Automated Retinal Layer Segmentation and Pre-Apoptotic Monitoring for Three-Dimensional Optical Coherence Tomography

Vedran Kajic

DOCTOR OF PHILOSOPHY

SCHOOL OF OPTOMETRY AND VISION SCIENCES

CARDIFF UNIVERSITY

2011

Acknowledgements

First, I would like to thank my first supervisor, Wolfgang Drexler, for providing me with his expert guidance, financial support and a lot of patience. Also, I want to thank my other two supervisors from the Computer Science department, David Marshall and Paul L. Rosin, who gave me all the computer science related support and advice, good recommendations and have arranged with Professor Drexler my PhD in Cardiff. Gavin Powell, also from the Computer Science department, helped me in the transition period at the start of the first year, provided me with early stage software and helpful explanations.

I would like to thank Boris Považay from the biomedical imaging group who was there whenever I was in need of any help, who gave me advice and many thought-provoking comments. Bernd Hofer from the group gave me plentiful signal processing advice and thorough theoretical explanation of principles of operation of various OCT systems. Boris Hermann has provided me with valuable support regarding various software and technical issues.

Working with Debbie Tudor and Irina Erchova on the apoptosis project has been a very useful and interesting experience that enabled me to see how computer science and OCT fit into multidisciplinary research that can enable invaluable biological insight.

I would also like to thank all other member of the group who helped me settle in, made me feel welcome and helped me both professionally and personally: Cristiano Torti, Alex Tumlinson, Ling Wang, Yaiza Garcia-Sanchez, Aneesh Alex, James Fergusson and Sara Rey.

Abstract

Optical Coherence Tomography (OCT) is an emerging biomedical imaging method that allows unprecedented, three-dimensional (3D) sub-surface insight into living tissue. In order to make these large 3D OCT data sets clinically useful it is necessary to objectively extract quantitative information. Therefore it is necessary to develop effective computer post-processing algorithms that include registration (alignment) of two-dimensional cross-sections (B-scans) of an acquired volume and automated segmentation (i.e. thickness quantification) of visualized layers of the investigated tissue. Registration is important since it allows for enhanced volume reconstruction, and enables speckle reduction and signal-to-noise (SNR) improvement by averaging through the stack (Drexler and Fujimoto, 2008). Segmentation is necessary to objectively detect subtle changes for early disease diagnosis or therapy monitoring.

The aim of this PhD thesis was to develop segmentation algorithm adapted and optimized to retinal OCT data that will provide objective 3D layer thickness which might be used to improve diagnosis and monitoring of retinal pathologies. Additionally, a 3D stack registration method was produced by modifying an existing algorithm. A related project was to develop a pre-apoptotic retinal monitoring based on the changes in texture parameters of the OCT scans in order to enable treatment before the changes become irreversible; apoptosis refers to the programmed cell death that can occur in retinal tissue and lead to blindness. These issues can be critical for the examination of tissues within the central nervous system.

It should be noted that most of the early texture segmentation work has been done before the start of this PhD, by a computer science colleague Gavin Powell, but is included since it has been continued and optimized by myself within the first year of the PhD and for completeness of this work. In the remainder of the PhD a novel statistical model for segmentation has been created and successfully applied to a large data set. A broad range of future research possibilities into advanced pathologies has been created by the results obtained. A separate model has been created for choroid segmentation located deep in retina, as the appearance of choroid is very different from the top retinal layers. Choroid thickness and structure is an important index of various pathologies (diabetes etc.).

As part of the pre-apoptotic monitoring project it was shown that an increase in proportion of apoptotic cells *in vitro* can be accurately quantified. Moreover, the data obtained indicates a similar increase in neuronal scatter in retinal explants following axotomy (removal of retinas from the eye), suggesting that UHR-OCT can be a novel non-invasive technique for the *in vivo* assessment of neuronal health.

Additionally, an independent project within the Computer Science department in collaboration with the school of psychology has been successfully carried out, improving analysis of facial dynamics and behaviour transfer between individuals. Also, important improvements to a general signal processing algorithm, dynamic time warping (DTW), have been made, allowing potential application in a broad signal processing field.

The four major computer science contributions are:

- dynamic programming shortest path algorithm that incorporates path derivative and edge orientation criteria (based on directional wavelet analysis), implemented efficiently to preserve the same time complexity O(E) as the standard shortest path approach (Section 7.3.1). The fundamental difference between this approach and others is that the algorithm works with limited path history, instead of only using static weights per edge to determine the distance function. Thus there is no need for complex and slow workarounds, as when trying to incorporate path information with static weights only.

- statistical shape and texture model developed for image segmentation; compared to the standard active appearance model (AAM) it features a normalized term added to the objective function that uses the shape prior information obtained by a simpler thresholding approach of prominent structures in an image. This approach can be generally applied to any image segmentation problem when certain information can be obtained prior to applying the full statistical model. The second stage does column-wise refinement (A-scan optimization) based on a separate model built by back projecting the training data through the full 2D shape model and using column wise offsets as input to the training of the A-scan model. This stage allows the model to be precise while at the same time being robust. For choroid segmentation a simple but effective objective function was used that maximizes the area covered by dark blobs found by using the maximum stable extremal region (MSER) algorithm as the blob detector (Sections 7.2.4 and 7.3.2).

- applied to pre-apoptotic analysis, efficient 3D texture features were derived, such as directional, but point of view axis rotationally independent, wavelet features; additionally, feature reduction criteria based on Bayesian Information Criteria (BIC) was proposed (Section 8.2).
- during the course of an independent project on facial dynamics (included in the appendix) a general improvement has been made to the dynamic time warping (DTW) algorithm, the standard algorithm for aligning signals. Real derivative information has been added, computed dynamically while preserving the same time complexity. Another addition is the graph based DTW; a shortest path approach that utilizes similar signals available in a dataset to improve the registration accuracy between the target and source nodes (Appendix).

The major medical contributions are:

- the retinal segmentation algorithm allows for monitoring of healthy eyes and early pathologies with an important distinction being, compared to the existing approaches, that it is robust to noise and other artefacts (e.g. shadows, out of focus retinal position) that often occur when applying OCT in clinical setting.
- a choroid segmentation algorithm has been developed that allows for the first time fully automatic segmentation (no other approach has been published so far). This enables clinicians to study the relationship between changes in choroid thickness and vascularisation and various pathologies. Diabetes is one of the important applications, while many other links to pathologies still remain to be fully explored; one of the reasons why this hasn't been possible earlier was the lack of an automatic segmentation algorithm.
- an apoptosis (programmed cell death that can occur in retinal tissue and lead to blindness) monitoring algorithm has been developed. Pre-apoptotic detection is critical to preventing blindness before the process becomes irreversible.

Contents

1 Introduction

In the following sections, OCT will first be explained in more detail followed by a critical literature review of available registration and segmentation algorithms with special focus on a variety of OCT data. Introduction also covers data pre-processing, machine learning, texture analysis, optimization and model building algorithms since they are important prerequisites to the successful registration, segmentation and pre-apoptotic monitoring.

In the second chapter, the main focus of the thesis, the segmentation and segmentation related work is presented.

It is divided into: work on registration algorithm (2.1) as it related to both the retinal (2.2) and choroidal segmentation (2.3). At the current stage the segmentation works on each image in a stack independently, thus registration is not really necessary. However, extending the segmentation to 3D would make it essential. Additionally, to interpret the result, create thickness maps etc. it is important to have a stack registered at least approximately. Thus, in the chapter 2.1 early work on registration is presented, dealing with combining existing algorithms (rigid body and elastic) into a package applicable to OCT data registration, applicable to a whole stack in automatic fashion.

The 2.2.1 subchapter presents an adaptive thresholding approach used for top and bottom foveal boundary detection, that sets a starting point for segmentation of the other inner layers.

In the following subchapter 2.2.2 early work on data segmentation based purely on texture analysis is explained, based on work by Gavin Powell.

Subchapter 2.2.3 deals with a shape fitting, adaptive template matching approach, as it was found that due to poor texture only discrimination between layers, texture analysis alone was insufficient. However it was found that this approach still did not give satisfactory results.

In subchapter 2.2.4 the final method that unifies both the texture and shape segmentation approaches, based on machine learning principles (utilizing the training data) is presented.

For segmentation of choroid affected often by extreme pathologies, it was necessary to develop a more robust top and bottom boundary detection algorithm. How it works and why it is applicable in large problem area outside OCT is explained in subchapter 2.3.1.

In subchapter 2.3.2 choroid segmentation is explained using a statistical model similar to that developed for the foveal segmentation (2.2.4)

In the third chapter, the pre-apoptotic monitoring method based on texture analysis is presented.

It is summed up in the Summary how all the methods that have been employed so far have lead to a unified model based solution. In the final chapter, an outlook on potential future work is presented.

In the appendix a detailed overview is given of the computer science based, facial dynamics project (published conference paper).

1.1 Optical Coherence Tomography (OCT)

OCT is an emerging non-invasive optical imaging modality that can perform high resolution (1 – 15 µm), cross-sectional imaging in materials and biological tissues by measuring the echo time delay and magnitude of back reflected or backscattered laser light (Drexler and Fujimoto, 2008). In medical applications the main advantage of OCT is that imaging of tissue structure is possible in situ and in real time, without the need to remove specimens as in conventional histopathology. In OCT, part of the beam directed to the sample will be absorbed or will multiple-scatter into the tissue, while part will be backscattered or reflected. The latter part will be acquired by the detector and using the interferometric property of the combined sample and reference beam that depends on the optical path difference, it is possible to reconstruct one-dimensional optical depth resolved back-scattering profile, called an optical axial scan (or optical A-scan, in analogy to ultrasound imaging). By scanning the optical beam in the transverse direction and performing adjacent axial scans along this direction it is possible to construct a two dimensional cross-section (or B-scan). By scanning along a longitudinal axis as well, it is possible to visualize volumes of the investigated tissue. Resolvable volume depth is limited to 1-2 mm by OCT penetration depth, which depends on absorption and multiple scattering and hence on the central wavelength used. In general, longer wavelengths will give higher tissue penetration due to reduced light scattering but high axial OCT resolution necessitates extremely broad optical bandwidth since OCT axial resolution l_c is mainly determined by the optical bandwidth of the used light source; $l_c = \frac{2\ln 2}{p}\frac{l_0^{\,2}}{\Delta l}$. Normally, superluminescent diodes are used for OCT because they are compact, user-friendly and relatively inexpensive. However, the optical bandwidth they provide is limited to 70-90 nm at full-width-at-half-maximum. Femtosecond lasers are used for ultrahigh resolution (UHR) OCT, because they provide optical bandwidth of up to 300 nm centred at 800 nm and therefore enable a significant increase in the axial OCT resolution. Comparison between in vivo standard and ultrahigh resolution OCT of the normal human retina is depicted in Figure 6 Simulation of the speckle noise created by the coherent detection process in 1D (Schmitt et al., 1999). a) shows the distribution of scatterers, with the incoherent signal "imagining" giving a close approximation to the original distribution. However, coherent light, such as used in the OCT, produces a bad approximation c) with the envelope shown in d)

Even though speckle is usually considered to be a corrupting factor in OCT images, it can also provide some insight into the tissue composition. Most of the algorithms are focused on the idea of treating speckle as noise and removing it as much as possible without losing too much of the original information. Several different algorithms have been applied to the speckle removal in OCT images. Standard pre-processing techniques are mean and median filtering. Median filtering has an important edge preserving property. Wiener filter might be used to reduce the noise in the signal if a reasonably accurate assumption about the spectral content of the signal and the noise can be made in advance. Usually the requirement is that the filter has to be physically realisable resulting in the casual solution. Performance criteria used is minimum mean-square error. Common use of Wiener filter is in Wiener deconvolution. It will suppress frequencies that have poor signal to noise ratio. Even though the exact noise and signal spectra are not known in advance, reasonable spectra assumptions can often be made. For instance, in the image processing applications it can usually be assumed that most of the signal is contained in the lower frequencies while the noise has usually a flat frequency spectrum. Kulkarni et al. built a linear shift invariant model describing coherent specimen-light interactions in OCT (Kulkarni, Thomas and Izatt, 1997). Based on this model, an iterative deconvolution algorithm was demonstrated for enhancing the sharpness of biological structures in OCT images. An example of algorithm developed for another purpose and then applied to the OCT field is CLEAN (Schmitt, 1998). An iterative point-deconvolution algorithm originally intended for use in radio astronomy, it was successfully used for restoring signal in OCT images and allowing observation of tissue morphology previously hidden under noise. Its performance degrades reasonably as the number of irresolvable scatterers increases. Derived deconvolution kernel is based on the theoretical point-spread function of an OCT scanner and depends on the properties of the imaging system and the scattering properties of the medium. It incorporates a modification based on inverse Wiener filter designed to reduce ripple artefacts in the images of tissue containing densely packed scatterers. The sticks method adapted from ultrasound imaging is based on fitting short linear segments to the boundaries of the vessels for example. Rotating kernel transformation has been applied to coronary images, based on matching edges in the image to a set of binary-valued two-dimensional square kernels. Undecimated wavelet filter transformation uses the logarithm of the magnitude to remove the uncorrelated speckle features at multiple scales. Logarithm scale is used since the multiplicative noise becomes additive noise in the logarithm scale. Another method that has been developed for this purpose simultaneously constrains the image magnitude to match a blurred version of the image in a least square sense. The constraint used is relative entropy called the I-divergence. Despeckling result of this method is shown in Figure 7.

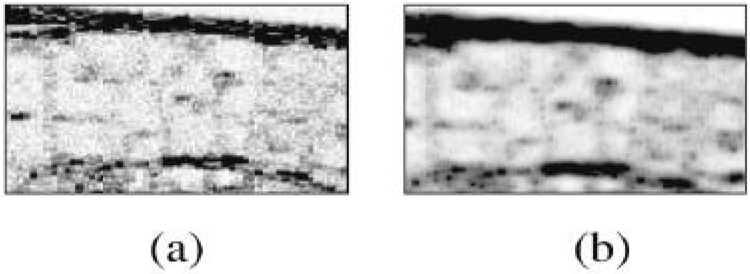


Figure 7 Comparison of the original and the despeckled image using I-divergence method (Drexler and Fujimoto, 2008)

Since in OCT images are usually formed from the envelope of the measured interference signal, computation of the absolute magnitude of the signal is a nonlinear process that destroys phase information. A speckle-reduction technique that works in the complex domain, zero-adjustment-procedure (ZAP), has been introduced (Yung, Lee and Schmitt, 1999). However, ZAP tends not only to reduce speckle but also to blur sharp edges between the image features. Anisotropic diffusion filtering in scale-space with the diffusion coefficient chosen in a way to encourage intraregion rather than interregion smoothing was presented (Perona and Malik, 1990). Scale-space is a kind of multiresolution image representation with certain feature preserving properties. Gilboa et al. demonstrated further improvements in denoising by extending the analysis from real to complex domain and combining the diffusion equation with the free Schrödinger equation (Gilboa, Sochen and Zeevi, 2004). Additional property is that the imaginary value can be used as a robust edge filter. These methods have been further extended by making the diffusion process truly anisotropic, in a sense that the diffusivity varies not only with the edge location but with the orientation as well (Joachim, 1999). The main advantage of this approach is that the diffusion rather acts along coherent structures and thus provides an efficient denoising method since noise has no preferred orientation. Fernandez et al. combined the complex diffusion and coherence-enhanced diffusion filtering and successfully applied it to the OCT data pre-processing as a step before the peak based image segmentation (Cabrera Fernández, Salinas and Puliafito, 2005).

1.1.1 Quantization

Quantization refers to mapping raw data to a series of fixed values. There are three reasons for signal quantization (Yu et al., 2001). The first reason relates to image compression, transmission and digital storage. Another purpose is to adapt the data for the human vision system and in this case subjective visual impression is more important than the absolute distortion. The third reason is for pixel level transformation. Four commonly used methods are equal-interval, equal-probability, minimum-variance and histogram hyperbolization. Equal-interval is a simple linear transformation from the original signal range to the new range. Equal-probability, or often referred to as histogram equalization, creates intervals, or bins, of variable sizes in a way that all bins contain equal number of pixels. Minimum-variance quantization creates the intervals so that the weighted sum of the variances of the intervals is minimized. Histogram hyperbolization adapts the image to the nonlinearity of the human vision system. The logarithm scale is usually used in OCT field as a mapping from the complex domain raw data to the final image. However, some detail will inevitably be lost in this process and more attention should be paid to the choice of the appropriate method. The simplest one is Direct Logarithm (DL) method. Truncation Logarithm (TL) method considers both the dynamic range determination and the noise present in the signal. TL method uses a fixed threshold to reduce the noise but it also degrades image quality. The main reason for choosing a logarithm-based method is to compress the dynamic range in respect to the exponential law that governs the attenuation of light in scattering materials. However, logarithm-based methods often suffer from poor contrast and loss of details in the image. Three quantization methods presented in (Yu et al., 2001) are Minimum Distortion (MD) and Truncated MD (TMD), Information Expansion (IE) and finally Maximum Entropy. MD method is based on the minimum distortion principle. MD method becomes equivalent to minimum-variance method when mean-square measure is used as a measure of distortion. TMD method can be used on data that contains large peaks. IE method is similar to histogram equalization except that it works on the raw data. The advantage lies in the fact that histogram equalization reduces the image entropy while entropy of the raw data remains unchanged during equalization. ME method is focused on preserving the hidden information in the raw data. A standard measure from the information theory that measures mutual dependability of two variables, mutual information, is used as a maximizing variable. When mutual information of data before and after quantization is maximal, loss of the information is minimal. Since the quantization function is deterministic, maximization of mutual information is equal to the maximization of the image entropy after the quantization and hence the name of the method.

1.2 Apoptosis and Pre-Apoptotic Monitoring

Apoptosis or "programmed cell death" is a highly organised form of cell death, that although essential for the maintenance of normal tissue homeostasis, has been implicated in a number of neurodegenerative diseases including Parkinson's disease, Alzheimer's disease and spinal muscular atrophy (Mattson, 2000), as well as in retinal neuronal pathologies including glaucoma, diabetic retinopathy (De Souza-Ramalho et al., 2009) and age-related macular degeneration (Dunaief et al., 2002). The quantification of cellular health is essential for benchmarking responses to experimental or therapeutic interventions.

The most recent advances in *in vivo* monitoring of cells and biological structures include development of confocal scanning laser ophthalmoscopy (cSLO) (Cordeiro et al., 2004; Schmitz-Valckenberg et al., 2008) and *in vivo* confocal neuroimaging (ICON) (Prilloff et al., 2010). These methods that still rely either on external ligands, for example the use of annexin V in oncology (Rottey et al., 2006; Rottey et al., 2007; Wong et al., 2008) and retinal research (Cordeiro et al., 2004; Reichstein et al., 2007; Schmitz-Valckenberg et al., 2008; Prilloff et al., 2010), or on the development of new contrast agents for the use in magnetic resonance imaging (MRI) (Yoo and Pagel, 2008; Terreno, Castelli and Aime, 2010). However, development of a non-invasive, non-toxic technique is key to the advancement of *in vivo* cell and tissue imaging.

A number of optical methods were developed to stage cells as they progress through apoptosis based on detailed knowledge of subcellular changes (Palade, 1953; Cereghetti and Scorrano, 2006). These ligand-based techniques are routinely used for histochemical staging of apoptosis including phosphatidylserine localisation in the outer plasma membrane (Vermes et al., 1995; Blankenberg et al., 1998), mitochondrial dysfunction(Green and Reed, 1998), activation of caspases (Thornberry, 1997, 1998), DNA fragmentation (Robertson, Orrenius and Zhivotovsky, 2000) and loss of membrane integrity. Yet while these technologies are robust for *in vitro* analyses, their use in *in vivo* studies can be challenging due to their potential toxicity, requirements for ligand delivery, and bias towards events that coincide with the commitment of a cell to death.

