, developments and applications of optical coherence tomography. *J Phys D: Appl Phys* 2005;38:2519–3253.

- [12] Barton JK, Izatt JA, Kulkarni MD, et al. Three-dimensional reconstruction of blood vessels from in vivo color Doppler optical coherence tomography images. *Dermatology* 1999;198:355–61.
- [13] Wang RKK, Jacques SL, Ma Z, et al. Three dimensional optical angiography. *Opt Express* 2007;15:4083–97.
- [14] Makita S, Hong Y, Yamanari M, et al. Optical coherence angiography. *Opt Express* 2006;14:7821–40.
- [15] Zhao Y, Chen Z, Saxer C, et al. Doppler standard deviation imaging for clinical monitoring of in vivo human skin blood flow. *Opt Lett* 2000;25:1358–60.
- [16] Nelson JS, Kelly KM, Zhao YH, Chen ZP. Imaging blood flow in human port-wine stain in situ and in real time using optical Doppler tomography. *Arch Dermatol* 2001;137:741–4.
- [17] Bonesi M, Churmakov DY, Ritchie LJ, Meglinski IV. Turbulence monitoring with Doppler optical coherence tomography. *Laser Phys Lett* 2006;4:304–7.
- [18] Bazant-Hegemark F, Stone N., Towards automated classification of clinical optical coherence tomography data. LIMS, in press, doi:10.1007/s10103-008-0615-6.
- [19] Weissman J, Hancewicz T, Kaplan P. Optical coherence tomography of skin for measurement of epidermal thickness by shapelet-based image analysis. *Opt Express* 2004;12:5760–9.
- [20] Rogowska J, Bryant CM, Brezinski ME. Cartilage thickness measurements from optical coherence tomography. *J Opt Soc Am A* 2003;20:357–67.
- [21] Gambichler T, Moussa G, Regeniter P, et al. Validation of optical coherence tomography in vivo using cryostat histology. *Phys Med Biol* 2007;52:N75–85.
- [22] Bazant-Hegemark F, Stone N. Near real-time classification of optical coherence tomography data using principal components fed linear discriminant analysis. *J Biomed Opt* 2008;13(3):034002.
- [23] Yang Y, Dubois A, Qin X-P, et al. Investigation of optical coherence tomography as an imaging modality in tissue engineering. *Phys Med Biol* 2006;51:1649–59.
- [24] Brereton RG. Experimental design, and pattern recognition, Chapters 1 & 4. In: *Chemometrics—data analysis for the laboratory and chemical plant*. John Wiley & Sons Ltd.; 2003. ISBN: 0-471-84977-6; pp. 15–53 and 191–224.
- [25] Wold S, Esbensen K, Geladi P. Principal component analysis. *Chemometr Intell Lab* 1987;2:37–52.
- [26] Zysk AM, Boppart SA. Computational methods for analysis of human breast tumor tissue in optical coherence tomography images. *J Biomed Opt* 2006;11:054015.
- [27] Iftimia NV, Bouma BE, Pitman MB, et al. A portable, low coherence interferometry based instrument for fine needle aspiration biopsy guidance. *Rev Sci Instrum* 2005;76:064301.
- [28] Hori Y, Yasuno Y, Sakai S, et al. Automatic characterization and segmentation of human skin using three-dimensional optical coherence tomography. *Opt Express* 2006;14:1862–77.
- [29] Proskurin SG, Meglinski IV. Optical coherence tomography imaging depth enhancement by superficial skin optical clearing. *Laser Phys Lett* 2007;4:824–6.
- [30] Vargas G, Readinger A, Dozier SS, Welch AJ. Morphological changes in blood vessels produced by hyperosmotic agents and measured by optical coherence tomography. *Photochem Photobiol* 2003;77:541–9.
- [31] Lanigan S. Improving the treatment of port wine stain birthmarks. *PDPDT* 2007;4:143–4.
- [32] Qin ZP, Li KL, Ren L, Liu XJ. Photodynamic therapy of port wine stains—a report of 238 cases. *PDPDT* 2007;4:53–9.
- [33] Yuan KH, Li Q, Yu WL, Zeng D, Zhang C, Huang Z. Comparison of photodynamic therapy and pulsed dye laser in patients with port wine stain birthmarks: a retrospective analysis. *PDPDT* 2008;5:50–7.