In view of these limitations the development of *in vivo* ligand-free approaches are of special interest. For example, apoptotic cells can be distinguished from healthy cells by flow cytometry as indicated by changes in forward scatter and side scatter properties (Vermes, Haanen and Reutelingsperger, 2000). Coherent anti-Stokes Raman scattering spectroscopy (CARS) also show promise but is limited by the need for a high numerical aperture and low signal to noise ratio (Wright et al., 2007; Balu et al., 2010). Currently, these methods are widely used for the analysis of cell cultures but are of limited use for monitoring of cellular events in living tissue.

Apoptosis is a cellular event involving many organelles including the mitochondria, endoplasmic reticulum, golgi and the nucleus, all of which can generate optical signal changes detectable by UHR-OCT. The mitochondria are key initiators of the apoptosis pathway. During apoptosis the mitochondrial membrane becomes “leaky” allowing the release of a number of pro apoptotic molecules including cytochrome c, leading to the activation of the caspase cascade and ultimately cell death (Frank et al., 2001; Cereghetti and Scorrano, 2006). Prior to this mitochondrial membrane dysfunction, neurons are subjected to prolonged periods of compromise during which dendrite pruning (McGowan, 2006) and remodelling occur prior to loss of the cell body. During this period of compromise, before the breakdown of the mitochondrial membrane, the mitochondria change morphology where they cease to be part of a continuous network, round off and become discrete (Karbowski and Youle, 2003). The dimensions of these discrete organelles are close to the wavelength of the light used in the 800 nm UHR-OCT system so the scattering coefficient will be maximal (Palade, 1953). This, in combination with the fact that mitochondria have a higher refractive index than the cytosol due to the high concentrations of proteins within them makes them ideal targets (Mourant et al., 1998) for the pre-apoptotic detection. OCT has been previously used to detect in vivo changes in the neuronal population of mouse cortex, thought to arise from light scattering by the mitochondria (Mourant et al., 1998), and other light scattering techniques have detected mitochondrial changes in in-vitro cells undergoing malignant change (Gourley et al., 2005).

The requirement for use of external ligands is overcome by recent developments in UHR-OCT suggesting that it may finally be possible to study localized light scattering within tissues and therefore map indices of cellular health to tissue structures. In particular, it has been shown that UHR-OCT is able to detect differences in light attenuation between apoptotic versus necrotic cells (van der Meer et al., 2010).

1.3 Machine Learning and Texture Analysis

Since the algorithms that are going to be described later in this thesis depend on the supplied training data to tune their parameters it is necessary to give a basic overview of the machine learning algorithms in general. Machine learning algorithms can be divided in two major classes of methods: supervised and unsupervised learning.

Supervised learning is a technique for learning a function based on training data, which consists of pairs of input objects, and desired outputs. The output of the function can be a continuous value, or a discrete class label. The task of the supervised learner is to predict the value of the function for any valid input object after having seen a number of training examples.

Unsupervised learning is a type of machine learning where manual labels of inputs are not given. The most common type of unsupervised learning is clustering. This classifies objects into different groups, so that the data in each subset share some common trait, usually proximity according to some defined distance measure. The task of classifying the data set into *k* clusters is often referred to as *k-clustering*.

It is difficult to give a comprehensive definition of what texture is. One approach is that a region in an image has a constant texture if a set of local statistics or other local properties of the picture function are constant, slowly varying, or approximately periodic (Chen, Pau and Wang, 1993). The goal of texture based classification is to produce a classification map of the input image where each uniform textured region is identified with the texture class it belongs to. Image texture is defined as a function of the spatial variation in pixel intensities. One immediate application of image texture is the recognition of homogeneous image regions using texture properties. In general, from each image, texture features are produced by applying various filters (mean, co-occurrence matrix, Laws', Haralick, Gabor, wavelets, etc.) of different window sizes and mapping every pixel from the original image into the created *n-dimensional* feature space. Another useful class of texture features are local binary patterns (LBP) that are especially suited for cases when the texture pattern is of interest and smaller intensity variations are to be ignored (Timo et al., 2002). LBP divides an image into cells. For each cell it looks at every pixel in that cell and assigns 0 or 1 to its neighbours, depending on whether they have a lower or higher intensity. In case of a typical 8-neighbourhood, an 8 bit number is produced. These are aggregated into a histogram, representing texture properties of a given cell.

Different methods are then used to separate classes in the feature space and using that information to classify new images pixelwise. Texture analysis is an important method that captures local information very well but it lacks global knowledge of the segmented structure which makes it sensitive to variations in the local appearance, such as introduced by speckle. Also, in case when structures that need to be segmented do not have a defining unique texture, as is the case with OCT layers, using contextual spatial information becomes necessary. It is important to mention that it is difficult to decide what feature set is optimal to use since the number of possible combinations of different feature types and window sizes is huge and it is computationally costly to use unnecessarily large set of features for the segmentation of new images. Thus, it is useful to perform a feature selection step in the training phase that will keep the best features from a possibly very large set of initial features. In this way the training phase of the algorithm will be lengthy, but once the optimal set of features has been found, segmentation of new images will proceed much faster. One method that can be used to perform this task, and was chosen in this work, is sequential forward search (Guyon and Elisseeff, 2003). It belongs to the class of supervised learning methods and works by iteratively expanding the optimal set by one starting from zero. This algorithm does not guarantee to give optimal set since choosing some of the features might give poor segmentation improvements individually, but they might work very well in combination. In most cases however, sequential forward search will give a good trade-off between computational cost of the training phase and the obtained reduced feature set which is usually close to the optimum.

To perform the actual separation of different classes in feature space several methods can be used. The most common choice is to use Mahalanobis distance, neural networks or support vector machines. These methods will be described in more detail.

Mahalanobis distance differs from Euclidean distance in that it takes into account correlations of the data set. It can be intuitively described as a hyperellipsoid (ellipsoid in *n-dimensional* space) that best represents the set's probability distribution and is estimated by building the covariance matrix of the samples. The Mahalanobis distance is simply the distance of the test point from the centre of mass of class distribution divided by the width of the hyperellipsoid in the direction of the test point. An illustration of the Mahalanobis distance in 2D case is shown in Figure 8.

A Neural network (NN) is a mathematical model or computational model based on biological neural networks. It consists of an interconnected group of artificial neurons and processes information where each connection between neurons has a weight, with the weights modulating the value across the connection. The training phase is performed to modify the weights until the network implements a desired function. Once training has completed, the network can be applied to data that was not part of the training set.

A Support vector machine (SVM) is a classification technique that not only separates data into different classes, but also leaves a maximal margin between those classes. Because of their maximum margin separation, SVMs can obtain better classification results, which minimize the risk of misclassification for data that is not shown in the training set. The maximal margin principle is shown on Figure 9 with the line H2 representing the maximal margin division. A useful technique that can be applied if the desired separation surface is of higher order (for example second or third order polynomial) maps the problem to a higher dimensional space and uses a hyperplane to mark the boundary. This hyperplane will correspond to the higher order surface in the original space. How to construct a soft margin hyperplane in cases when training data cannot be separated without errors is demonstrated in (Cortes and Vapnik, 1995).

With NNs, training is done repeatedly to obtain iteratively better results for classification and it is difficult to determine when to stop training or when the best network is generated. On the other hand, SVMs handle the entire training data set simultaneously and provide the optimal solution when the training is complete (Fan-Yin, 2005).

A large number of texture based segmentation methods have been published. Some of the representative and relevant approaches will be covered.

A six-stimulus basis for stochastic texture perception was presented in (Jan-Mark and Arnold, 2005). Fragmentation of the scene by a chaotic process (adding new objects to the scene randomly) causes the spatial scene statistics to conform to a Weibull-distribution, while the parameters of the Weibull distribution characterize the spatial structure of uniform stochastic textures of many different origins completely. The perceptual significance of the Weibull parameters was reported with perceptual parameters being: regularity, coarseness, contrast, roughness, and directionality.

Viola et al. achieved high speed texture segmentation using effective image representation (integral image), feature selection algorithm that filters thousands of simple features based on AdaBoost and a cascade approach to ignore the areas of the image which are not likely to contain objects of interest early in the process (Viola and Jones, 2001).

Gossage et al. presented good texture segmentation based results on mouse and fat tissue, as well as healthy and diseased mouse lung tissue (Gossage et al., 2006). They used energy, entropy, correlation, local homogeneity and inertia features. However, it is important to note that the speckle pattern differences in the differentiated tissue were visible by naked eye which is not the case with retinal layers.

Texture analysis has been successfully applied to resolve collagen tissue with and without endothelial cells and shows that tissues that contain scatterers of different sizes and distributions can be differentiated with reasonable accuracy using speckle information only (Kirk et al., 2006). However, not all retinal layers contain enough speckle information for texture analysis based segmentation to be successful alone. Results obtained with texture analysis so far are presented later on.

An interesting approach to the problem of segmentation has been presented in (Ojala and Pietikäinen, 1997). By splitting an image into small consistent regions and then merging them into larger areas, a higher level of spatial consistency has been achieved. Even though they performed the analysis on images with much more apparent difference in texture than is the case with OCT data, it is still a highly relevant approach.

1.4 Numerical Optimization Algorithms

Optimization algorithms are used to solve problems that are either impossible or impractical to solve analytically. They are essential building blocks of registration and nontrivial segmentation methods and thus have to be studied in detail. In this work they are used for minimizing a distance function in image registration algorithms and for minimizing an objective function that defines success in fitting an adaptive template or model to data.

An important optimization example is the least square problem which is used to model numerical data obtained from observations by adjusting the parameters of a model to get an optimal fit to the data. The best fit is that instance of the model for which the sum of the squared residuals has the minimal value, the residual being the difference between an observed value and the value given by the model. All of the optimization methods share the same principle of iteratively minimizing (or maximizing) the objective function. Since both data registration and segmentation approaches often fall into this domain it is important to give an overview of some of the many existing optimization methods. The number of individual optimization methods is huge so it is important to give a review of the different categories and which might be of interest to the practical applications presented in this work.

1.4.1 Classification of Optimization Methods

Objective functions can differ significantly in its nature. They can be linear, nonlinear or convex. A convex set is a set for which any line segment connecting any two points in the set lies in that set. A convex function is a function whose domain is a convex set and for any two points in this domain, its graph lies below the line segment. In respect to smoothness, functions can be differentiable or nondifferentiable. If the function is both smooth and convex, then its graph is bowl shaped. If the function satisfies these conditions, optimization problem becomes much easier. Most of the algorithms guarantee convergence to a stationary point in finite time but for convex, smooth functions it is also certain that this local optimum is global optimum as well.

Considering the reach of the optimization in the domain space, algorithms can be defined as local or global. Most of the algorithms focus on finding a local minimum, a point for which all the neighbouring points hold higher values. For the neighbourhood to be defined, the domain space has to be topological. Global algorithms try to find a global solution, a minimum in the whole domain space of the objective function. In the most general case it is impossible to guarantee that the global solution has been found. It is also very difficult to be certain that the minimum that has been found really is a global one. However, global optimization methods often demand certain type of objective function in order to make the search reasonably effective. In a special case, when the objective function is convex, local solution is equal to the global solution (Nocedal and Wright, 1999).

One important difference that can be made between optimization methods is whether they put constraints on the variables or not. Unconstrained optimization deal with problems that require no constraints on the variables or the constraints on the variables arise naturally and are usually safe to dsregard. Another common case when unconstrained problems appear is when constrained problems are reduced to unconstrained problems by replacing the constraints with penalizing terms that are added to the objective function. In general, unconstrained problems are easier to solve. Constrained problems arise when dealing with models that contain explicit constraints on the variables. The set of constraints can either be equalities or inequalities that again can be simple bounds on the variables or complex inequalities among the variables often represented by a matrix. Set of constraints which are active in respect to the current point is called Active set. At each iteration, the active set class of constrained optimization algorithms searches for and removes inactive constraints and then approximately solves the equality problem defined by the active set.

Optimization methods can also be separated based on the domain of values that the variables can hold, continuous or discrete. For example, in some problems variables can hold only integer values. If the discrete domain fact is ignored and the problem is optimized by allowing variables to hold real values and then rounding up is performed at the end, solution is by no means guaranteed to be optimal. This problem is properly handled by adding a new constraint to the set of constraints and forcing the variable set to hold discrete values (Nocedal and Wright, 1999). In general, continuous problems are easier to solve because the smoothness of the function allows for use of objective and constraint information at some point to deduce information about the point neighbourhood as well.

Depending on the certainty of the model definition, optimization algorithms can be deterministic or stochastic. Deterministic algorithms are applied to completely defined models, when no random behaviour is expected. In many situations behaviour of the objective function cannot be completely predicted. However, probabilities can be usually applied to different scenarios. In many economic and financial planning models this situation occurs. In this problem domain best solution is to use stochastic optimization methods. Stochastic optimization algorithms can be also used in solving deterministic problems if difficult nonlinear, constrained objective function has to be optimized and some benefits are expected from the random behaviour of the optimization algorithm in hope of avoiding local minima. It is important to define certain concepts regarding stochastic optimization since they will be used later on when describing some stochastic algorithms in more detail. Monte Carlo methods are a broad term that refers to a class of computational algorithms that rely on repeated random sampling to compute their results and are usually used when it is infeasible or impossible to compute an exact result with a deterministic algorithm. Stochastic optimization methods are examples of Monte Carlo methods and as such are based on random walks. The algorithm moves around a marker in a multi-dimensional space, tending to move in directions that lead to a lower function, but occasionally moving against the gradient. A Markov chain is a stochastic process with the Markov property, which means that given the present state, future states are independent of the past states. In other words, the description of the present state fully captures all the information that may influence the future evolution of the process while the future states are reached through a probabilistic process instead of a deterministic one. At each step the system may change its state from the current state to another state through any of the applicable transitions, or remain in the same state, according to a certain probability distribution of the transition probabilities. An example of a Markov chain is a simple random walk where the state space is a set of vertices of a graph and the transition steps involve moving to any of the neighbours of the current vertex with equal probability. Markov chain Monte Carlo (MCMC) algorithms are used for sampling from probability distributions based on constructing a Markov chain that has the desired distribution as its equilibrium distribution. Equilibrium distribution is a vector for which holds that a probability of state S is equal to the sum of multiplies of probabilities of all states that S is connected to and their transition probabilities. The state of the chain after a large number of steps is then used as a sample from the desired distribution with the quality of the sample improving as a function of the number of steps.

Another distinction that can be made among the optimization algorithms is how large are the sets of variables that they can handle effectively. Some algorithms can be defined as medium-scale and are not effective when used with large variable sets, thousands or hundreds of thousands of variables. Large-scale algorithms should be used in such cases.

In respect to the iterative approach that is taken, optimization methods fall into Line search or Trust region classes. Line search algorithms determine first the direction of the step to be taken and will then determine the optimal size of the step. On the other hand, trust region algorithms first define an initial area of search and will collapse or expand it depending on how well this area approximates the model function, which is usually a quadratic.

Optimization algorithms use different strategies for finding the optimal solution. Some will rely only on the objective function value at a certain point; others will use first or second derivatives, while some will even use the information from previous iterations. However, it should be common to all the good optimization methods that they are robust, efficient and accurate (Nocedal and Wright, 1999). It is usually difficult to achieve all of those goals since they may conflict so a trade-off is usually sought. In the following some of the most popular optimization methods will be shortly described, as well as those, which were considered as potentially useful to this work.

1.4.2 Deterministic Optimization

Gradient descent is one of the simplest and most intuitive optimization methods (Cauchy, 1847). If the function is differentiable it will decrease fastest in the direction of the negative gradient, so to find a local minima, steps proportional to the negative of the gradient of the function, or to the approximate gradient, are taken at the current point. However there are two significant weaknesses of gradient descent. Firstly, it may take the algorithm many iterations before it converges towards a local minimum, in cases the curvature in different directions varies a lot. Secondly, finding the optimal step length per iteration can be time-consuming. Conversely, using a fixed step length can yield poor results.

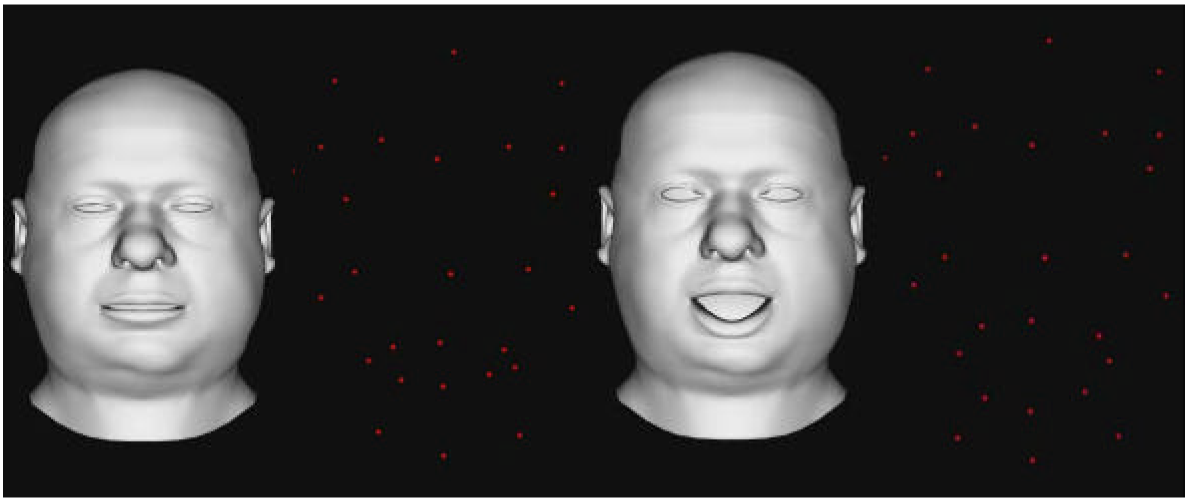
Newton's method is based on General Newton's method for finding roots of equations in one or more dimensions by iteratively finding roots of the tangent at the current point. It can be easily used to find local maxima or minima of functions by noticing that if any real number is a stationary point of a function, then that number is also the root of the derivative of the function, and therefore it is possible to find stationary points by applying General Newton's method to the derivative of the function. The geometric interpretation of Newton's method is that at each iteration function is approximated by a quadratic function around the point, and then a step is taken towards the minimum of that quadratic function. If the function happens to be a quadratic, then the exact minimum is found in only one step. Newton's method converges much faster towards a local minimum than the gradient descent. However, finding the inverse of the Hessian matrix required by the algorithm is an expensive operation.

Sequential Quadratic Programming (SQP) is a constrained optimization method that has the basic idea analogous to Newton's method for unconstrained minimization (Pshenichny and Danilin, 1983). At each step, a local model of the optimization problem is constructed and solved, yielding a step towards the solution of the original problem. However, instead of using Taylor expansion of the objective function as a quadratic approximation, Lagrangian is used instead, which gives better update results of both Lagrange multipliers and optimizing variable. Lagrange multipliers are used to force the optimization procedure to stay close to the given constraints.

Unlike Newton's method, in Quasi-Newton class of methods the Hessian matrix of second derivatives of the function to be minimized does not need to be computed at any stage. Hessian is updated by analyzing successive gradient vectors instead. Quasi-Newton methods are a generalization of the secant method to find the root of the first derivative to multidimensional problems. The secant method is a root-finding (function zeros) algorithm that uses a succession of roots of secant lines to better approximate a root of the function. In a multi-dimensional space, the secant equation is under-determined and members of a class of quasi-Newton methods differ in how they constrain the solution. The Broyden-Fletcher-Goldfarb-Shanno (BFGS) method for solving unconstrained nonlinear optimization problems is one of the most successful members of this class (Broyden, 1970).

The Levenberg–Marquardt Algorithm (LMA) provides a numerical solution to the problem of minimizing a function, generally nonlinear, over a space of parameters of the function (Levenberg, 1944). These minimization problems arise often in least squares curve fitting and nonlinear programming. LMA interpolates between the Gauss-Newton algorithm (GNA), which is a variation of the Newton's method that can only be used to minimize a sum of squared function values, and the method of gradient descent. This shifting is achieved by using a dynamic damping parameter that sets LMA closer to gradient descent or GNA. LMA is more robust than GNA, which means that in many cases it finds a solution even if it starts far off the final minimum. On the other hand, for well-behaved functions and reasonable starting parameters, LMA tends to be a bit slower than GNA. This algorithm was used for stack registration, Chapter 7.1.

Conjugate gradient method is an algorithm for the numerical solution of particular systems of linear equations, namely those whose matrix is symmetric and positive definite, but it has broader use since quadratic can be used on the quadratic form as well. It can be also used to solve unconstrained optimization problems such as energy minimization. An example of the convergence rates of conjugate gradient and steepest descent is given in Figure 10 with the green line showing the steps of steepest descent method, while the red line shows the rapid conjugate gradient convergence. It is based on the notion of conjugate vectors, vectors that are orthogonal in respect to some matrix. This property allows an easy extraction of vector coefficients from the representation of the linear equation in the basis of conjugate vectors. If the conjugate vectors are chosen carefully, then usually not all of them are needed to obtain a good approximation to the solution. That is why it is useful to regard the conjugate gradient method as an iterative method. It also allows solving of large systems where the direct method would take too much time. This suggests taking the first basis vector to be the gradient of the function at the starting point, while the other vectors in the basis will be conjugate to the gradient. In other words, the conjugate gradient method can be considered as a steepest descent method with an extra constraint that all steps have to be conjugate, which leads to the faster convergence. One illustrative example where conjugate gradient method was applied on a real-life problem was minimizing the distance between motion capture data and the Dylan Thomas model to be driven by that data. This work was done as a part of my side project at Cardiff University Computer Science department and even though not directly related to this PhD work, it demonstrates well the importance of the optimization algorithms. In Figure 11 four images are shown captured from the Dylan Thomas video sequence. On the right is a set of points extracted from the positions of the markers on the actor's face (not shown). On the left is a Dylan Thomas expression generated to best match the marker's position. This is done by combining the set of static face expressions and performing the optimization to get the set of expression weights that will produce the minimum distance between the model and the actual marker positions.



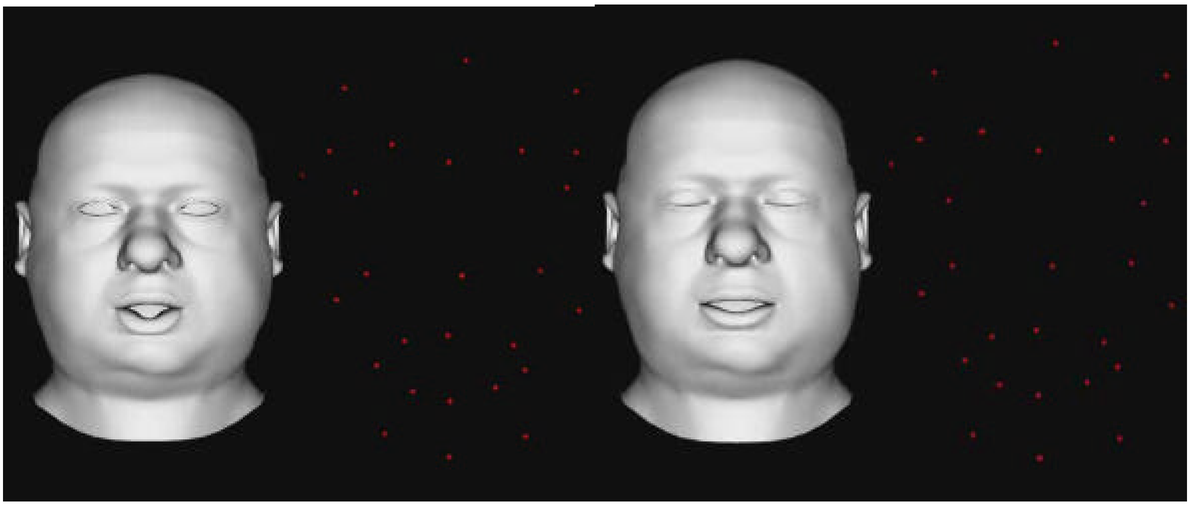


Figure 11 Conjugate gradient method applied to driving a Dylan Thomas model. Four different example expressions are produced by extracting the markers from an actor's face (right) an applying them to the model (left)

Nelder-Mead method or downhill simplex method uses the concept of a simplex, which is a polytope of N + 1 vertices in N dimensions; a line segment on a line, a triangle on a plane, a tetrahedron in three-dimensional space etc. The method approximately finds a local optimum to a problem with N variables when the objective function varies smoothly. As each run of the simulation is expensive, it is important to make a good decision when placing a starting point. The Nelder-Mead algorithm generates a new test position by extrapolating the behaviour of the objective function measured at each test point arranged as a simplex. At each iteration, the algorithm chooses to replace one of these test points with the new test point. The simplest option is to replace the worst point with a point reflected through the centroid of the remaining N points. If this point is better than the best current point, then it can be stretched out exponentially along that line. On the other hand, if that new point is not much better than the previous value, then the algorithm is stepping across a valley, so the simplex shrinks towards the best point.

The Simplex algorithm is a popular algorithm for numerically solving linear programming problems. Linear problem can be represented in geometric terms as a closed convex polytope, defined by intersecting a number of half-spaces in n-dimensional Euclidean space; each half-space is the area which lies on one side of a hyperplane. In this case it can be shown that the optimum value is attained on the boundary of the polytope. Methods for finding this optimum point on the polytope work in several ways: some attempt to improve a possible point by moving through the interior of the polytope (interior point methods), while others start and remain on the boundary when searching for an optimum. Simplex algorithm follows the latter methodology. The idea is to move along the facets of the polytope, point to point, in search for the optimum. However, unless the optimum occurs on an edge or face that is parallel to one of the hyperplanes, the optimum will be unique and occur at a vertex of the polytope. If the optimum is found on an edge or a face that is parallel to one of the hyperplanes, then the optimum is not unique and can be obtained at any point on that edge or face. Since the simplex algorithm is concerned only with finding a single optimal point, it is possible to look solely at moves skirting the edge of a simplex, ignoring its interior. Starting at some vertex of the polytope, at every iteration an adjacent vertex is chosen so that the value of the objective function does not decrease. If no such vertex exists, a solution to the problem has been found. However, often such an adjacent vertex is nonunique, and a pivot rule must be specified to determine which vertex to pick next. Various pivot rules exist to solve this problem but there is no need to go into more detail.

Interior point methods (or barrier methods) are a certain class of algorithms to solve linear and nonlinear convex optimization problems. These algorithms have been inspired by Karmarkar's algorithm, developed by Narendra Karmarkar for linear programming (Karmarkar, 1984). The basic principle of the method consists of using a self-concordant barrier function used to encode the convex set. Contrary to the simplex method described earlier, it reaches an optimal solution by traversing the interior of the feasible region. As mentioned earlier, any convex optimization problem can be transformed into minimizing (or maximizing) a linear function over a convex set.

The primal-dual schema is a discrete optimization technique that is well known in the combinatorial optimization field. It started as a very general technique for solving linear programming problems but it was realized that it could be a very powerful tool for deriving approximation algorithms to problems of linear integer programming as well. An interesting application of this method is Markov random field optimization problem that works on a discrete set of objects, all of which are vertices in a graph while the edges of this graph encode the objects' relationships. One label from the set of labels has to be assigned to each object and a price has to be paid when assigning a label to the object and to the pair of objects. The goal is to minimize the sum of those prices. The Fast-PD is an efficient approach for solving the primal-dual problem in Markov random field optimization by ensuring that the number of augmentations decreases significantly over time (Komodakis, Tziritas and Paragios, 2008).

1.4.3 Stochastic Optimization

So far only standard, deterministic and relatively simple optimization algorithms have been presented. In the following stochastic and global optimization algorithms will be reviewed, since it is often important to escape local minima when dealing with difficult objective functions. These algorithms are often imaginative and draw the fundamental principles of operation from the natural phenomena. An unfortunate drawback of many of these algorithms is that they tend to be relatively slow.

The Simultaneous Perturbation Stochastic Approximation (SPSA) algorithm is capable of an effective loss-function minimization when only, often noisy measurements, of the loss function are available. No measurements of the gradient of the loss function are assumed available. SPSA successfully addresses one of the major shortcomings of these finite-difference stochastic approximation algorithms that take derivative steps in each direction by significantly reducing the number of measurements required in many multivariate problems of practical interest. In theory, the execution speed up in respect to the standard stochastic approximation methods is equal to the number of the dimensions. SPSA has also been extended to the second-order approximation based on estimating both the loss function gradient and inverse Hessian matrix at each iteration. The aim of this approach is to emulate the acceleration properties associated with deterministic algorithms of Newton form, particularly in the terminal phase where the first-order SPSA algorithm slows down in its convergence. The second-order SPSA algorithm requires only five loss function measurements at each iteration, independent of the problem dimension (Spall, 1997).

Wang (Wang and Spall, 1998) presented a stochastic approximation algorithm based on the penalty function method and a simultaneous perturbation gradient estimate for solving stochastic optimization problems with general inequality constraints based on using a penalty function. A very general convergence result for the proposed algorithm has been presented, however some of the requirements on the problem are difficult to satisfy in practice.

Simulated Annealing (SA) is a global optimization method, designed for locating a good approximation to the global optimum of a given function in a large search space. It is often used when the search space is discrete. The name and inspiration come from the annealing technique in metallurgy, a technique involving heating and controlled cooling of a material to increase the size of its crystals and reduce their defects. The heat causes the atoms to become unstuck from their initial positions and wander randomly through states of higher energy, while the slow cooling provides more chance for finding configurations with lower internal energy than the initial one. By analogy with this physical process, each step of the SA algorithm replaces the current solution by a random near solution, chosen with a probability that depends on the difference between the corresponding function values and a global temperature parameter T that is gradually decreased during the process. The dependency is such that the current solution changes almost randomly when T is large, but behaves increasingly "downhill" as T approaches zero. The allowance for "uphill" moves saves the method from becoming stuck at local minima.

Quantum Annealing (QA) is a general method for finding the global minimum over a given set of candidate solutions by a process analogous to quantum fluctuations. It is used mainly for problems where the search space is discrete and with many local minima. In quantum annealing, a current state is randomly replaced by a randomly selected neighbour state if the latter has a lower energy (value of the objective function). The process is controlled by the tunnelling field strength, a parameter that determines the extent of the neighbourhood of states explored by the method. The tunnelling field starts high, so that the neighbourhood extends over the whole search space and is slowly reduced through the computation, until the neighbourhood shrinks to those few states that differ minimally from the current state.

Parallel Tempering, also known as replica exchange, is a simulation method aimed at improving the dynamic properties of Monte Carlo method simulations and Markov chain Monte Carlo sampling methods. For example energy updating simulation based on the temperature that runs as a single Monte Carlo simulation can suffer from critical slowing down when the system is highly correlated. This can be improved by running for example two parallel simulations with slightly different temperatures so that the generated distributions overlap. Another way to interpret this overlap is to say that system configurations sampled at one temperature are likely to appear during a simulation at another temperature. At a given Monte Carlo step, the global system can be updated by swapping the configuration of the two systems, or alternatively swapping the two temperatures. If temperatures and number of systems are carefully chosen it is possible to achieve an improvement in the mixing properties of a set of Monte Carlo simulations that exceeds the extra computational cost of running parallel simulations. Parallel tempering method can also be used as a super simulated annealing that does not need restart, since a system at high temperature can provide new local optimizers to a system at low temperature, allowing tunnelling between metastable states and improving convergence to a global optimum.

Tabu search enhances the performance of a search method by using memory structures. Once a potential solution has been determined, it is marked as tabu so that the algorithm would not visit that possibility repeatedly.

A Genetic Algorithm (GA) is a global, heuristic search technique used in computing to find exact or approximate solutions to optimization and search problems. Genetic algorithms are a particular class of evolutionary algorithms that use techniques inspired by evolutionary biology such as inheritance, mutation, selection, and crossover.

The Ant Colony Optimization (ACO) algorithm is a probabilistic technique for solving computational problems that can be reduced to finding good paths through graphs. They are inspired by the behaviour of ants when finding paths from the colony to food. In the real world, ants initially wander randomly and upon finding food return to their colony while laying down pheromone trails. If other ants find such a path, they are likely not to keep travelling at random, but to instead follow the trail, returning and reinforcing it if they eventually find food. ACO is basically a computer simulation of that behaviour.

Particle swarm optimization is a stochastic, population-based, evolutionary optimization algorithm based on the principle of social optimization. The communication structure or social network is defined, assigning neighbours to each individual to interact with. A population of individuals is defined as random guesses when the problem solution is initialized. These individuals are also known as the particles, hence the name particle swarm. The particles iteratively evaluate the fitness of the candidate solutions and remember the location where they had their best success. They also make this information available to their neighbours. Movements through the search space are guided by these successes, with the population usually converging to a problem solution better than that of a non-swarm approach using the same methods.

Differential Evolution (DE) is an optimization technique similar to GA but adapted to continuous problems. Floating-point instead of bit-string encoding and arithmetic vector operations instead of logical ones are used. These recasts have changed Genetic Annealing from a combinatorial into a continuous optimizer. The crucial idea behind DE is a scheme for generating trial parameter vectors by adding the weighted difference between two population vectors to a third vector.

Harmony search is a music phenomenon-inspired algorithm that was further developed by considering ensemble among music players. It conceptualizes a group of musicians together trying to find better state of harmony, where each player produces a sound based on one of the three operations: random selection, memory consideration, and pitch adjustment.

Memetic algorithm, similar to genetic algorithms, represents one of the recent growing areas of research in evolutionary computation. It is widely used as a synergy of evolutionary or any population-based approach with separate individual learning or local improvement procedures for problem search.

Bees algorithm is a population-based search algorithm that mimics the food foraging behaviour of swarms of honey bees. In its basic version, the algorithm performs the neighbourhood search combined with the random search and it can be used for both combinatorial and functional optimisation.

1.5 Statistical Model Building for Registration and Segmentation

Using a statistical model based on the training data is a potentially effective tool for both segmentation and registration (Cootes and C.J.Taylor, 2004). Its main advantage is that knowledge of the problem can be used to resolve the confusion caused by structural complexity, provide tolerance to noisy or missing data, and provide a means of labelling the recovered structures. The idea is to apply the knowledge of expected shapes of structures, their spatial relationships, and their texture appearance to restrict the automated system to plausible interpretations. To be useful, a model needs to be specific, capable of representing only legal examples of the modelled object. Statistical models can reproduce specific patterns of variability in shape and texture by analysing the variations in shape over the training set. It is difficult to achieve this whilst allowing for natural variability and thus it is essential what features will be supplied from the training set to the model building phase. The key step of the statistical model training phase is the dimensionality reduction on the large set of features of the training data set. The reason for dimensionality reduction is to reduce the computational cost of the optimization method that is actually used to apply the model to the real data. The idea behind this concept is to find statistical dependencies between the produced features and reduce the dimensionality of the space by selecting only a certain number of the most prominent directions in the data set, the most important eigenvectors. It is important to explain in more detail techniques that are used to perform this step.

1.5.1 Dimensionality Reduction

Principal component analysis (PCA) is the standard vector space transform technique used to reduce multidimensional data sets to lower dimensions for analysis. It works by calculating the eigenvalue decomposition of a data covariance matrix or singular value decomposition of a data matrix. Usually, a relatively small number of eigenvectors with greatest eigenvalues can describe the original data well. A technique similar to PCA is Local Discriminant Basis (LDB) algorithm, a powerful algorithmic framework that was developed as a technique for analyzing object classification problems. The main advantage of LDB over other similar techniques such as PCA is its lower computational cost (Hazaveh and Raahemifar, 2003). An adapted waveform framework called best basis algorithm had been developed prior to the development of LDB, mainly for signal compression problems. However, PCA and LDB work well on linear data but do not model well the non-linearities or discontinuities in the data. There are also non-linear dimensionality reduction methods, such as the constraint shape space point distribution model that automatically clusters the data into a combination of Gaussian distributions.

1.5.2 Model Types

Two main classes of statistical models will be briefly discussed. An Active Shape Model (ASM) deforms a shape model to describe the location of structures in a target image. It works by iteratively finding a new best match for each point of the model, updating the parameters to best fit the model to the new found points and repeating this cycle until convergence. These models can be combined with more complex texture feature models into efficient discriminative models (Cristinacce and Cootes, 2007). An Active Appearance Model (AAM) manipulates a model capable of synthesising new images of the object of interest by finding the model parameters which generate a synthetic image as close as possible to the target image (Cootes and C.J.Taylor, 2004).

Statistical shape models usually work by analysing a set of landmark points. However, it is often desirable to have smooth and invertible deformation fields, or diffeomorphisms. Unfortunately it is difficult to interpolate the deformation field between the points and ensure that the resulting mapping is diffeomorphic. One effective way of representing diffeomorphic deformation fields is by composing simple diffeomorphisms and using those to build a model (Cootes et al., 2008).

Statistical models have not been applied to the OCT segmentation problem yet, so that would be a novel approach to the problem. Additionally, constructing an OCT model is a novel, specific task and thus might yield not only specific but also general contribution to the statistical model building that could be related to some other problem.

1.6 Image Registration

Image registration refers to aligning source image to the target image in order to get the best match. Registration is necessary in order to be able to compare or integrate the data obtained from different measurements. Registration can differ based on the transformations used that can be either affine: combination of translation, rotation, scaling and shear. Affine transformations are global in nature, thus not being able to model local deformations which often need to be corrected in medical imaging. The second category includes elastic transformations that allow local warping of the image features, thus providing support for local deformations.

1.6.1 Classification of Registration Algorithms

Registration algorithms can be divided into two main groups: area based methods and feature based methods. For area based image registration methods, the algorithm looks at the structure of the image as a whole by using for example correlation metrics, Fourier properties and other means of structural analysis. Feature based methods, instead of looking at the overall structure of images, fine tune their mappings to the correlation of image features: lines, curves, points, line intersections, boundaries and others.

Another useful classification is between intermodal and multimodal registration algorithms. Intermodal registration algorithms are intended to register images acquired using the same kind of imaging device, while multimodal registration algorithms are capable of registering images acquired using different imaging devices. There are several examples of multimodal registration in the medical imaging field. Registration of brain CT/MRI images is the most notable one. In the OCT field registering OCT volume to a scanning laser ophthalmoscope (SLO) fundus image is an interesting approach.

Registration can be mathematically expressed as using one of the optimization algorithms to minimize the distance function between the source and the target. Several distance metrics are used in practice. The most common ones are cross-correlation, mean-square difference and mutual information. Mutual information is a measure of mutual dependency of two variables and thus suited for multimodal registration while the other two metrics are usually used for intermodal registration.

Most of the image registration methods operate in the spatial domain, using features, structures, and textures as matching criteria. This is also the most intuitive approach since in the spatial domain images look as the human eye might perceive them. However, there is a class of algorithms that use the properties of the frequency-domain to directly determine shifts between two images. By applying the Phase correlation method to a pair of overlapping images, it is possible to use the produced image which contains a single peak as the relative translation between the two images. Unlike many spatial-domain algorithms, the phase correlation method is resilient to noise, occlusions, and other defects typical of medical images. Additionally, the phase correlation uses the Fast Fourier transform to compute the cross-correlation between the two images, resulting in large performance gains. This method can be also extended to determine affine rotation and scaling between two images by first converting the images to log-polar coordinates. Due to properties of the Fourier transform, the rotation and scaling parameters can be determined invariant to the translation. This single feature makes phase-correlation methods highly attractive when compared to typical spatial methods, which must determine rotation, scaling, and translation simultaneously, often at the cost of reduced precision in all three.