ELSEVIER



www.elsevier.com/locate/pdpdt

HOW-TO-DO-IT

Endoscopic photodynamic therapy for oesophageal disease

Hugh Barr FRCS, FRCSE*, Catherine Kendall, Florian Bazant-Hegemark, Nicholas Stone

Biophotonics Research Group, Cranfield Postgraduate Medical School, Gloucestershire Royal NHS Foundation Trust, Great Western Road, Gloucester GL1 3NN, UK

KEYWORDS

Photodynamic therapy;
Photofrin;
Aminolaevulinic acid;
Barrett's oesophagus

Summary Photodynamic therapy is a very important technique for the eradication of widespread oesophageal mucosal disease which has the potential to degenerate to cancer. Patients unsuitable or unwilling to undergo radical therapy can be cured using photodynamic therapy. We predominantly treat patients with high-grade dysplasia in Barrett's oesophagus. Lesions that are visible macroscopically are removed using endoscopic mucosal resection. The remaining area is then treated in 5 cm segments at 3 monthly intervals with separate photosensitisation using endoscopic photodynamic therapy.

© 2006 Published by Elsevier B.V.

Introduction and patient selection

We predominantly use photodynamic therapy for the treatment of patients with high-grade dysplasia, and early mucosal cancer of the oesophagus. There is also a group of patients with localised recurrence after radical therapy, or microscopic marginal involvement after radical resection in which photodynamic therapy can be very useful in eradication of potential disease that may recur locally. In addition patients, who are unwilling to accept the uncertainty of a surveillance programme for dysplastic disease, or are unfit or unwilling to have radical therapy, are offered photodynamic therapy for early cancer. The diagnostic dilemmas

associated with identification and diagnosis of dysplasia are very important. The level of uncertainty means that photodynamic ablation is a very attractive option for patients.

Counselling the patient. The diagnostic dilemmas

Cancer in both the squamous oesophagus and the columnar-lined oesophagus develops through a multi-step process progressing through low- to high-grade dysplasia, which currently remains the best marker of cancer risk. As many as 40% (range 0–73%) of patients with high-grade dysplasia may already have a coexistent cancer, and between 5 and 60% of patients will develop cancer during surveillance over 1–7 years.

There are profound problems associated with the staging and assessment of dysplasia and cancer

* Corresponding author. Tel.: +44 8454 226679;
fax: +44 8454 226813.

E-mail address: hugh.barr@glos.nhs.uk (H. Barr).

[1]. The division into the four groups of negative for dysplasia; combined indefinite for dysplasia and low-grade dysplasia; high-grade dysplasia; and carcinoma has revealed that there are poor levels of agreement among histopathologists (intra-observer kappa values of 0.64, inter-observer kappa values of 0.43) [2]. The vital separation is that of high-grade dysplasia from intramucosal carcinoma. This classification is also difficult with inter-observer agreement between all pathologists and specific gastrointestinal pathologists for high-grade dysplasia and intramucosal carcinoma having a kappa value of 0.42 and 0.56, respectively [3]. Since the problem in the oesophagus is limited to a depth of 0.6 mm; photodynamic therapy of the oesophagus eradicates the disease, preventing progression and removing the tissue and the diagnostic dilemma.

Clinical assessment

Full clinical assessment of the patient and evaluation of the extent of the diseased area is essential prior to photodynamic therapy. We use standard radial endosonography (7.5 MHz), performed to confirm the superficial nature of the disease and exclude deep muscular invasion and lymph node involvement. High-frequency endosonography (20 MHz) is particularly useful to identify more superficial invasion into the submucosa, since this may be associated with lymph node metastasis. The patient may need to consider more radical options if curative treatment is to be considered.

It is our routine procedure in patients with high-grade dysplasia or early carcinoma to perform extensive biopsy of all macroscopic lesions. Specifically in patients with columnar-lined Barrett's oesophagus we perform quadrant biopsies at 1 cm intervals of the entire columnar-lined segment to attempt to determine whether the lesions are focal or multifocal, and map the extent and distribution of the disease. Magnification chromoendoscopy with dye spraying (indigo carmine solution 0.4–0.8%) helps to identify areas of dysplasia.

If there is a macroscopic nodule or irregularity (Fig. 1), we remove it for histology using endoscopic mucosal resection. We are also investigating optical coherence tomography for the assessment of dysplasia (Fig. 2). The patient is offered photodynamic therapy after staging. There is no absolute contra-indication in elderly or frail patients with co-morbid disease. The management of mul-

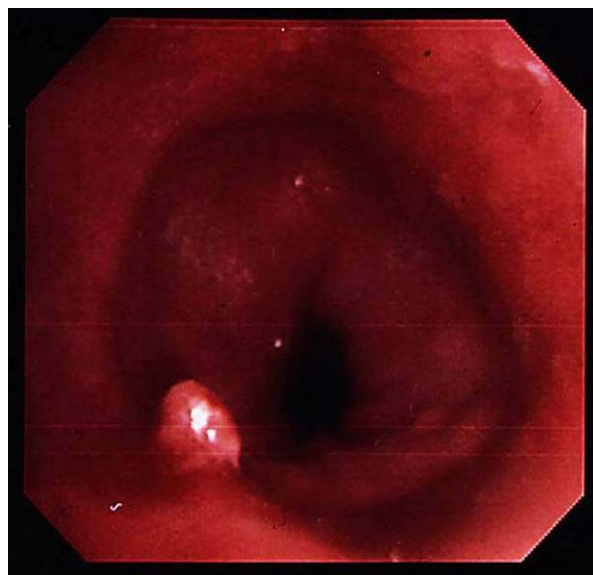


Figure 1 Mucosal nodule in a segment of Barrett's oesophagus. This was removed using endoscopic mucosal resection. The whole area of Barrett's is then ablated using photodynamic therapy.

tifocal areas of high-grade dysplasia are particularly suitable for treatment with photodynamic therapy.

Method of photosensitisation

Photofrin and tetra(*m*-hydroxyphenyl)chlorin (mTHPC)

The patient receives light irradiation 48 h after the administration of 2 mg/kg of Photofrin (porfimer sodium), by slow intravenous injection. The agent is reconstituted from powder immediately prior to injection. It is very important to avoid extravasations so a wide bore intravenous cannula is inserted. The clinical protocol for mTHPC proposes a drug dose of 0.15 mg/kg administered intravenously 4 days before irradiation. Photofrin photodynamic therapy is used if there have been areas of macroscopic abnormality or concern about disease being deeper than the mucosa. It is now our usual practice to use Photofrin photodynamic therapy for all patients following the results of the randomised clinical trial [4].

Aminolaevulinic acid (ALA)

The patient receives 60 mg/kg ALA dissolved in orange juice [5]. We have seen some patients feel