1.6.2 Registration Metric

In the registration process, the metric typically compares intensity values in the fixed image against the corresponding values in the transformed moving image. When a point is mapped from one space to another by a transform, it will in general be mapped to a non-grid position. Therefore, interpolation is required to evaluate the image intensity at the mapped position (Ibanez et al., 2005). The simplest interpolation technique is nearest neighbour interpolation which uses the intensity of the nearest grid position. Linear interpolation assumes that intensity varies linearly between grid positions and unlike nearest neighbour interpolation, the interpolated intensity is spatially continuous. However, the intensity gradient is still discontinuous at grid positions. B-spline interpolation uses spline coefficients that are computed using recursive filtering. Sinc window interpolator corrects the well known signal processing phenomena which states that the process of sampling a spatial function using a periodic discrete grid results in a replication of the spectrum of that signal in the frequency domain. It can be shown that recovering the continuous signal from the discrete sampling is equivalent to the removal of the replicated spectra in the frequency domain. This can be done by multiplying the spectra with a box function that will set to zero all the frequencies above the highest frequency in the original signal which is equivalent to convolving the spatial discrete signal with a sinc function (sin(x)/x). The sinc function has infinite support, which of course in practice cannot really be implemented so the sinc is usually truncated by multiplying it with a window function.

1.6.3 Pyramidal Approach

Performing image registration using a multi-resolution pyramidal approach is widely used to improve speed, accuracy and robustness. The basic idea is to perform the registration at a coarse scale where the images have fewer pixels. The spatial mapping determined at the coarse level is then used to initialize registration at the next finer scale and the process is repeated until it reaches the finest possible scale. This coarse-to-fine strategy in general greatly improves the registration success rate and also increases robustness by eliminating local optima at coarser scales (Ibanez et al., 2005). Some methods use arbitrary coefficients for multi-resolution pyramid levels while using scale-space instead can be advantageous because of its image feature preserving properties.

1.6.4 Model-based Approach

Another approach that can be taken is registering a geometrical model to an image. Using this approach, a geometrical model is built first and a number of parameters are identified in the model. This same approach can be considered as a segmentation technique, since once the model has been optimally superimposed on the image, pixels can be labelled according to their associations with specific parts of the model. One of the main advantages of this approach is that, as opposed to image-to image registration, it actually provides insight into the anatomical structure contained in the image (Ibanez et al., 2005). This shows well how interconnected registration and segmentation problems are.

1.6.5 Registration of OCT Data

Registration in respect to the OCT in general refers to the registration of the stack that makes up the imaged volume. This is a more difficult task than a simple registration between the two images since it is important to perform this process in regard to the stack as a whole. Otherwise, if registered sequentially, serious errors can result as a consequence of the small errors that accumulate during the slice to slice registration. Another important issue is that the slices might actually overlap at some parts while some parts of the retina might end up being missed. The reason for this is the eye motion during the scan procedure that can occur in three degrees of freedom, including the rotation along the z axis. Thus, the ideal registration method would take all these facts into account and would correct for those transformations by applying reverse transformations and interpolating between the non-scanned areas. One approach would be to use a fundus image without the motion artefacts as a reference for the OCT stack registration. In the following, several methods will be reviewed that are specific to the OCT field or are relevant to this work in some other respect.

An automatic sub pixel registration algorithm that minimizes the mean square intensity difference between a reference and a test data set, which can be either images or volumes was presented in (Thevenaz and Unser, 1996). It uses an explicit spline representation of the images in conjunction with spline processing, and is based on a pyramid approach. The minimization is performed according to a fast, custom variation of the Marquardt–Levenberg algorithm for nonlinear least-square optimization. The geometric deformation model is a global three-dimensional affine transformation that can be optionally restricted to rigid-body motion, combined with isometric scaling. Good results were obtained for the registration of intramodality positron emission tomography and functional magnetic resonance imaging data. Based on this paper, a plug-in for ImageJ software was developed which was used as a basis for a modified version that is described later on.

Sorzano et al. presented an elastic registration algorithm for the alignment of biological images that represents deformation field as a B-spline model, which allows for a rich variety of deformations (Sorzano, Thevenaz and Unser, 2005). The registration problem is solved by minimizing a pixelwise mean-square distance measure between the target image and the warped source. The problem is further constrained by way of a vector-spline regularization that provides control over the deformation divergence and its curl. The algorithm is also capable of handling soft landmark constraints, which is useful when parts of the images contain very little information. An optimal analytical solution is proven in the case when only landmarks and smoothness considerations are taken into account. This paper was also a base for the ImageJ plug-in which was modified and used in the registration work.

Interesting registration and segmentation results were presented by applying primal-dual schema and MRF representation to a discrete grid laid over the images (Glocker et al., 2008). It guarantees finding a solution close to the global optimum in a reasonable time. However, the number of variables has to be relatively low. Although applicable to the MRI segmentation, this approach is not directly suitable for the OCT segmentation since the grid is too constrained to allow for a large variety of possible deviations from the normal retinal model.

An additional benefit for the stack registration would be use of an independent fundus image free from motion artefacts that would be used as a reference to register a fundus image created from the OCT stack. It was shown that an effective approach for generating a fundus image from the OCT signal in real time is to use the non-interference terms in the frequency spectra of the OCT signal to generate an intensity image (Jiao et al., 2005). The same author demonstrated use of these fundus images produced from the low density OCT scan as a reference for the registration of the high density OCT slices of the same subject (Jiao et al., 2006).

A method for mosaicing with correction of motion distortions and tissue deformations for in vivo fibered microscopy that is relevant to OCT stack registration has been presented (Vercauteren et al., 2006). It deals with a global registration problem that solves the error accumulation effect which occurs when registering a stack sequentially. However, some of the transformations used are specific to the fibered microscopy imaging so it cannot be used directly on the OCT stacks. The approach is based on a hierarchical framework that is able to recover a globally consistent alignment of the input frames, to compensate for the motion distortions and to capture the non-rigid deformations by modelling the relationship between the motion and the motion distortions. The proposed global alignment scheme is seen as an estimation problem on a Lie group, since the space of rigid body transformations is not a vector space but rather a Lie group that can be considered as a Riemannian manifold. A Lie Group is a smooth manifold together with a smooth composition map and a smooth inverse map, which satisfy the group axioms: closure, associativity, existence of a neutral element and existence of an inverse. Riemann manifold is a real differentiable manifold M in which each tangent space is equipped with an inner product, a Riemannian metric, which varies smoothly from point to point.

1.7 Data Segmentation

Since the focus of this work is on segmentation it is essential to give a detailed segmentation algorithms overview of both general and specific methods applied to the medical and OCT field in particular.

Segmentation refers to the process of partitioning a digital image into multiple regions. The general goal of segmentation is to simplify the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects or boundaries in images. The result of image segmentation is a set of regions that cover the entire image, or a set of contours extracted from the image. Pixels that belong to the same region are bounded by strong edges or have similar features or both. Some of the practical applications of image segmentation are in medical imaging, locating objects in aerial or satellite images, face recognition, fingerprint recognition and machine vision. In medical imaging segmentation can be used to locate tumours and other pathologies, measure tissue volumes, assist in computer-guided surgery, diagnosis, treatment planning, study of anatomical structure, etc.

In OCT in particular, existing published approaches and results vary in respect to the number of the layers segmented and on the tolerance to pathological cases. In general they tend to be very sensitive to pathologies and able to segment only the most prominent cases. The goal of this work is to develop an algorithm capable of segmenting all the visible layers and to be able to work on pathologies.

1.7.1 Classification of Segmentation Methods

Many general-purpose algorithms have been developed for image segmentation. Since there is no general solution to the image segmentation problem, these techniques often have to be combined with domain knowledge in order to effectively solve an image segmentation problem.

Histogram-based methods are simple and computationally very efficient when compared to other image segmentation methods because they typically require only one pass through all the pixels. An image histogram is computed and the peaks and valleys in the histogram are used to locate the clusters in the image. Usually colour or intensity is used as the measure. It is possible to extend this technique by recursively applying the histogram-seeking method to clusters in the image in order to divide them into smaller clusters. Main disadvantages of the histogram-seeking method are that it may be difficult to identify significant peaks and valleys in the image and that it does not take into account structural information.

Gaussian Mixture Model is a density model that is comprised of a number of Gaussian component functions that are combined to provide a multimodal density. It can be employed to segment an image based on the image statistics. Each mixture component is a Gaussian function with defined mean and covariance matrix.

The edge detection methods are based on gradient calculation. The assumption is that pixels with a low gradient constitute the background regions while pixels with high gradient values make edges or boundaries. Since the definition of low and high gradient value is obviously relative, a threshold value needs to be defined. The gradient is calculated using discrete differencing, in other words convolving the image with a differencing block that approximates continuous gradient at each point in the image (Hazaveh, 2003). However, the edges identified by edge detection are usually disconnected. To segment an object from an image it is necessary to extract closed region boundaries. This problem is usually solved by bridging the discontinuities if the distance between the two edges is within some predetermined threshold.

The region growing methods rely on expanding the initial regions. They can be divided into two basic subclasses: seeded and unseeded methods. Seeded method takes a set of seeds as input along with the image with each seed marking one of the objects to be segmented. The regions are iteratively grown by comparing neighbouring pixels to the attached region. The difference between a pixel's intensity value and the region's mean is usually used as a measure of similarity. This process continues until all pixels are allocated to one of the regions. The segmentation results are heavily dependent on the choice of seeds and in cases that seeds are placed in noisy regions, it can prevent the algorithm from advancing properly. Unseeded region growing is a modified algorithm that does not require explicit seeds and starts with a single region. It considers the neighbouring pixels in the same way as seeded region growing. If the neighbouring pixel differs from the region’s mean by less than a predefined threshold, it is added to the region. If not, the pixel is considered significantly different from all current regions and a new region is created containing that pixel.

The watershed transformation considers the gradient magnitude of an image as a topographic surface that is used for a water potential simulation. Pixels having the highest gradient magnitude intensities correspond to hills, which represent the region boundaries. Water placed on any pixel enclosed by common high ground flows downhill to a common local intensity minimum. Pixels draining to a common minimum form a catchment basin which represents one of the regions. Even though intuitively attractive, this method is very sensitive to noise present in the image.

Active contour (snake, curve propagation) is a class of popular techniques for object segmentation and tracking. The central idea behind this approach is to evolve a curve towards the lowest potential of a cost function, where its definition reflects the task to be addressed and imposes certain smoothness constraints. The contour is parameterized according to some sampling strategy and each element is evolved according to the image potential and internal smoothness terms. While such traditional techniques can be very efficient, they have a serious drawback that limits the evolution in special cases. One example is the case of multiple curves merging or splitting that is notoriously difficult to properly describe. The level set method can be used to efficiently address this problem. The central idea is to represent the evolving contour in a higher dimension, where its zero-level corresponds to the actual contour. According to the motion equation of the contour, it is possible to derive a similar flow for the implicit surface that when applied to the zero-level will reflect the propagation of the contour. The basic principle of operation is shown in Figure 12. However, if initialized far from the optimum, level set method can be extremely slow. It is also sensitive to noise and can get stuck at fake edges caused by noise or can evolve through the holes caused by noise present at the real edges.

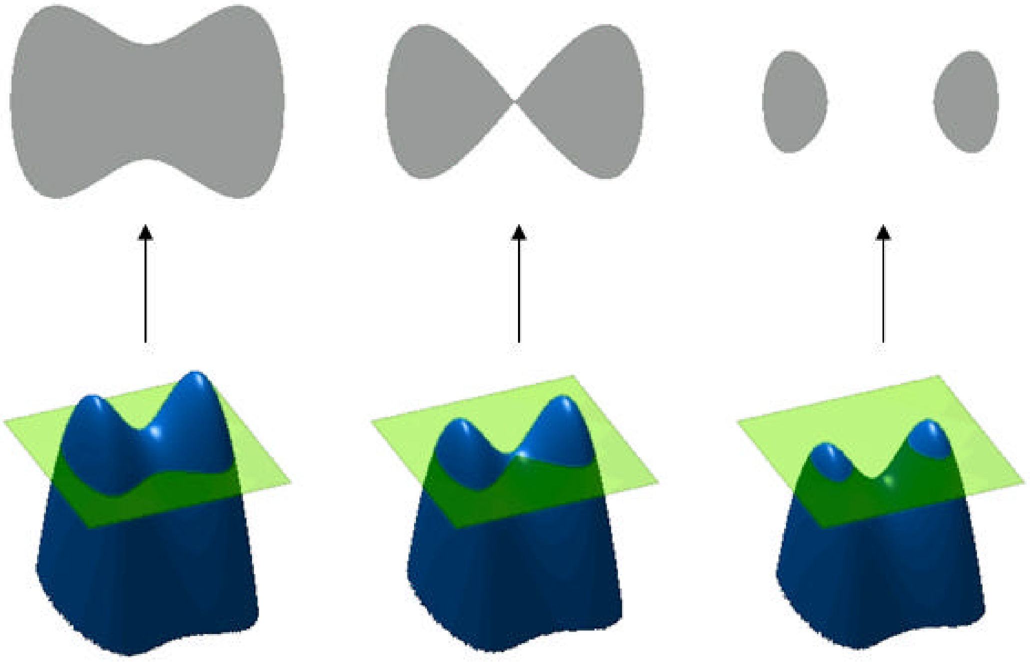


Figure 12 The level set method keeps the internal state of an evolving surface embedded in 3D (below). The plane cross-section produces the required segmentation (above). This type of representation allows for natural splitting of segmented areas, as the actual surface does not actually get split

Graph partitioning methods segment an image by modelling it as a weighted undirected graph. Each pixel is represented as a node in the graph and an edge is formed between every pair of pixels. The weight of an edge is chosen to represent the measure of the similarity between the pixels. The image is partitioned into disjoint sets by removing the edges connecting the segments. The optimal partitioning of the graph (the cut) is the one that minimizes the weights of the edges that were removed.

1.7.1.1 Model Based

Model based approach uses the variation obtained from the training set and imposes those constraints when segmenting an unseen image. This guarantees that the segmentation will be plausible and less sensitive to noise. However, it is extremely important to have a large, representative training set that includes all possible variation. It was described in the separate section that explains how statistical models can be built.

An extension to the standard AAM, called Robust AAM (RAAM) has been proposed in (Beichel et al., 2005) in order to enable more robust approach in presence of outliers. A two stage extension has been introduced; in the first step residuals are analyzed, while in the second stage optimization is used to detect gross outliers. This approach works well for cases when parts of the image are occluded or corrupted by strong noise, while the texture in the remaining part of the image is close to a prediction of the standard AAM model. Unfortunately, this approach is not directly applicable to the problem of OCT data segmentation, as the strong speckle variation is present in the whole image.

Interesting model based approaches have been applied in the face recognition and scene segmentation and classification fields. Cao and Fei-Fei presented a spatially coherent model that utilizes a visual bag of words representation of images but introduces spatial coherency and hierarchy of objects in an image (Cao and Fei-Fei, 2007). This approach could be useful if patches need to be segmented in OCT data that are not necessarily connected, but still obey some spatial consistency, which is often the case with layer detachments and fluid patches in pathologies. Another paper from the same research group performs scene classification, segmentation and annotation of objects utilizing a generative model (Li, Socher and Fei, 2009). This could provide very useful diagnostic benefit of not only segmenting various retinal layers but also providing a general prognosis for the patient. Kohli and Torr have presented a graph cuts approach in Markov random fields (Kohli and Torr, 2007). They applied it to a sequence of image frames in a video, utilizing the graph constructed from a previous frame to segment faster the current frame, without the need to construct a graph from scratch for each new image. This could potentially be speed wise beneficial to segmenting OCT stacks, as it can be noted that each slice in a stack is similar to its neighbours, just as videos frames are. These methods have been applied to problems where objects are usually well distinguishable, unlike the OCT data. Still, some of the ideas could be applied, modified, combined and used for OCT segmentation.

1.7.1.2 Extensions

Multi-scale approach is equally useful when applied to a segmentation problem as it is for registration. Segmentations are iteratively computed at multiple scales in scale-space and usually propagated from coarse to fine scales. This ensures that the process is both faster and less susceptible to noise.

Semi-automatic segmentation is often applied to difficult problems to enable the user to correct mistakes that might have been produced by the automatic algorithm or to initialize the algorithm and thus make it both faster and more accurate.

1.7.2 Segmentation of OCT Data

In the following, a series of publications will be reviewed that apply segmentation algorithms to the OCT and medical field in general or are relevant in some other respect.

A simple, yet rather effective, pre-processing algorithm for shadow compensation has been presented in (Fabritius et al., 2009a). Shadows are sometimes produced by the highly reflective upper layers of the retina, causing poor visibility of the lower layers. While a shadow compensation step is purely numerical and thus cannot remove the problem completely, it can alleviate some of the difficulties posed for the subsequent segmentation task.

Zawadzki et al. used a semi-automatic algorithm for the OCT segmentation where user could paint the areas of interest in any slice of the volume. For segmentation SVM was used with a feature vector that contained intensity, location, mean of the neighbourhood, standard deviation and gradient (Zawadzki et al., 2007).

A 3D graph search approach to OCT retinal layer segmentation was presented in (Garvin et al., 2009). The algorithm first aligned all the slices and straightened the RPE layer. After that, an optimal graph cut was performed with weights describing both edge and regional information. Good results were obtained, however they were obtained on slight pathologies and good signal quality data only. It would be too computationally demanding to apply this approach to more varied cases.

Fernandez et al. presented segmentation results of the peak finding algorithm (Cabrera Fernández et al., 2005). Since it is an iterative thresholding algorithm, it is sensitive to noise and deviation from the normal retinal data. It would be also difficult to extend it to any non-typical case since the training phase does not exist and the algorithm procedure is hard coded instead. Even though some good results were obtained, it is prone to failure and it is possible that detected boundaries will overlap.

Mujat et al. used the deformable spline algorithm (active contour) to determine NFL thickness (Mujat et al., 2005). They also detected blood vessels by detecting intensity holes in the RPE layer.

Vermeer et al. (Vermeer et al., 2011) used texture type classification, labelling each pixel independently, followed by a level set contour refinement stage. This approach suffers from an inefficient execution time and sensitivity to noise, the result of the level set method being unable to predict the probable shape based on training data.

Mayer et al. (Mayer et al., 2010) used a standard peak based detection approach for normal and glaucoma retina, however, for more demanding boundaries they used an energy minimisation term, compensating for the discontinuities and low contrast.

For rodent retinas Molnár et al. (Molnár et al., 2011) have presented an efficient ILM and RPE sliding window detection algorithm that assumes relatively flat boundaries, which in rodents is generally true. They then refine the results using an active contour approach.

Another rodent segmentation approach, in this case retinal degeneration, by Sarunic et al. (Sarunic et al., 2010) used a semi-automatic approach; user supplies a few initial boundary estimate points after which an active contour algorithm finds the boundary.

Yazdanpanah et al. (Yazdanpanah et al., 2009) used an active contour approach, minimizing a region term and assuming a consistent image intensity within one layer, with added terms for shape prior and regularization term to enforce smoothness. The initial position of the boundaries is user initialized, followed by the least squares optimization. Apart from this approach being semi automatic rather than fully user independent, it also assumes that layers are of uniform brightness, which is often not the case.

Rossant et al. (Rossant et al., 2009) combined several image processing and computer vision methods into a rather complicated, but slow algorithm. They first find the brightest boundaries by using a recursive low-pass filter and finding the maxima. For finding the weak boundaries, at roughly the same offset from the already determined boundaries, they use Kalman filtering, utilizing the principle of prediction/verification. Finally, remaining layers are found by performing texture classification using k-means clustering, after which the final boundaries are determined by optimizing a consistent area function using simulated annealing (SA).

A Markov boundary model was applied to connect the extracted boundary edge primitives in (Koozekanani, Boyer and Roberts, 2001). Even though more robust than standard column wise thresholding methods, it still relies on connecting 1D points. This makes it sensitive to noise and thus the detected layer boundary can easily drift off from the real boundary. Special rules have to be applied to correct for such cases which makes the whole approach less general.