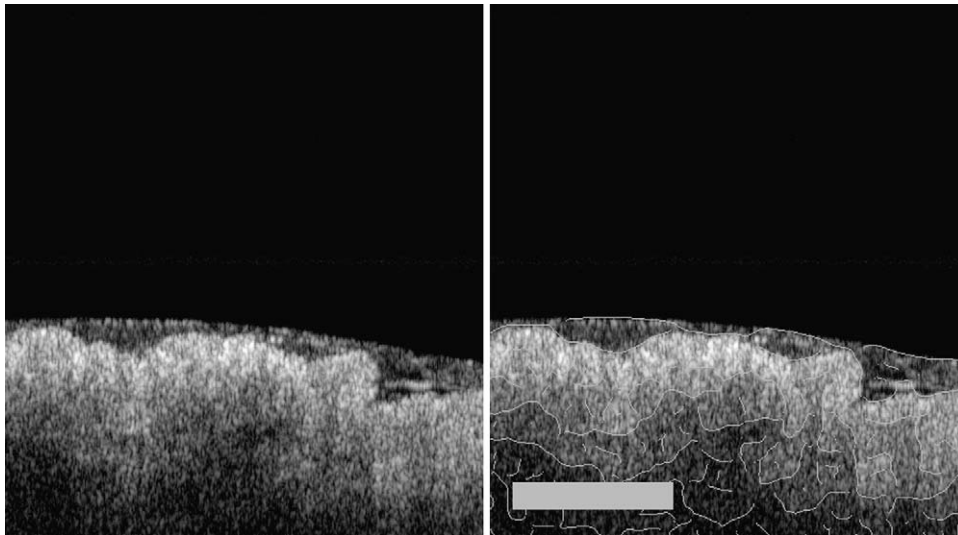


Figure 2 Optical coherence tomography of area of dysplastic Barrett's oesophagus. The bar represents 1 mm measurement. Areas of interest are shown on the right with white lines showing optical transitions which may indicate histological changes.

quite nauseous and have become mildly hypotensive. The ALA is administered between 4 and 6 h before transfer to endoscopy for endoscopic light delivery. Patients with no macroscopic abnormality with smooth appearing Barrett's oesophagus with no suggestion of deeper disease are treated with ALA.

Method of treatment

Precautions are taken to avoid excessive light exposure in transportation of the patient to the endoscopy suite, and exposure to bright operating lights. Treatment is performed with topical anaesthesia with Xylocaine pharyngeal spray and intravenous sedation of between 1 and 10 mg of midazolam. In our practice we have found that analgesia is occasionally administered (Pethidine 50–100 mg intravenously). Patients photosensitised using ALA often require a prolonged endoscopic light administration. It is very noticeable that there is considerable local discomfort and irritation during light irradiation. The patients usually have no recollection of this but are sometimes very difficult to keep well sedated. The discomfort is manifest with considerable agitation and restlessness during the procedure. Throughout treatment, oxygen is delivered via a nasal sponge at a rate of 4–5 l a minute. Repeat sedation may be necessary as the light treatment can last 20–30 min.

The treatment times are considerably shorter for Photofrin (10–15 min) and tetra (*m*-hydroxyphenyl)

chlorin mTHPC (2 min) photosensitisation than for ALA, and there appears no problem of discomfort. It is very important to pay close attention to light dosimetry and use an appropriate light-centring device. The calculation of light dosimetry and delivery is performed by medical physics [6]. The aim is to deliver an even light dose to a defined circumferential area of the organ. To treat long areas repeated sequential areas are irradiated in 5 cm lengths in subsequent sessions. We use a windowed balloon (Wilson Cook Inc). The balloon is passed over a guide wire. Inhomogeneous and inconsistent illumination is a problem and occurs due to oesophageal motility, respiratory movements, and patient restlessness. To avoid this we pass a small video endoscope down beside the device to ensure positional stability throughout treatment.

The laser fibre with 7 cm diffusing tip is inserted and the correct wavelength of light chosen PpIX (generated from ALA)-630 nm, Photofrin-630 nm, and mTHPC-652 nm. We do adjust the position of the irradiating device during the procedure so the laser treatment is interrupted on occasion. The power density used for Photofrin is between 25 and 30 mW/cm² (100–400 mW/cm) to provide an energy density of 25 J/cm² from the diffuser to the mucosa [6]. Treatment can be performed on an outpatient basis and patients with Barrett's oesophagus must all receive profound acid suppression with proton pump inhibitor therapy. Direct sunlight and other intense lights must be avoided for a period of 4–6 weeks, and an advice sheet is given.

Post-operative care and follow-up

During the clinical trial [4], we performed a second endoscopy 2 days after the initial treatment session to irradiate any areas that had been missed at the first treatment. It is now not our practice to do this, we repeat a diagnostic endoscopy 4–6 weeks after the initial treatment.

The patient is counselled that they will often feel no effect until days 3–7 after photodynamic therapy when retrosternal pain and discomfort is often apparent. This is treated with antacids. The patients are asked to report any persistent difficulty in swallowing since stricture formation may occur. All patients have continuous endoscopic surveillance initially at 3 months and then yearly when two consecutive endoscopies with multiple biopsies have shown no dysplasia. If further treatment is required it is performed at 3 monthly intervals.

References

- [1] Schlemper RJ, Riddell RH, Kato Y, et al. The Vienna classification of gastrointestinal epithelial neoplasia. *Gut* 2000;47:251–5.
- [2] Montgomery E, Bronner MP, Goldblum JR, et al. Reproducibility of the diagnosis of dysplasia in Barrett's oesophagus: a reaffirmation. *Hum Pathol* 2001;32:368–78.
- [3] Ormsby AH, Petras RE, Henricks WH, et al. Observer variation in the diagnosis of superficial oesophageal adenocarcinoma. *Gut* 2002;51:671–6.
- [4] International Photodynamic Group for High grade Dysplasia in Barrett's Esophagus et al. Photodynamic Therapy with porfimer sodium for ablation of high-grade dysplasia in Barrett's esophagus: international, partially blinded, randomized phase III trial. *Gastrointest Endosc* 2005;62:488–98.
- [5] Hauser T, Szeimies RM, Scholmerich J, et al. Side effects of 5-aminolevulinic acid photosensitization in photodiagnosis of precancerous lesions of the gastrointestinal tract. *Gastroenterology* 1999;114:G3092.
- [6] Stone N. Standardizing dosimetry in esophageal PDT: an argument for use of centering devices and removal of misleading units. *TCRT* 2003;2:76–84.

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

FEATURED ARTICLES

Electrostatic Forward-Viewing Scanning Probe for Doppler Optical Coherence Tomography using ...

Ryerson University (Mar 29 2008) [Cardiology](#), [Endoscopy](#)

Optical Coherence Tomography Imaging as a Quality Assurance Tool for ...

The Annals of Thoracic Surgery (Mar 29 2008) [Cardiology](#), [Endoscopy](#)

Background: Concerns about intimal disruption and spasm have limited enthusiasm for endoscopic radial artery harvest (ERAH), although the risk of these problems after this procedure remains uncertain. Radial artery conduits were screened intraoperatively before and after ERAH vs open harvest ...

Artifact removal in Fourier-domain optical coherence tomography with a piezoelectric ...

opticsinfobase.org (Mar 28 2008)

Sébastien Vergnole, Guy Lamouche, Marc L. Dufour. We describe an artifact removal setup swept-source optical coherence tomography (OCT) system that enables high-speed full-range imaging. We implement a piezoelectric fiber stretcher to generate a periodic phase shift between successive A-scans, thus ...

Endoscopic Optical Coherence Tomography.

(Mar 28 2008) [Endoscopy](#)

We develop advanced MEMS probes for endoscopic optical coherence tomography (OCT) to obtain high quality 3D imaging of biological tissues.

Neointimal coverage of sirolimus-eluting stents 6 months and 12 months ...

Chinese Medical Journal (Mar 28 2008) [Cardiology](#)

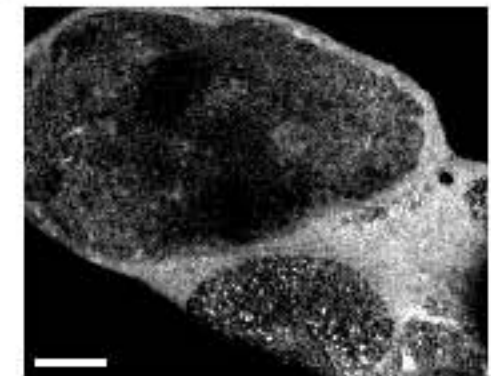
Background Optical coherence tomography (OCT) is a new imaging modality with resolution of approximately 10 μm and can be employed to visualize intracoronary characteristics. Sirolimus-eluting stents (SES) are susceptible to late thrombosis due to delayed re-endothelialization over the stent struts, ...