In (Chiu et al., 2010) a similar approach has been used; first the layer boundary endpoints are detected, followed by a 1D tracking step. In this case optimal path is found using standard Dijkstra's algorithm. However, this approach suffers from exactly the same drawbacks as the previous one.

Yang et al. (Yang et al., 2010) used dynamic programming shortest path approach, based on working in dual scale edge; one image was created using axial gradient, while fine scale image was produced using the Canny edge detector. Dual scale approach alleviates to some extent the problems with shadows and other discontinuities present in OCT images.

A useful multiscale approach to segmentation is using wavelets since the multiscale image edges will correspond to the local maxima of the wavelet transform (Deng et al., 1993).

Highly regular wavelets that may be of general interest have been introduced for the detection of clustered micro calcifications in mammograms (Lemaur, Drouiche and DeConinck, 2003). Since they have higher regularity than classical wavelets, their performance is better for real signals that cannot be represented by lower order polynomials.

A novel approach to epidermal segmentation using shapelets has been demonstrated (Weissman, Hancewicz and Kaplan, 2004). Shapelet decomposition can be thought of as a two-dimensional wavelet decomposition in which an image containing a particular shape is used as a basis kernel rather than as an orthogonal or bi-orthogonal 1D or 2D waveform. After the shapelet transform has been performed, standard column wise search looks for the first maximum to locate the boundary.

An elegant approach to retinal segmentation based on spectral rounding was introduced in (Tolliver et al., 2008). It is a graph-partitioning algorithm that is based on the eigenvector calculation to determine the oscillation steps that represent the retinal edges. It performed well considering that no a priory information was available to the algorithm, simply dividing an image iteratively along the oscillation boundary. However, the number of extracted layers was low and it is unlikely that it would be possible to extract layers with weaker signal without using some structural information.

Based on this literature review it can be concluded that even though many segmentation methods and approaches have been introduced, a general and reliable solution for the OCT layer segmentation has not yet been developed.

2 Segmentation

Retinal and choroidal segmentation are the key to early pathology detection and therapy monitoring. Several novel methods have been developed to deal with real world, clinical, large datasets, consisting of normals and pathologies, compromised by noise and imaged using different commercial and non commercial OCT systems.

The chapter is divided into: work on registration algorithm (2.1) as it related to both the retinal (2.2) and choroidal segmentation (2.3). At the current stage, the segmentation works on each image in a stack independently, thus registration is not really necessary. However, extending the segmentation to 3D would make it essential. Additionally, to interpret the result, create thickness maps etc. it is important to have a stack registered at least approximately. Thus, in chapter 2.1, the early work on registration is presented, dealing with combining existing algorithms (rigid body and elastic) into a package applicable to OCT data registration, applicable to a whole stack in automatic fashion.

Retinal segmentation section contains description of an adaptive thresholding approach used for top and bottom foveal boundary detection (as a starting point for segmentation of the other inner layers), early work on data segmentation based purely on texture analysis, adaptive template matching approach to segmentation and finally a statistical model that unifies both the texture and shape information based on machine learning principles.

For segmentation of choroid affected often by advanced pathologies, it was necessary to develop a more robust top and bottom boundary detection algorithm. How it works and why it is applicable in large problem area outside OCT is explained in subchapter 2.3.1. In subchapter 2.3.2, choroid segmentation is explained using a statistical model similar to the one developed for the foveal segmentation (2.2.4)

2.1 Registration

Image artefacts resulting from motion may degrade image quality and cause inaccurate clinical interpretation of images. Motion artefacts arise when the object being imaged is moved during data acquisition but is assumed stationary in the image reconstruction process. Axial motion refers to a situation where the longitudinal distance from the probe to sample is changed in time, resulting in SNR reduction (Yun et al., 2004). Transverse motion results in both SNR reduction and degradation in transverse resolution, visible as blurring of an image along the direction of motion (Yun et al., 2004). Motion artefacts can be reduced by increasing the imaging speed, by applying eye tracking based OCT systems or by applying registration.

2.1.1 Stack Registration Algorithm

Stack registration algorithm has been developed based on the papers already mentioned in the registration section (Thevenaz and Unser, 1996), (Sorzano et al., 2005). The basic rigid body algorithm is used to register a source to target image based on a pyramidal approach. The stack registration algorithm uses the rigid body registration method and calls it iteratively to perform the stack registration. The elastic registration algorithm warps a source to target image and is not used by the stack registration algorithm.

These methods were modified in order to extend the existing functionality. The goal was to obtain a unified solution that will allow for approximate rigid body registration and fine warping producing a smooth stack. The first thing that had to be addressed, was the limited registration range of the rigid body algorithm, which meant that if there was a larger jump in the original stack it could not be registered properly. This issue was solved by increasing the number of pyramid levels. The next step was to apply elastic registration to each iteration of the stack registration procedure. To accomplish this, the elastic registration was successfully added to the stack registration loop. However, a problem appeared when registering larger stacks and using the warping sequentially. Errors would accumulate in each iteration due to numerical inaccuracies which meant that subsequent slices would be substantially deformed. To solve this problem, masking was applied to exclude the background from biasing the objective function and warping parameters were constrained heuristically to prevent unrealistic deformation of slices. However, this temporary solution did not solve the problem completely. The solution to this problem is planned for future work and has to be accomplished using a comprehensive solution that works simultaneously on the stack as a whole. However, since registration is not the focus of this PhD, this temporary registration solution is for now deemed satisfactory.

2.1.2 Results

The registration results presented compare the smoothness of the original input stack and stacks after the registration using different parameters. Results were subjectively judged based on the smoothness and geometrical distortions. No objective criteria were used for the evaluation, since it is a work in progress. Also, it is difficult to perform an objective comparison for this task since the motion artefacts occur during the scan of living subjects and thus performing the eye histology is not feasible. Has the histology been available one could evaluate the results either automatically or using a manual approach by marking the corresponding structures in two sets (OCT and histology).

Figure 13 shows a resliced side view of a stack obtained by a 800 nm system that is given as input. It is an appropriate representation because the motion artefacts along the stack are easily perceived. Figure 13 a) shows the original, unregistered stack and motion artefacts are very prominent. In Figure 13 b), the stack is displayed after the stack rigid body registration was applied using the modified rigid body plug-in. Most of the motion artefacts are corrected but the registration is still not completely satisfactory. The remaining motion artefacts are well visible at the top boundary where smaller jumps and roughness still remain. Subtle motion artefacts remain along the hidden axis as well. Figure 13 c) shows the stack after a modified stack registration plug-in that uses both the rigid body and elastic registration was applied. The stack is smoother but accumulated errors resulted in the progressive deformation and thinning of the whole stack. Even though useful results were obtained, a real 3D registration algorithm that works on the whole volume concurrently would be preferable.

2.2 Retinal Segmentation

Retinal segmentation was done mostly with the goal of segmenting normal and close to normal retinas for purpose of early disease detection. Adaptive thresholding approach used for top and bottom foveal boundary detection (a starting point for segmentation of the other inner layers, replaced later by a more advanced approach described in 2.3.1), early work on data segmentation based purely on texture analysis, adaptive template matching approach to segmentation and finally the statistical model that unifies both the texture and shape information based on machine learning principles are presented in the subsequent sections.

2.2.1 Top and Bottom Retinal Boundary Detection

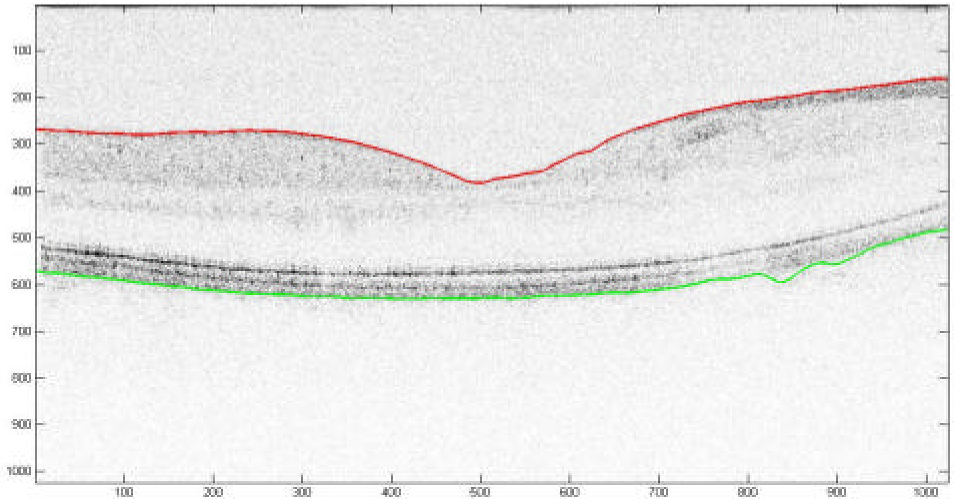
The next task to be solved was developing an algorithm that would detect top and bottom retinal boundaries since this information is necessary as a starting point for a subsequent multiple layer segmentation algorithm. This cannot be solved by simply thresholding column wise since the noise present in OCT data can cause either false edges or gaps in the real boundary. The first method tried was using a Level set approach, with force pushing the surface outwards in the normal direction and the inner force constraining the smoothness. This approach gave decent results but was not completely reliable (the curve would occasionally fall through in cases of noisy data) and could be extremely slow (in cases when there would be substantial amount of noise around the real boundary edges which would slow the curve down).

The adaptive thresholding algorithm starts by intensity thresholding, ILM boundary is found first by starting the thresholding process from the top of the image, while RPE boundary is found next by starting from the bottom. However, an image is divided into segments, allowing for different thresholds to be used in different parts of the image, as the image intensity is often not consistent throughout the image. Thresholding range and number of segments are chosen heuristically. Next, the algorithm converges to a close strong edge after the first estimate, additionally using constraints on distances between the boundaries. Robust polynomial fitting is afterwards used to eliminate outliers, followed by interpolation along the remaining points.

2.2.1.1Results

The boundary detection methods seek a top and a bottom retinal boundary that is a prerequisite for the segmentation algorithm since it constraints the segmentation space. Results obtained by both the level set and adaptive thresholding methods are presented. In general the distance between the drawn curve and the boundary perceived by the observer is used as a subjective measure. Since faster and more robust methods have been developed for the same purpose in the later stages of the PhD, no further work was put into further refining and evaluating this method. Images come from 800 nm and 1060 nm systems, with mostly healthy retinas analyzed, as well as one age related macular degeneration (AMD) and two patients with changes in RPE. One stack comes from a time domain system.

Figure 14 contains three snapshots taken during the top boundary extraction level set run on a macular hole image. It can be noted that results are generally good. However, it took several minutes to obtain this simple extraction, while on the other hand if parameters are changed to make the evolution more aggressive there is a serious risk of contour penetrating the real boundary. Figure 15 shows a sequence of healthy retinas with the top and bottom boundaries marked as extracted by the adaptive thresholding. Processing times per B-scan are about 2 seconds, while the extracted boundaries match real boundaries very well. In Figure 16, three pathological B-scans are shown with difficult, irregular macular holes at the bottom. For all three cases, boundary detection performed equally well as in healthy subjects. Figure 17 shows a difficult retina that contains significant shadowing on the left which makes the NFL boundary hardly visible. Some noise is present around the boundaries as well. The algorithm did not perform as well as in previous cases but the extracted boundaries still roughly match real boundaries. With accurate stack registration applied, the algorithm could be extended to 3D, making the top and bottom surface extraction even more reliable.



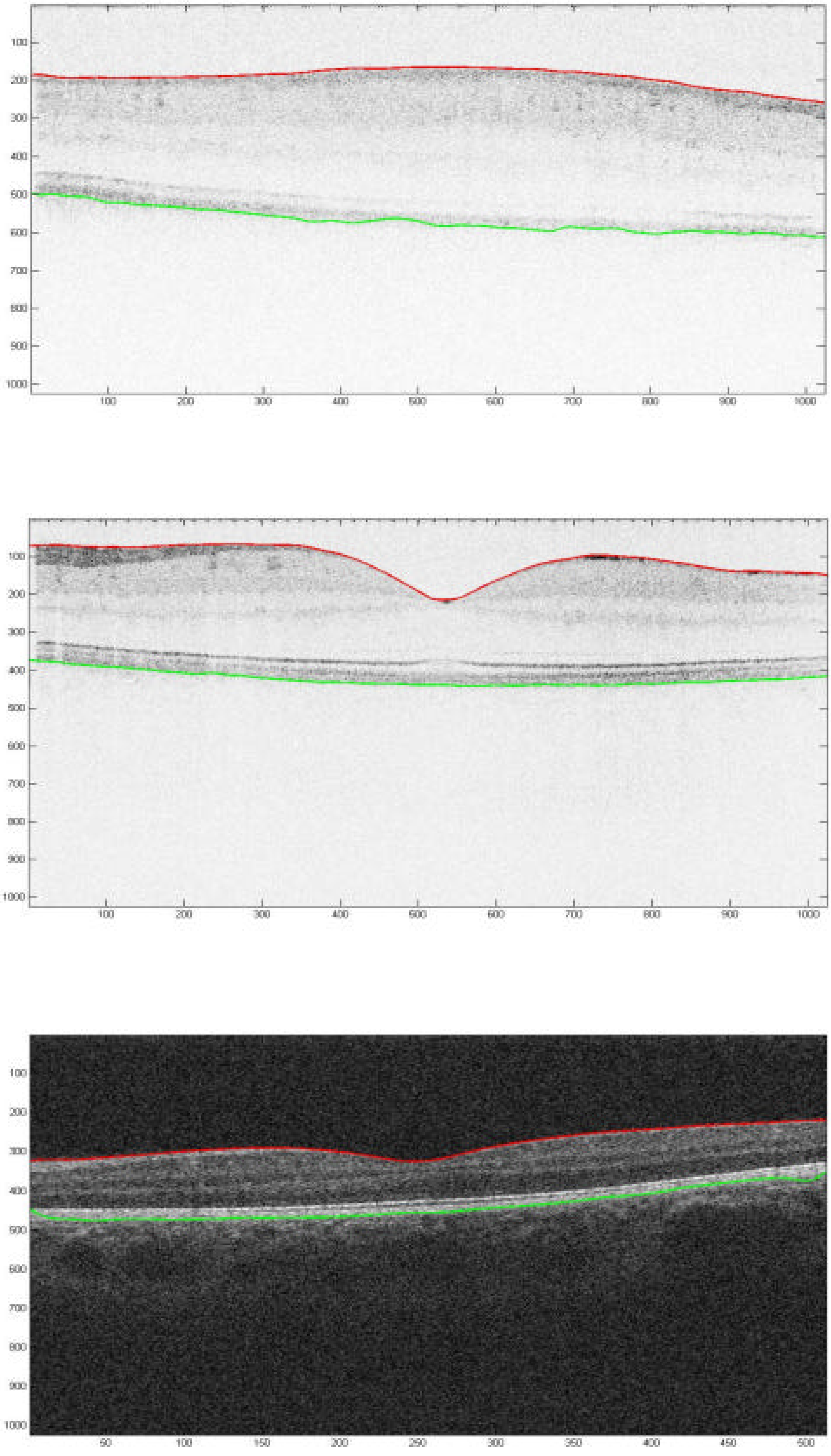


Figure 15 Top (red) to bottom (green) boundary detection using adaptive thresholding on healthy retinas imaged by two different OCT setups (high resolution prototype a) to c) and Zeiss commercial system d))

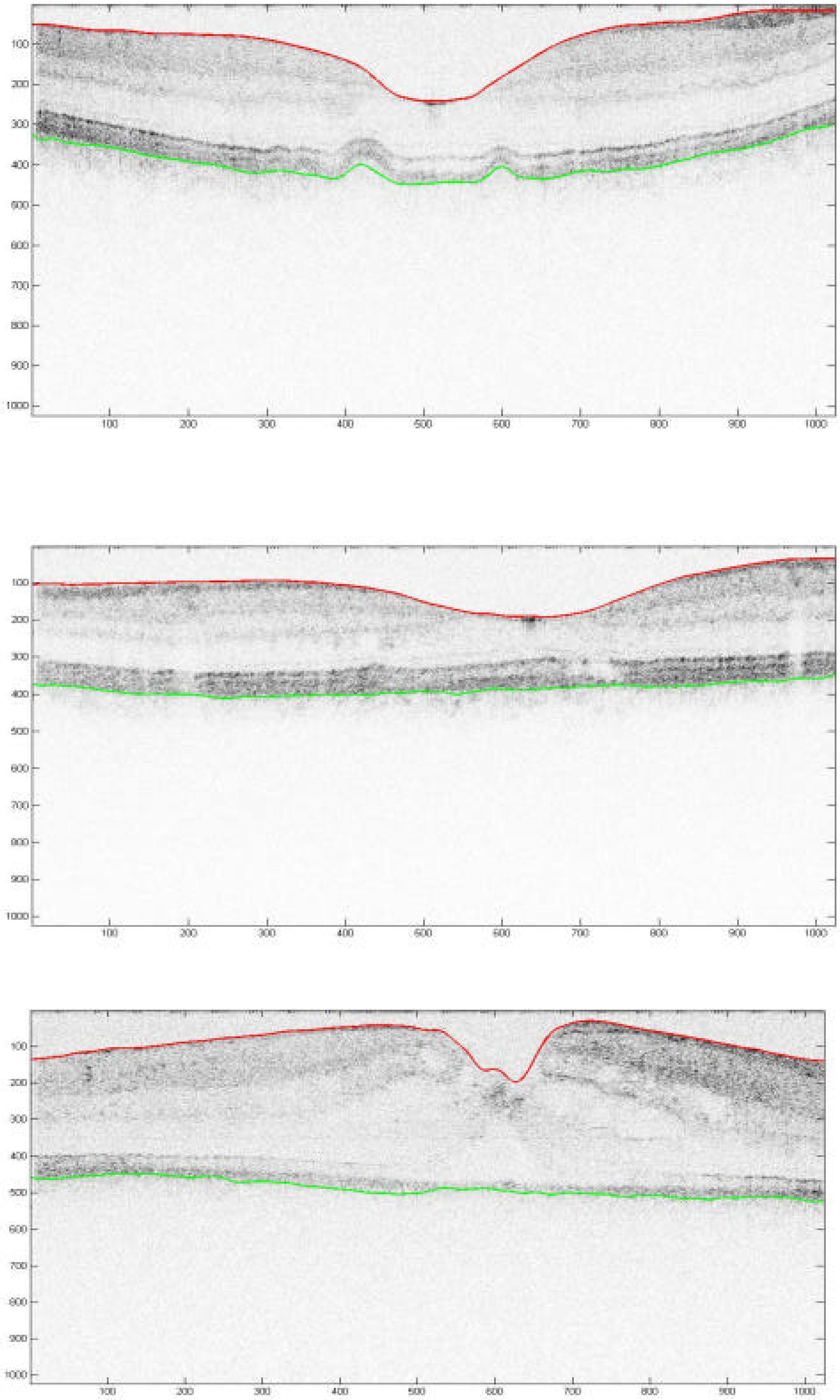


Figure 16 Top to bottom boundary detection on pathological retinas using adaptive thresholding (high resolution prototype OCT system). a) and b) show changes in the RPE, while c) is an AMD

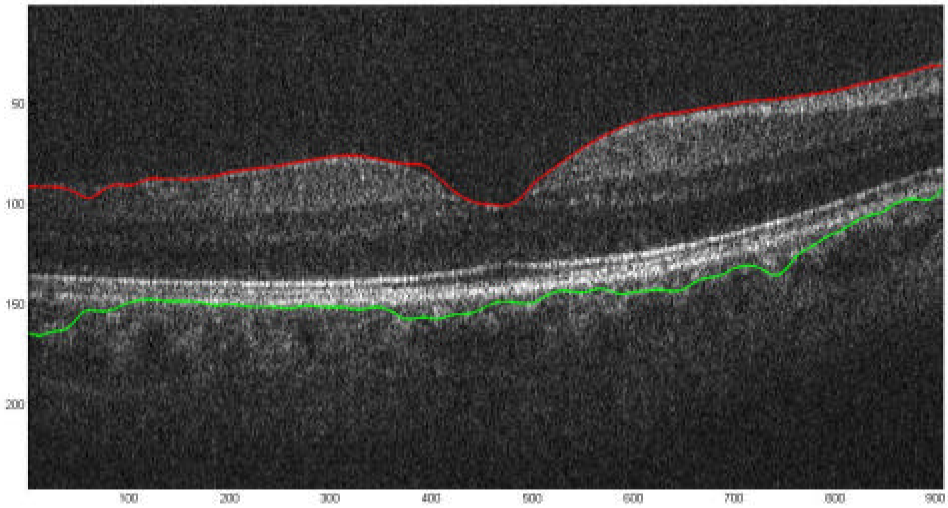


Figure 17 Top to bottom boundary detection using adaptive thresholding on a difficult, shadowed image of a healthy retina imaged by a Time domain OCT system

2.2.2 Texture-based Segmentation

Texture based segmentation, based on earlier work by Gavin Powell, was tried first. It is performed only on the region enclosed by the top (ILM) and bottom (RPE) retinal boundary. The basic algorithm scheme is given in Figure 18.