Optical coherence tomography evaluation of posterior capsule opacification related to ...

ScienceDirect (Mar 26 2008) [Ophthalmology](#)

Purpose To evaluate posterior capsule opacification (PCO) and the impact of different

IMAGE OF THE WEEK



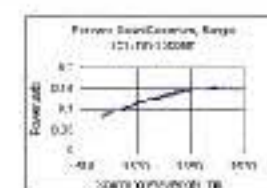
En face ex vivo OCT image of a bivalved lymph node obtained with the Michelson Diagnostics EX1301 OCT microscope at 1300 nm as part of a study to assess the merit of OCT for guiding histopathology. Image reconstructed from a data volume using a volume viewer. Tissue specimen subjected to histopathology for lymph node staging. The fibrous capsule shows that lymph node is available in its entirety. The sample also contains connective and adipose tissue, which is not of interest for this histopathologic assessment. The image measures 6.6mm x 5.2 mm at 7.5 μm resolution (Scale bar 1mm). Image courtesy of Mr Florian Bazant-Hegemark, Clinical Scientist at Gloucestershire Royal Hospital NHS Foundation Trust.

Submit your figure, image, or video via Feedback link below.

ORGANIZATIONS IN THE NEWS

Volcano
Heidelberg Engineering
University of Western Australia
Carl Zeiss, Inc.
Optovue
Heidelberg Retina Tomograph
Optoelectronics and
Measurement Techniques
Laboratory
Bascom Palmer Eye Institute
Massachusetts Institute of
Technology
Zeiss Meditec

PICTURE GALLERY



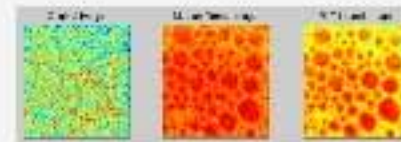
Rotating Kernel Transformation

Download Now:

Rating: N/A

Code Metrics: [Full report](#)

[What is this?](#)



[View fullsize image](#)

Author: [Florian Bazant-Hegemark](#)

Summary: Filter emphasising the shape of the kernel (eg straight line) within an image

MATLAB Release: R12.1

Required Products: Image Processing Toolbox

Description: The Rotating Kernel Transformation convolves an image with several orientations of a kernel. For a line-kernel, straight lines are emphasized. For larger kernels, it becomes computationally expensive.

- The zip-archive contains 5 files:
- 1) rktfilt.m ... function for filtering
 - 2) rktkern.m ... for creating a kernel
 - 3) rkttest.m ... script for testing the example
 - 4) rktexample.mat ... sample image
 - 5) example.png ... example screenshot

Please use 'help rktfilt' or 'doc rktfilt', 'help rktkern' or 'doc rktkern' for help.

Submit a File
Now accepting MATLAB Published M-files.

- [Author Index](#)
- [Most Recent Files](#)
- [Most Downloaded](#)
- [Screenshot Index](#)
- [Most Recent Comments](#)
- [Files By Product](#)
- [Metrics and Reports](#)

▪ [Download this File](#)

Notify me when changes are made to this file

E-mail:

[How do I unsubscribe?](#)

E-mail this page to a colleague

File Details

File Id:	14977	Average rating:	0.0
Size:	167 KB	# of reviews:	0
Submitted:	2007-05-14	Downloads:	33
Subscribers:	0		
Keywords:	Rotating Kernel Transformation, speckle, noise reduction, feature enha		
Zip file contents:	rktexample.mat, rktfilt.m, rktkern.m, rkttest.m, example.png		

User Reviews

Number of Reviews: 0

Be the first to rate this file!

Review this Submission

[Guidelines for Reviewing a Submission](#)

Bold Indicates Required Information

Overall Rating:

General Comments