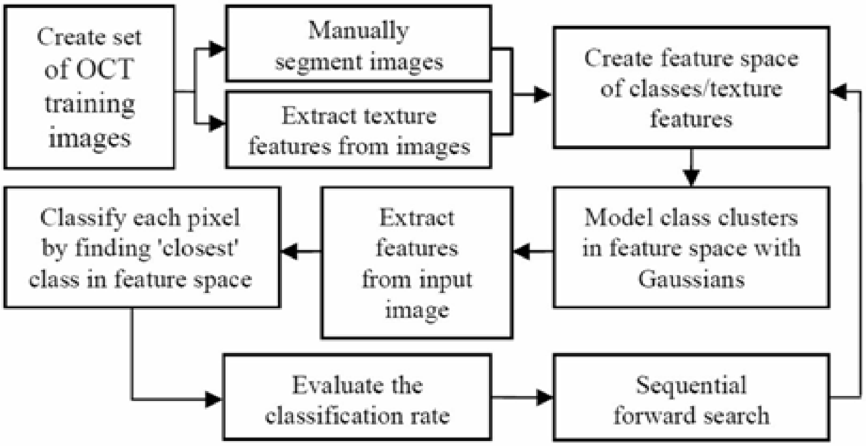


Figure 18 Texture based segmentation overview

The algorithm consists of a training and an application phase. In the training phase, a representative set of manually segmented images is used. The purpose of the training phase is to reduce the number of used features from currently 206 (including mean, entropy, Haralick, convex shell) to 70 by using sequential forward search algorithm (already mentioned in the segmentation section) and to obtain class distributions in the feature space, with every class representing a layer. Once the training is complete, the algorithm can be applied to the unseen data. Every pixel is simply mapped to the feature space and Mahalanobis distance is computed to the centre of each class. The smallest distance is chosen as the index to the belonging class. It is important to note that some geometrical features were used, that measure pixel's distance to the boundaries. Without these features, segmentation results tend to be patchy and inconsistent with layers overlapping. However, using geometrical features means that the training data layer geometry will be imposed on the unseen data. If the training set is too specific, the segmentation is too rigidly constrained. On the other hand, a varied training set does not constrain segmentation sufficiently to prevent the patchy, inconsistent results. This is illustrated best in the next section. Therefore, texture segmentation showed some potential but was unable to perform reliably on its own.

Texture based segmentation classifies each pixel as belonging to one of the layers and colours it accordingly. Results of the automatic segmentation can be compared to the manually segmented data and numerical evaluation can be used to rate the accuracy. While segmenting data that was very similar to the training data gave reasonably accurate results, the algorithm performed rather poorly on more diverse data (Figure 21). Since the errors were quite large, numerical measurements were not performed and focus was placed on extending and developing new segmentation approaches.

2.2.2.1 Results

Texture segmentation results of a healthy subject from the data set closely matching the training set are presented in Figure 19 with retinal and individual layer thickness maps shown in Figure 20.

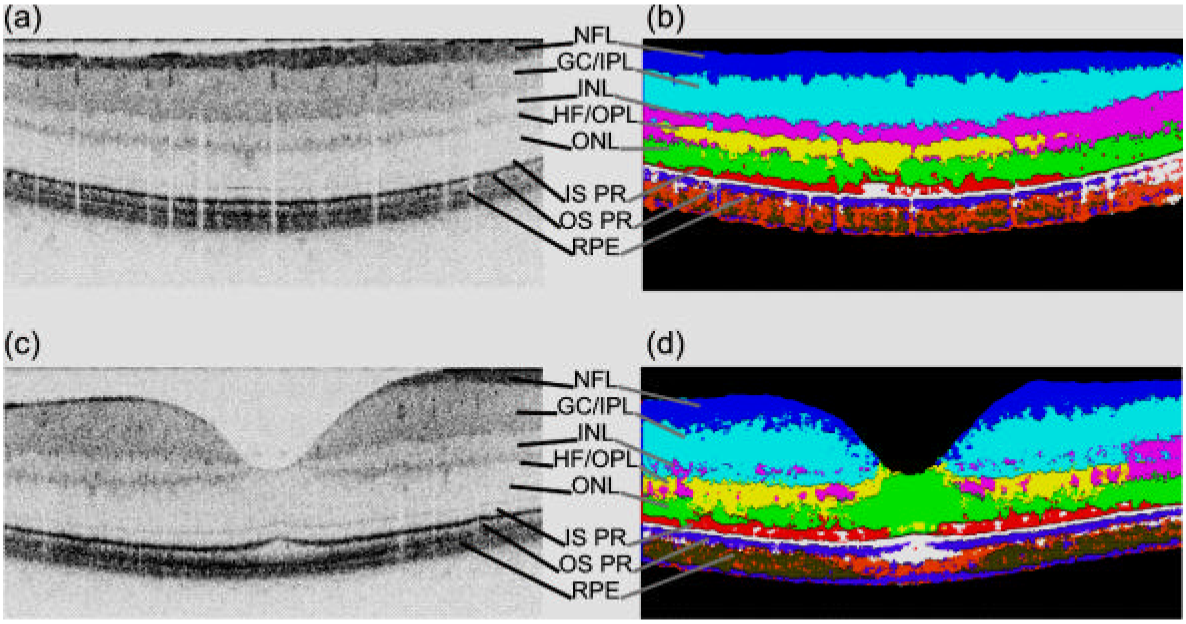
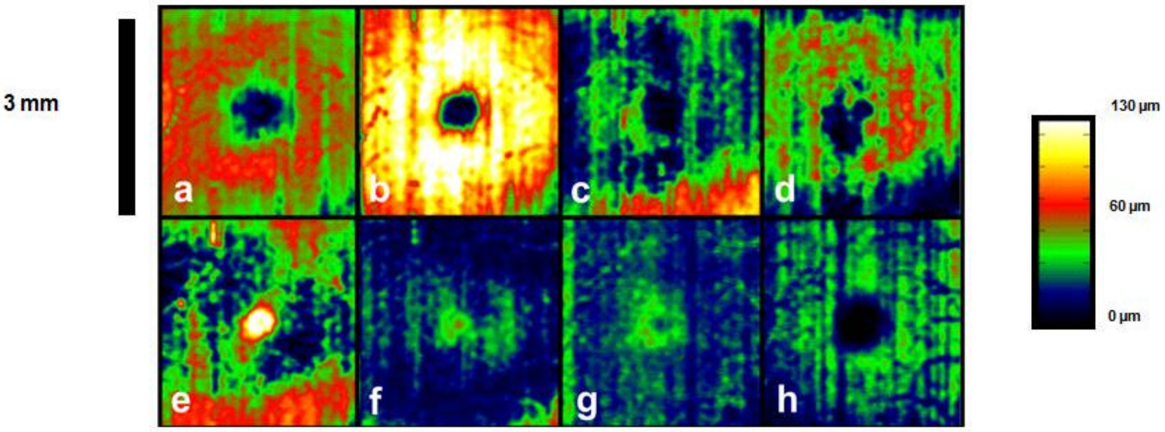


Figure 19 Texture segmentation results on a healthy retina



(a) NFL (b) GCL+IPL (c) INL (d) HF+OPL (e) ONL (f) IS PR (g) OS PR (h) RPE

Figure 20 The retinal thickness maps for a healthy subject

The obtained results show a decent performance of the texture analysis with segmented layers roughly corresponding to the real positions. However, as mentioned in the previous section, it is clear that the boundaries are not smooth but patchy since the classification of every pixel is independent. This effects mostly the segmentation of thin layers such as inner segment (IS), IS/OS, outer segment (OS) and RPE. It should be noted that some of the used features were geometrical, measuring the top and bottom retinal distance.

In Figure 21 results of the algorithm applied to two pathologies are shown. In Figure 21 a) the discontinuities in the IS, IS/OS and OS are visible in the segmented image. However, the segmentation in general is not accurate enough. It can be additionally seen how the geometry was imposed on the outer plexiform layer (OPL), ganglion cell layer and inner plexiform layer (GCL/IPL), with IPL/INL (inner nuclear layer) boundary not following the real boundary through the whole image.

Figure 21 b) shows the algorithm applied to the macular hole retina using the geometrical features, while c) displays segmentation results of the same image without the geometrical features being used. While in the b), results show what the ordering of the layers is correct, the segmented areas do not match the real image very well since the geometry of the training set is intrinsically imposed on the unseen image. In the c), row segmentation matches the real structure better, but the ordering of the layers is inexistent, with areas detected as NFL appearing in what is actually RPE because of similarities in texture.

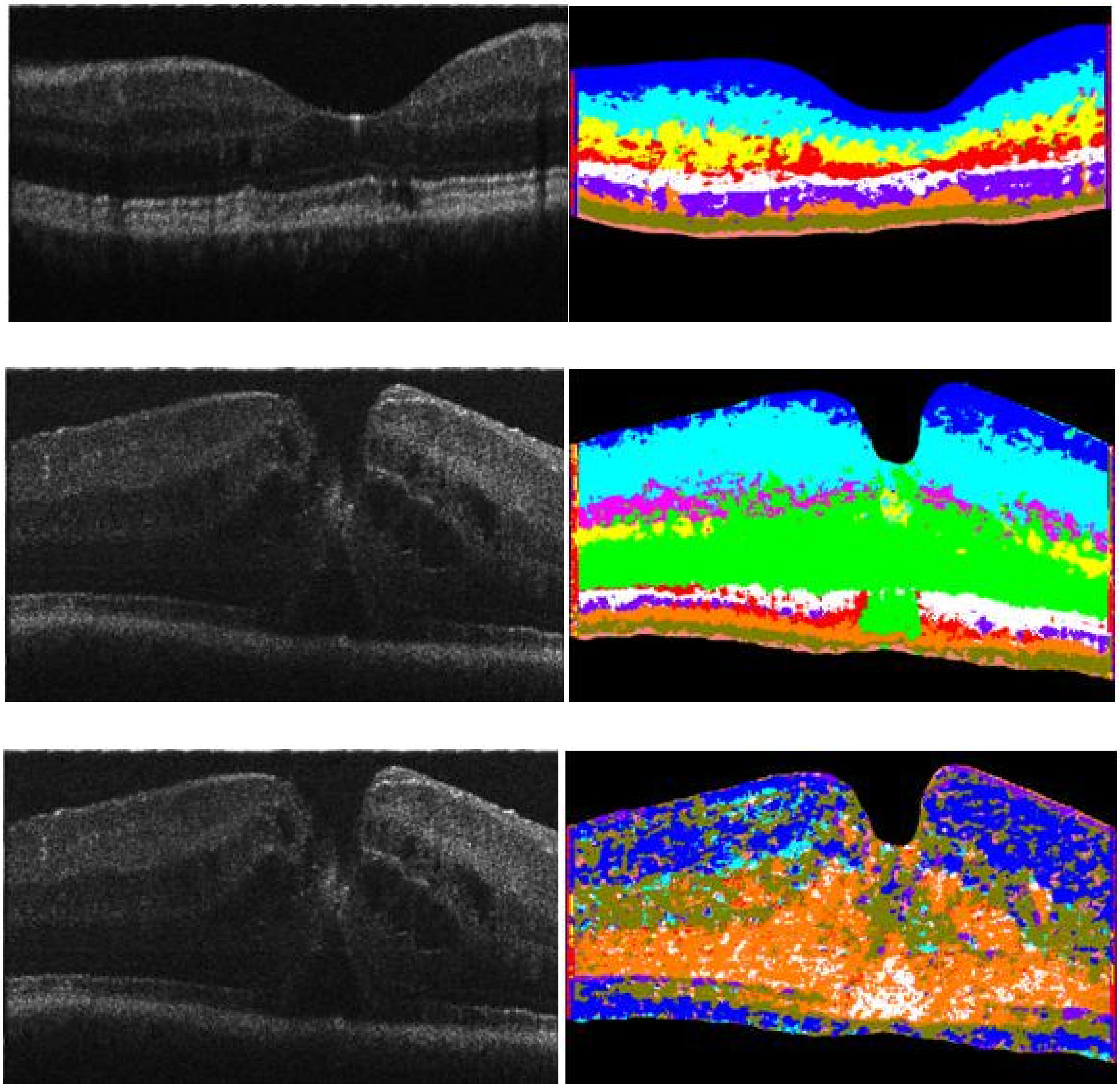


Figure 21 Texture analysis of pathologies with the algorithm trained on a healthy subject data set. a) shows result on a slight pathology, with localized damage to the RPE (arrow). The damaged area can be spotted in the segmentation result. However, the spatial constraints cause poor performance close to the fovea. In b), spatial constraints also cause poor performance, as the AMD geometry is significantly different from a healthy retina. In c), the same image is segmented without using the spatial features. Even though the layers are patchy and mixed, the segmented areas follow more closely the actual underlying geometry.

2.2.3 Adaptive Template Matching-based Segmentation

To address this problem, a geometry based, adaptive template matching solution was developed. The basic idea is to try to fit adaptive intensity based profile to each A-scan and assume that it will deform to best match the existing layer boundary. Figure 22 shows how a set of step adaptive templates are imposed on the A-scan intensity profile. The modelled layers include: nerve fibre layer (NFL), ganglion cell layer and inner plexiform layer (GCL+IPL), inner nuclear layer (INL), outer plexiform layer (OPL), outer nuclear layer (ONL), connecting cilia (CL), external limiting membrane (ELM), post-ELM (PELM), outer segment (OS), two segments of retinal pigment epithelium (RPE1 and RPE2). An active set MATLAB optimization method is used to deform them and best value is chosen as correct segmentation. Layer boundaries are simply the step transitions of the template. The optimization objective function is obviously the key for the algorithm performance. It is defined as a sum of normalized square difference of the template and the signal, the penalty for changing the template intensity and the penalty for changing the template step width with weights chosen heuristically. The optimization problem is constrained by forbidding height inversion of the neighbouring template steps. Also, the template step horizontal boundaries are not allowed to overlap. However, an individual A-scan is usually extremely noisy and it is very difficult to determine correct boundaries. To visualize this it is best to zoom into a retinal image and see how an A-scan looks like, Figure 23.

It is very difficult to determine correct layer boundaries looking at one A-scan even though by looking at the whole image this task is quite easy. Therefore, the algorithm has one more step. After every A-scan was separately segmented the connection phase will join the neighbouring A-scans into growing segments.

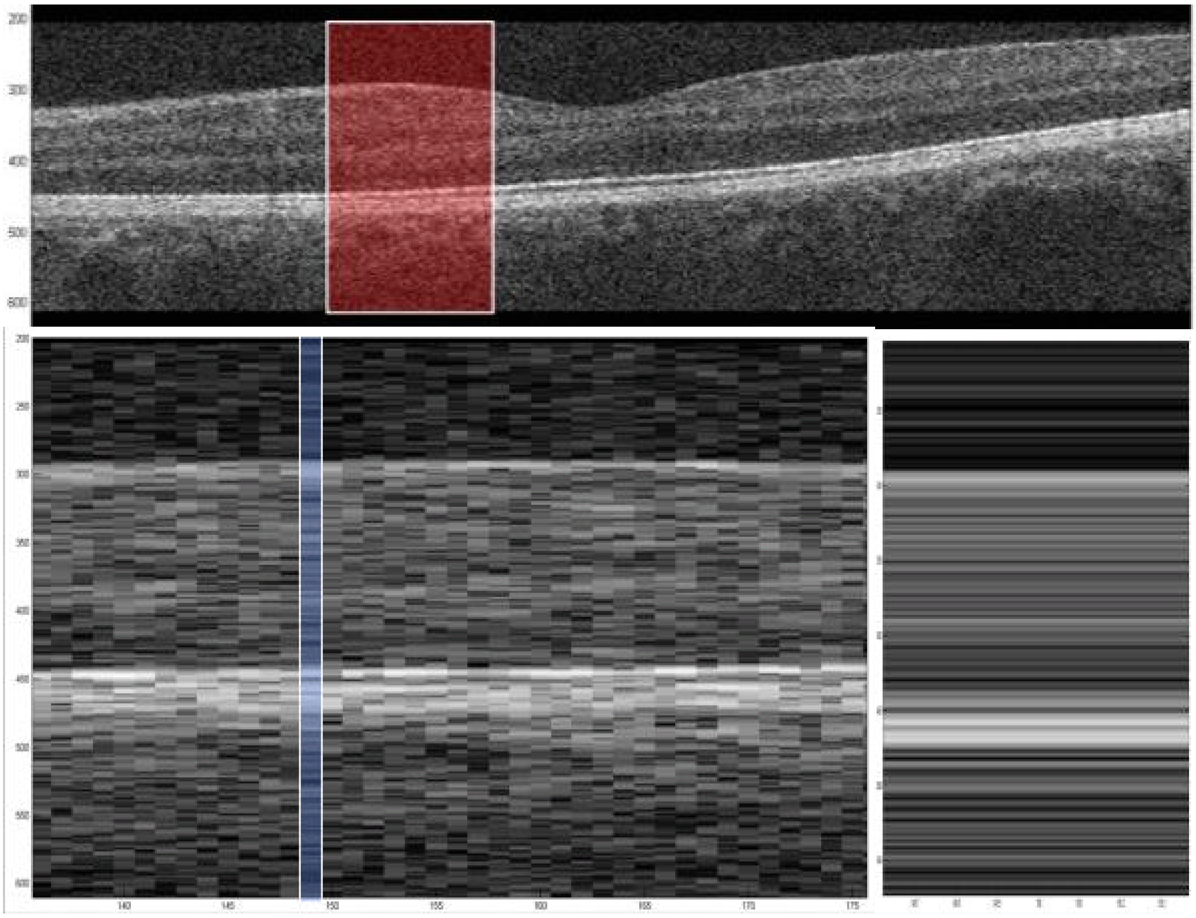


Figure 23 Zoom in to the A-scan level

This progression is shown in Figure 24. The left and right boundaries between the segments are shown in red and blue respectively, with rest of the template positions shown in green. First, the segments will start expanding to the “free” territory. As the segments come into contact they will merge if they are boundary-wise close enough, in other words below certain threshold. If no more segments can be merged the threshold is adaptively relaxed by a fixed step until all the segments are connected. When two segments actually connect, neighbouring values of each segment propagate through the other segment and replace the worse existing values.

However, this correction scheme does not correct the problem thoroughly. As registration should be done on the whole stack simultaneously, segmentation should be performed simultaneously on the whole image or volume using smoothness term in the objective function. To illustrate why this is important consider Figure 25. This figure illustrates two imaginary neighbouring A-scans. The blue line follows the strong edge while the red line follows the weak edge with reasonable smoothness. Using independent A-scan segmentation, the blue line would be chosen as the proper boundary, even though it is actually the red line that approximates it best. Therefore, the segment connection phase cannot correct for such a problem since it would impose one of the A-scan segmentations on the neighbour, while the correct solution is neither of those. Thus, to have a reliable method it is necessary to perform segmentation simultaneously. However, this introduces a new difficult problem to the optimization method used. The number of variables becomes large, even when reducing the original resolution for the rough fit. Even worse, the large number of inequality constraints makes the objective function domain extremely difficult to optimize. Trust region method implemented in MATLAB for the large scale problems and SPSA algorithm have been tried without any success. This result is not unexpected and leads to the model based approach as probably the optimal solution. It is important to note that the work done so far is a natural evolution of trying first the separate methods and then combining them together. It would be ungrounded to conclude at the start that a complex statistical model is needed if the simpler solutions have not been thoroughly tested.

2.2.3.1Results

The principle of adaptive template matching based segmentation is different than that of the texture segmentation. Instead of classifying each pixel individually, boundaries are adaptively fitted to give the best match according to the objective function. Thus, on the images shown, boundaries are marked with coloured contours and ten inner layer/interface boundaries were classified. The numerical comparison to the manually segmented data was not performed, as in some cases considerable inaccuracies were already visible from the visual inspection of the results. Thus, the subsequent model based approach was developed and evaluated.

A series of results obtained using adaptive template matching is shown in Figure 26 with thickness maps of two normals displayed in Figure 26. These results demonstrate that by using global geometrical information smooth boundaries can be obtained that match roughly the real boundaries. For major layers the obtained boundaries are usually close to their real values while the less prominent layers are less accurately positioned. In some areas matched templates do not follow the real boundary very well that can be explained by recalling that A-scans individually are very noisy. As was mentioned in the previous section, penalty term was used to contain too large shifts from the original template shape. That explains why the template will often not match well to the real boundary since this boundary has to be strong in order to attract the template despite the penalizing term. However if that term was not used, noise in the individual A-scans would be overwhelming, and likely cause even greater inaccuracies. Another problem is that, usually, in the whole processed stack, segmentation of some slices fails significantly in certain areas. This can be seen in thickness maps where certain spots are visible that actually do not match any real pathological thinning or displacement since normal subjects were segmented. Those spots correspond to the areas were the algorithm failed. The two pathologies shown in the last two rows of Figure 26 demonstrated the performance of the algorithm in difficult cases. A macular hole is especially difficult since this example shows an advanced pathology. It is encouraging that the RPE detachment was detected despite the major dissimilarity between this case and a normal retina.

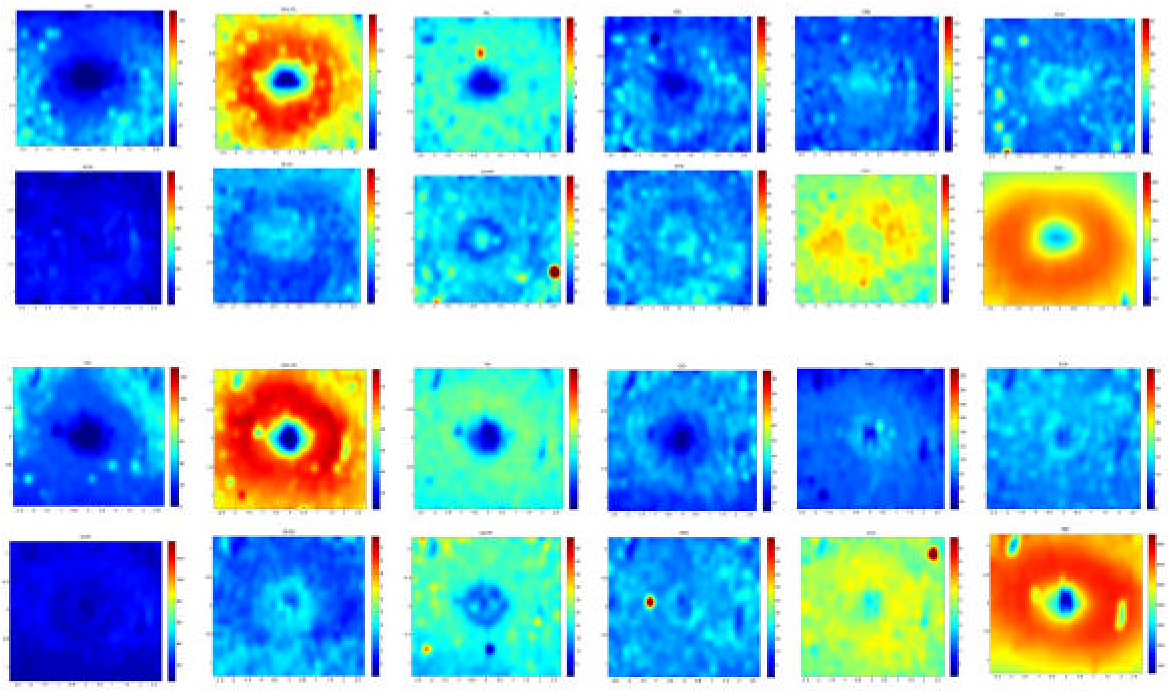


Figure 27 Thickness maps of eleven layers (nerve fibre layer (NFL), ganglion cell layer and inner plexiform layer (GCL+IPL), inner nuclear layer (INL), outer plexiform layer (OPL), outer nuclear layer (ONL), connecting cilia (CL), external limiting membrane (ELM), post-ELM (PELM), outer segment (OS), two segments of retinal

pigment epithelium (RPE1 and RPE2)) and retina obtained by applying adaptive template matching to two healthy subjects.

2.2.4 Model based retinal segmentation applied to normal human fovea

The idea behind the model based approach is to utilize information from the manually segmented example images (training data) to constrain the space of possible solutions to a plausible set. Generated speckle noise was progressively added to the data to test the robustness of the algorithm. Also, accurate segmentation is achieved in cases even when large shadows and low signal are present.

A three-dimensional OCT system was used for imaging. It uses a superluminescent light source, with 840nm central wavelength and 50nm optical bandwidth. Axial resolution is 5-6 microns, while transverse resolution is 15-20 microns. Data acquisition speed was 27 klines/sec. Optical power was 500 µW and SNR was 96dB with a sensitivity roll off - 6dB/mm. Depth range was 3.5mm and axial sampling 2.3 µm/vx.

2.2.4.1Algorithm overview

As seen in the overview of the algorithm (Figure 28), one can observe that the pre-processing stage is performed for both the training step and the segmentation of the unseen data. Once the variation parameters have been learned from the manually segmented training data, they can be used to drive the model to perform segmentation of unseen data. The actual segmentation process is essentially an optimization run that changes the model parameters in order to minimize the objective function that defines the difference between the model and a given unseen image that is to be segmented.

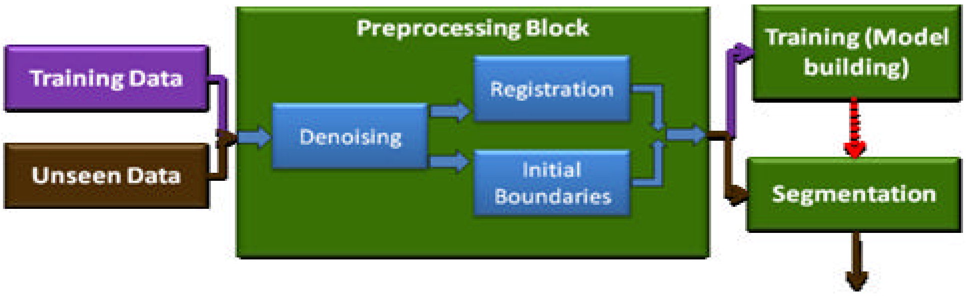


Figure 28 Algorithm overview: manually segmented data is used as the input to the training phase of the algorithm. After passing the pre-processing block, a statistical model is constructed that captures the variance in the training data, which can be then used to segment unseen data.

2.2.4.2 Pre-processing

Before the segmentation process, dual-tree complex wavelet (DTCW) denoising is applied to the data. The denoising algorithm exhibits very good performance, while being computationally efficient (Selesnick, Baraniuk and Kingsbury, 2005). This reduces the speckle noise present and thus makes the subsequent segmentation tasks easier. However, even without this step the model based approach outperforms the previously developed methods.

Denoising based on quasi-random nonlinear scale space described in (Mishra et al., 2010a) and applied to OCT speckle reduction in (Wong et al., 2010) would likely be more effective. It is an effective and fast method based on formulating the denoising problem as a general Bayesian least-squares estimation problem. A quasi-random density estimation approach is introduced for estimating the posterior distribution between consecutive scale space realizations. However, the relatively small performance difference (larger speed difference) in not significant for the performance of the statistical model, thus a well-tested and freely available DTCW code was used.

After that, registration of the stack and segmentation of the three initial well-defined boundaries (ILM, connecting cilia (CL) and end of RPE) is performed. Registration and initial boundary location finding are currently independent since detection of the initial boundary location operates on each B-scan independently.

A stack registration algorithm has been developed based on B-spline multi-resolution pyramid registration approach (Thevenaz and Unser, 1996) and (Sorzano et al., 2005). The basic algorithm for translation and rotation is used to register source to target image (7.1).

ILM, CL and end of RPE boundaries are found using an adaptive thresholding algorithm (auto adjusts to appropriate power, similar to 7.2.1) that converges to a close strong edge after the first estimate, additionally using constraints on distances between the boundaries. Robust polynomial fitting is afterwards used to eliminate outliers, followed by interpolation along the remaining points. ILM boundary is found first by starting the thresholding process from the top of the image, while RPE boundary is found next by starting from the bottom. The CL boundary is determined the last and depends on the positions of the already found ILM and RPE boundaries. It is found starting from the top after eliminating the pixels in the neighbourhood of the already found ILM boundary and imposing constraints on the distance from the RPE boundary. An example with a large shadowed area is shown in Figure 29.

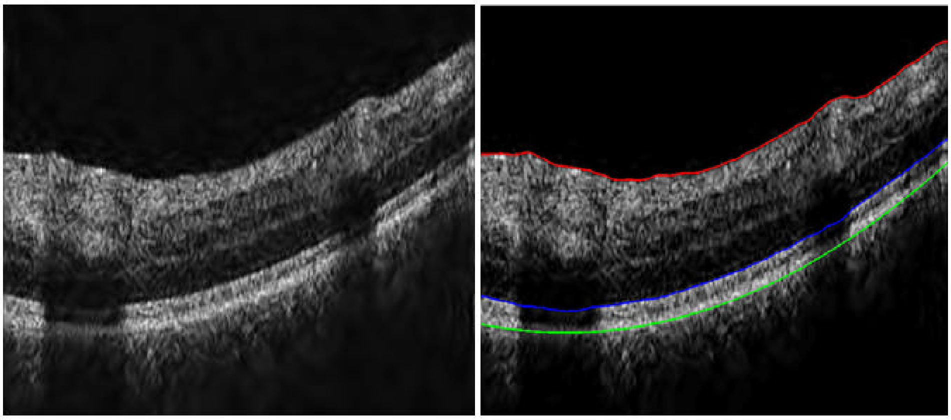


Figure 29 Initial segmentation step of a despeckled OCT frame (on the left) after adaptive thresholding boundary detection demarking (on the right): internal limiting membrane (ILM, red), connecting cilia (CL, blue), retinal pigment epithelium (RPE, green).

Initial boundary detection would also be possible based on a decoupled active contour (DAC) approach as presented in (Mishra, Fieguth and Clausi, 2010b). The level set active contour approach was tested and discarded for its slow convergence. However, DAC is both robust and fast, as it decouples the measurement (solved by using Hidden Markov Model (HMM) and Viterbi search) and prior active contour energy terms. As it was found that the initial boundary estimation approach was sufficient for the current application, not all of the other available methods were tested. For future work, however, algorithms such as DAC could prove valuable.

2.2.4.3Model building

After the pre-processing stage the statistical model is first trained on a set of manually segmented images and can be then applied to the unseen data. Using a statistical model based on the training data is a potentially effective tool for both segmentation and registration (T.F.Cootes, Edwards and C.J.Taylor, 2001). Its main advantage is that knowledge of the problem can be used to resolve the confusion caused by structural complexity, provide tolerance to noisy or missing data, and provide a means of labelling the recovered structures. The idea is to perform supervised learning by applying knowledge of the expected shapes of structures, their spatial relationships, and their textural appearance to restrict the automated system to plausible interpretations. Supervised learning is a type of machine learning for learning a function based on training data, which consists of pairs of input objects, and desired outputs. The task of the supervised learner is to predict the value of the function for any valid input object after having seen a number of training examples. To be useful, a model needs to be specific, capable of representing only legal examples of the modelled object.

From the manually segmented images, the shape and texture features were extracted and for each image all the extracted shape features are arranged into one vector and all the texture features into another vector. Separate models for shape and texture are constructed similarly, so only the shape model construction will be explained. If there are *m* training images, for each layer (*n* layers) one vector of offsets $\mathbf{v}$ is obtained per layer, per image of width w, which stacked together for all the layers define $\mathbf{x}$. All of the manual segmentations then comprise the matrix $\mathbf{X}$ Eq. (1).

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_m \end{pmatrix} = \begin{pmatrix} \mathbf{v}_{11} & \cdots & \mathbf{v}_{1n} \\ \vdots & \ddots & \vdots \\ \mathbf{v}_{m1} & \cdots & \mathbf{v}_{mn} \end{pmatrix} \tag{1}$$

$$\mathbf{v}_{ij} = [off_1 \ldots off_W]$$

Shape features that are used are sparsely sampled distances of the boundaries from the top boundary (ILM). Texture features that are currently used are simple, although it is trivial to include additional features if needed to further increase performance in cases of vessels, large shadows and pathological tissue; currently used features are the mean of all the pixels for each of the layers in the original image, standard deviation and mean of all the pixels for each of the layers in the median filtered image, as well as the multiple-scale (a pyramid of Gaussian filtered versions of the image) edges sampled along the boundaries. In practice, for an image of width 512, each boundary was sampled at 26 positions. Thus there are 26 spatial features and 4 texture features per each layer, and for eight layers, 208 spatial and 32 texture features were obtained.

Statistical models can reproduce specific patterns of variability in shape and texture by analyzing the variations in shape across the training set. It is difficult to achieve this selectivity, whilst allowing for natural variability, without using very large descriptors and thus it is essential to select good features from the training set for the model-building phase. The key step of the statistical model-training phase is the dimensionality reduction of the large set of features from the training data set. The reason for dimensionality reduction is to reduce the computational cost of the optimization method that is used to fit the model to the real data later on. The idea behind this concept is to find statistical dependencies between the produced features and reduce the dimensionality of the space by identifying only a certain number of the most prominent properties in the data set, represented by the most important eigenvectors.

Principal component analysis (PCA) is the standard vector space transform technique used to reduce multidimensional data sets to lower dimensions for analysis. It works by calculating the eigenvalue decomposition of a data covariance matrix or singular value decomposition of a data matrix. Usually a relatively small number of eigenvectors with greatest eigenvalues can describe the original data well. If $\mathbf{X}$ is the original data matrix, as defined in Eq. (2), after the decomposition only L principal components can be selected and in that way project the data into a reduced dimensionality space to get $\mathbf{Y}$ Eq. (2).

$$\begin{aligned} \mathbf{X} &= \mathbf{W}\Sigma\mathbf{V}^T \\ \mathbf{Y} &= \mathbf{W}_{\mathrm{L}}^T\mathbf{X} \end{aligned} \tag{2}$$

$\mathbf{W}$ is a matrix of eigenvalues of $\mathbf{XX}^T$, $\mathbf{V}$ is a matrix of eigenvalues of $\mathbf{X}^T\mathbf{X}$, while Σ is a diagonal matrix with nonnegative numbers on the diagonal.

However, rather than PCA, neural network based dimensionality reduction was used since it offers nonlinear eigenvectors and therefore can reduce the space more compactly if the data is nonlinearly distributed than the linear representation obtained by PCA (Scholz, Fraunholz and Selbig, 2007). The shape features proved to be nonlinear and thus a more compact representation was obtained using nonlinear dimensionality reduction that required a smaller set of eigenvectors than the PCA to capture the same variability within the training set. A Neural network (NN) is a mathematical or computational model based on principles found in biological neural networks. It consists of an interconnected group of artificial neurons and processes information where each connection between neurons has a weight, with the weights modulating the value across the connection. The training phase is performed to modify the weights until the network implements a desired function. Once training has completed, the network can be applied to data that was not part of the training set. It is useful to note that a special type of neural network (inverse) (Scholz, 2005) can be used to perform dimensionality reduction on the training feature set that is produced which contains missing values. Missing values occur when no data value is stored for the variable in the current observation. The generating function is used to produce larger dimensionality data $\mathbf{X}$ from the parameters $\mathbf{z}$ (equivalent to $\mathbf{W}_L$ in PCA) Eq. (3). The extraction function does the reverse.

$$\begin{aligned} \Phi_{gen} &: \mathbf{z} -> \hat{\mathbf{X}} \\ \Phi_{extr} &: \mathbf{X} -> \mathbf{z} \end{aligned} \tag{3}$$

The problem of missing data occurs because during the registration process slices are moved, and since input to the dimensionality reduction step has to be a rectangular matrix, it is necessary to fill the missing values. In practice these can be set as “not a number” (NaN) and perform the nonlinear PCA (Figure 30). After that, a reduced number of variables is obtained which can reasonably well describe any variation observed in the training data. The dimensionality of the original spatial feature space was reduced from 208 to 12, and the texture feature space from 32 to 2. This number of eigenvectors allowed for an efficient optimization in the subsequent steps, while still preserving the original data variation well.

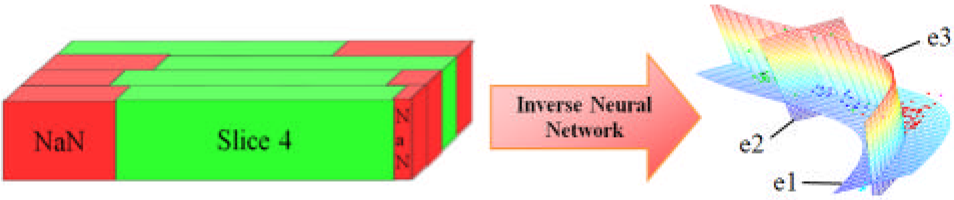


Figure 30 Filling the gaps after the registration with NaNs and applying inverse neural network nonlinear PCA dimensionality reduction. In the case of the example data shown on the right, it can be seen that already the first eigenvector (e1) captures most of the variance in the original data set. This illustrates the idea behind the dimensionality reduction.

The approach is based on a similar concept to the Active Appearance Model (AAM). For completeness, it will be first explained how the basic AAM model works, followed by an explanation of how the proposed statistical model differs from that concept. An AAM manipulates a model capable of synthesising new images of the object of interest by finding the model parameters which generate a synthetic image as close as possible to the target image (T.F.Cootes et al., 2001). An AAM will, based on learned shape deformation, generate a new image with a texture learned from the texture variation and then compute the distance between the synthesized and the given image that is to be segmented. $\mathbf{x}$ is the shape vector (which is normalized by subtracting the mean shape and rescaling, Eq. (4)) and $\mathbf{g}$ is the texture vector obtained from an image $\mathbf{I}$ and the shape vector (it is also normalized) Eq. (5).

$$\mathbf{x} \longrightarrow (\mathbf{x} - m(\mathbf{x})\mathbf{1})/s(\mathbf{x}) \tag{4}$$

$$\mathbf{g} = G(\mathbf{x}, \mathbf{I}) \tag{5}$$

Function $S(\mathbf{s})$ produces new shape vectors by adding the shape parameters $\mathbf{s}$ multiplied by the shape matrix $\mathbf{Q}_s$ (a matrix of sorted eigenvectors learned from the training set, usually produced by PCA decomposition) to the mean shape vector $\bar{\mathbf{x}}$ Eq. (6). The same procedure is used to generate new texture vectors.

$$\begin{aligned} \mathbf{x} &= S(\mathbf{s}) = \bar{\mathbf{x}} + \mathbf{Q}_s \mathbf{s} \\ \mathbf{g} &= T(\mathbf{t}) = \bar{\mathbf{g}} + \mathbf{Q}_g \mathbf{t} \end{aligned} \tag{6}$$

However, unlike the AAM which compares pixelwise synthesized images, the layer boundaries produced by the model were used during the optimization to compute texture features of the bounded area and compare them to the expected texture properties of each layer learned from the training set. This approach is used since unlike the areas in which AAMs are usually applied, the texture of retinal OCT scans varies so much within one layer that the direct comparison with a synthesized image is unusable. The objective function (Eq. (7)) evaluates how well the model matches real data and is minimized during the optimization.

$$f(\mathbf{s},\mathbf{t}) = \left| T^{-1}(G(S(\mathbf{s}),\mathbf{I})) - \mathbf{t} \right| + \frac{b * \sqrt{\sum (S(\mathbf{s})_b - \mathbf{b}_{init})^2}}{w} \tag{7}$$

b is the number of boundaries, w is image width and $\mathbf{b}_{init}$ defines the initial three boundaries positions found by the adaptive thresholding algorithm. T^{-1} is the inverse of T; T is defined in Eq. (6). T^{-1} returns the model texture parameters **t** that are most likely to generate a given vector of texture features **g**. The first term of the objective function defines the main measure for evaluation of the model fitting, determined by the difference of the model texture parameters and the texture parameters extracted from the image regions defined by the model shape parameters. The second term penalizes deviations from the initial boundary as found by the initial three boundaries algorithm and the one produced by running the optimization function for the statistical model. This is an important novelty, when compared to the standard AAM, which helps to constrain the optimization process to valid solutions. Additionally, the optimization process does not start from the mean of the model, but rather the median distance between ILM and RPE boundaries is found by the adaptive thresholding algorithm, as well as the ratio of the foveal pit distance to the greatest thickness found in the image. Using these values, the closest example from the training set is chosen and used to obtain the parameters for the initial model position. This way a faster and more robust convergence is ensured.

Another novelty is introduced in the second stage of the algorithm based on fitting a model for each independently used A-scan (depth-scan) to further improve the accuracy. This stage starts from the position defined by the result of the first stage B-scan fitting. The image area was heuristically divided into four segments and built an A-scan model for each segment since different types of variation can be expected at different offsets from the foveal depression. The A-scan model is trained on offsets produced by back projecting the manual segmentation data using the main B-scan model and computing the boundary offsets between the back projections and the original segmentations Eq. (8) (n is the number of layers and u is the number of A-scans from all the images in the given segment).

$$\mathbf{A} = \begin{pmatrix} aOff_{11} & \cdots & aOff_{1n} \\ \vdots & \ddots & \vdots \\ aOff_{u1} & \cdots & aOff_{un} \end{pmatrix} \tag{8}$$

In Figure 31 it can be seen how the second refinement stage of the algorithm improves precise tracking of the layer boundaries.

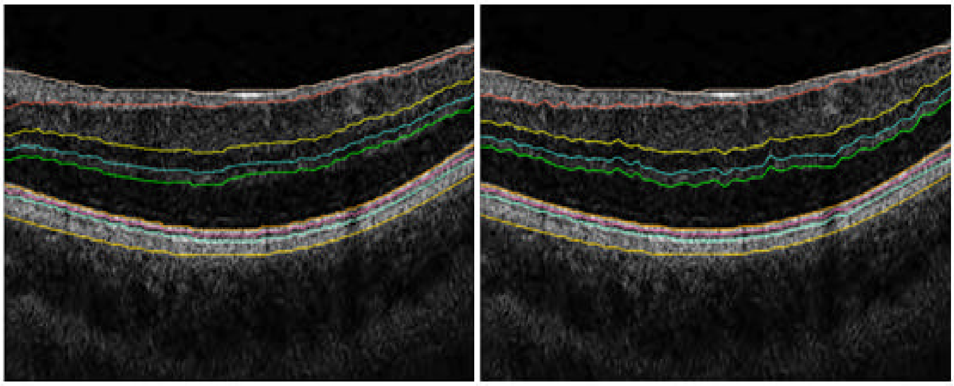


Figure 31 On the left is the result after the global low-res optimisation followed by, on the right, the refined result by the A-scan optimization.

2.2.4.4 Mechanical Turk

A large training data set has been efficiently obtained via an Amazon service called the Mechanical Turk (AMT), designed to offer a large international work force for completion of user defined tasks (Human Intelligence Task or HIT). Using this approach it was possible to evaluate the algorithm performance to a real golden standard, unlike the standard approach which can only evaluate repeatability of the algorithm in comparison to itself. In principle, the task of segmenting was divided to the detailed description of the task by a skilled person, manual delineation of the interfaces by a large number of less skilled workers, the comparison of multiple results for the same task and the supervision of the whole process by the skilled operator. Two account types are used: worker and requester. The worker account type is used for performing the tasks, while the requester type is used for submitting them. Submitted tasks are usually simple but it is possible to define criteria for the workers and in that way use skilled workers, for larger payments, of course. In this case no testing was performed for selection of the workforce apart from the general ranking of a worker based on previous performance recorded by the AMT-system. However, it was necessary to supervise the work relatively often and update the instructions based on the input from workers and give bonuses for good work to stimulate reliable workers to continue doing the provided tasks.

The architecture of the whole system is comprised of a web page with JavaScript to handle the user input that was designed through the AMT interface. Inside a Java applet was embedded through which the workers perform the segmentation. For storage of the B-scans, example images and results to be saved, Amazon S3 storage service was used. 505 B-scans were submitted and each image was set to be segmented twice by different workers respectively. That way, the inter worker variability can be computed and inaccurate results left out, while still having another one which is usually good. Inter worker variability was computed only on the images for which both results were deemed to be accurate. Bonuses were paid out for good work, approximately equivalent to the initial payment. In cases of inaccurate or inappropriate results it is not necessary to pay the worker. Since the behaviour of the AMT system can be better described by the rules of sociology than simple mathematical relations, the processing speed is nonlinear. It is important to note that while half of the results were obtained in just a few days; it usually took significantly longer to get all the tasks completed. That is not a problem since it is possible to use results as they are produced without having to wait for the completion of the whole batch. One most likely reason for the reduced work speed is that workers use the default sorting for viewing the available tasks, based on the number of available tasks. It took four weeks to complete the segmentation of about 2700 images. Workers seemed to be more interested in the task once the purpose of the work was given in the introduction and it was pointed out that it serves a valuable medical goal. A few questions were included in the form of a web form so that workers can give the feedback on the work that they were doing.

2.2.4.5 Results

For evaluation purposes 466 manually segmented B-scans were used, almost (in some cases the manual segmentation had to be discarded) uniformly sampled from 17 eyes (each stack contains 128 B-scans). The performance of the algorithm was tested on this data set using the leave-one-out test; all data from one person was iteratively left out, the model trained on the remaining data and then the performance tested on the data from the person left out. This procedure is performed for each person in the training set. This way it was ensured that the performance of the algorithm is tested on the “unseen” data.

For evaluation, automatic segmentation results were compared to manual segmentation done by the AMT workers. Two types of error measures were used, computed for each boundary *i* separately and from these error measures for an entire B-scan or for an individual layer were computed, Eq. (9).

$$ {}^{i}E_B = \sum_{j=1}^{j=w} \left| yAut_{ij} - yRef_{ij} \right|, {}^{i}E_{LDEV} = \sqrt{w * \sum_{j=1}^{j=w} (yAut_{ij} - yRef_{ij})^2} \tag{9} $$

E_B (Basic) is the basic error measure that defines the number of misclassified pixels. E_{LDEV} (Layer DEViation) uses the $\sqrt{w}$ term for normalization so that for the special case when the two boundaries are equally distant from each other along their whole length ($yAut_{ij} - yRef_{ij} = d$ for all *j*), it is equal to E_B (proved in Eq. (10)).

$$ \begin{aligned} {}^{i}E_B &= \sum_{j=1}^{j=w} \left| yAut_{ij} - yRef_{ij} \right| = \sum_{j=1}^{j=w} |d| = w * |d| \\ {}^{i}E_{LDEV} &= \sqrt{w * \sum_{j=1}^{j=w} (yAut_{ij} - yRef_{ij})^2} = \sqrt{w * \sum_{j=1}^{j=w} d^2} = \sqrt{w * w * d^2} = w * |d| = {}^{i}E_B \end{aligned} \tag{10} $$

For all other cases E_{LDEV} is larger than E_B. Thus E_{LDEV} will penalize large deviations from the reference position of a boundary, unlike E_B which only measures the number of misclassified pixels. E_{LDEV} is therefore useful for penalizing specific types of poor algorithm performance which could show as, for example, a large jump in a boundary position that could be narrow and thus not affect E_B significantly since the misclassified area would be relatively small.

The error for a whole image (this refers to both E_B and E_{LDEV}) is defined in Eq. (11).

$$E = \frac{\sum_{i=1}^{i=b} {}^{i}E}{A} \qquad (11)$$

${}^{i}E$ is the error for each boundary and A is the area between top (ILM) and bottom boundaries (RPE/CH).

In the case when the error for layer *k* is expressed separately, instead of summing up across all boundary errors, only the two boundaries that define a layer are added and divided by the sum of the layer area as given by the automatic segmentation (A_A) and the layer area as given by the reference segmentation (A_R), Eq. (12). This approach is used to normalize for double counting of misclassified pixels, as each layer is bounded by two boundaries.

$$E = \frac{\sum_{i=k}^{i=k+1} {}^{i}E}{A_A + A_R}, 0 < k < b \qquad (12)$$

A confidence measure could be introduced based on the values returned by the objective function after the optimization step. Large values are proportional to the low confidence in the boundary positions determined by the model fitting. This would be useful for the operator to decide whether the obtained results are reliable.

In Table 1 the inter-worker variability of the manual segmentations used in training is presented for each boundary and in total, while in Table 2 and Table 3 results are presented for both the initial segmentation and after the second step refinement.

Table 1 Variability of manual segmentations on 75 B-scans in percent (the data has been previously examined and "bad" results left out)

Error Type	NFL	GCL +IPL	INL	OPL	ONL	CL	OS	RPE	Total
E_B	13.6	11.4	22.8	25.0	6.0	28.0	23.3	18.7	16.1
E_{LDEV}	17.9	14.4	28.4	31.3	7.4	35.4	28.6	22.5	19.9

Table 2 Error values on 466 B-scans at various positions from 17 eyes in percent before the A-scan optimization

Error Type	NFL	GCL +IPL	INL	OPL	ONL	CL	OS	RPE	Total
E_B	23.2	14.3	31.6	41.9	8.6	35.2	32.1	22.1	22.4
E_{LDEV}	31.9	17.4	39.7	55.4	10.7	47.3	41.0	27.0	27.8

Table 3 Error values on 466 B-scans at various positions from 17 eyes in percent after the A-scan optimization

Error Type	NFL	GCL +IPL	INL	OPL	ONL	CL	OS	RPE	Total
E_B	20.0	10.1	22.1	31.6	7.1	34.9	30.8	21.6	18.7
E_{LDEV}	29.2	13.2	30.4	46.4	9.3	47.1	39.5	26.5	24.2

It can be seen that the total error rates (especially E_B which is the main measure) are close to the inter-operator variability (18.2% compared to inter-operator's 16.1%). E_{LDEV} difference is somewhat larger. Thus, it can be concluded that the algorithm performance is almost the same as ground truth.

The algorithm performs well even when artefacts are present, such as strong shadows, which can cause problems for less robust algorithms (Figure 32).

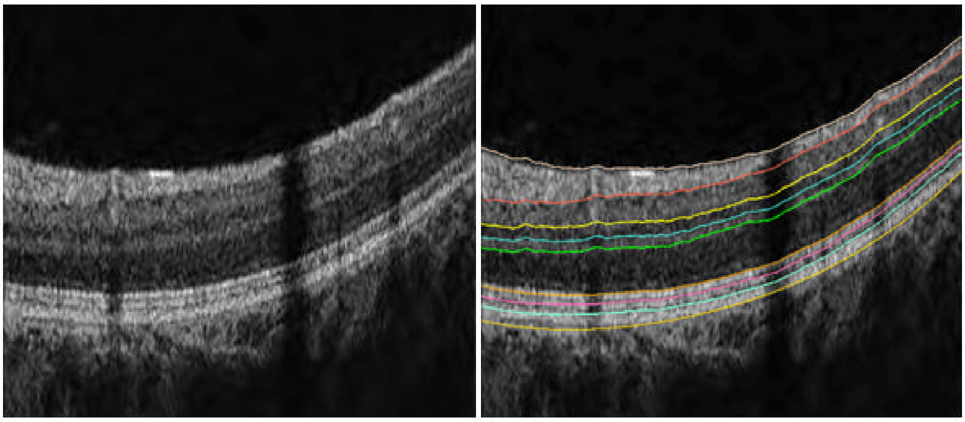


Figure 32 Robust performance for all the layers is achieved even in presence of shadowing. A despeckled image is shown on the left; the segmented image is on the right.

In Figure 33 thickness maps are shown for 17 different eyes after registering them and computing median and coefficient of variation (expressed as absolute variation in pixels), since it would take too much space to present the results for each eye individually. It can be seen that despite the data being affected by artefacts, the results are accurate and show larger variation only around the foveal pit region, as can be expected.

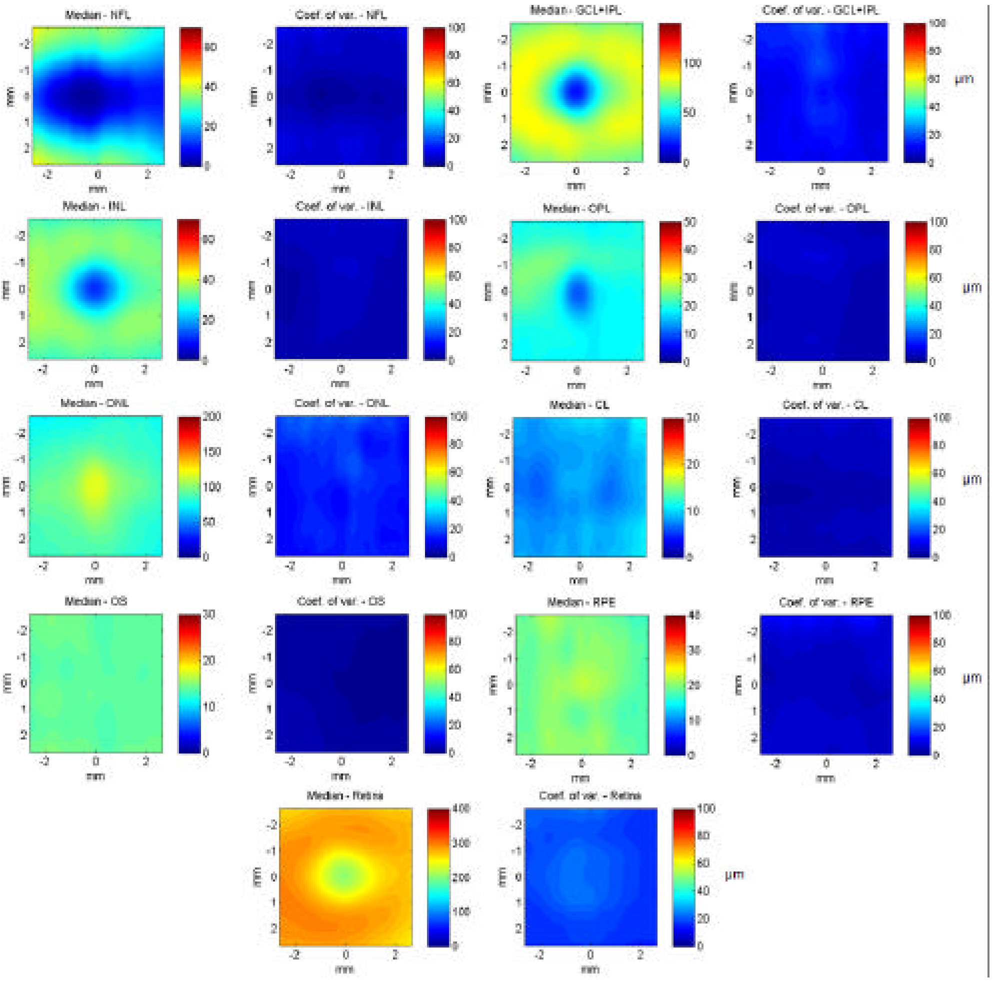


Figure 33 Median and coefficient of variation computed on thickness maps of all the individual layers (nerve fibre layer (NFL), ganglion cell layer and inner plexiform layer (GCL+IPL), inner nuclear layer (INL), outer plexiform layer (OPL), outer nuclear layer (ONL), connecting cilia (CL), outer segment (OS), retinal pigment epithelium (RPE)), as well as the retina, obtained from 17 eyes.

To evaluate the performance of the algorithm in conditions of increased noise (reduced dynamic range) that frequently occur in clinical measurements for a number of reasons (opaque cornea of cataract lens, residue in vitreous humour, non optimal imaging conditions, etc), background noise (speckle, multiplicative random noise) has been added to tomograms (Figure 34) and results plotted on a graph. The background was generated using a texture synthesis approach (Efros and Freeman, 2001). This enabled for efficient creation of a different speckle noise pattern for each image even though they are all based on the same physical speckle template, which is only one image of background noise with the typical spatial frequency distribution. Using this approach an arbitrary number of synthetic, but uncorrelated and realistic, images of background noise can be generated that are added subsequently to each given image to simulate low dynamic range. Using this novel approach it was possible to evaluate the performance in an arbitrary range of realistic speckle conditions. The algorithm shows robust performance under such conditions shown by a gentle rise of the error/dynamic range curve.

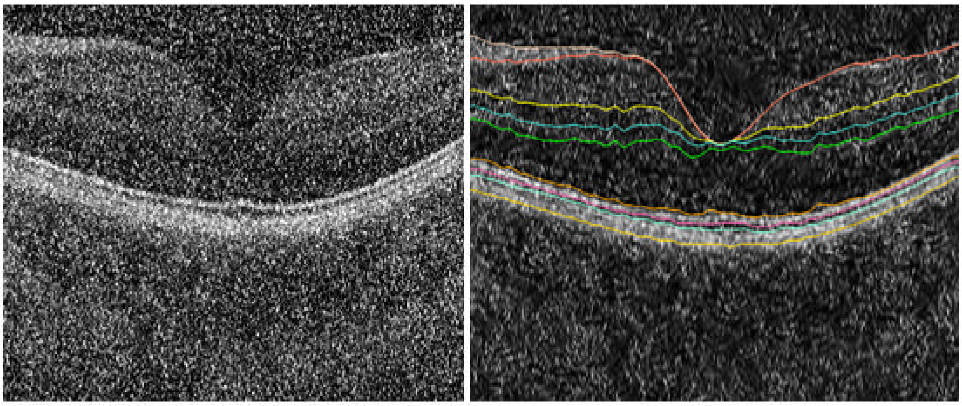


Figure 34 Segmentation in a case of added strong noise. Left original image. Right filtered, denoised image with segmentation results superimposed.

This can be seen in two graphs showing error rates E_B and E_{LDEV} plotted versus the dynamic range for a set of images for all the layers combined and with the confidence interval (1.96 std. dev.) plotted as dashed lines (Figure 35), as well as two graphs showing the error rates for each individual layer (Figure 36). The individual boundaries most affected by decreasing dynamic range are those defining INL and OPL, as could be expected since these layers exhibit normally significant variation and have weak boundaries which are affected early by the noise increase. Also, the boundaries between CL, OS and RPE are difficult to determine.

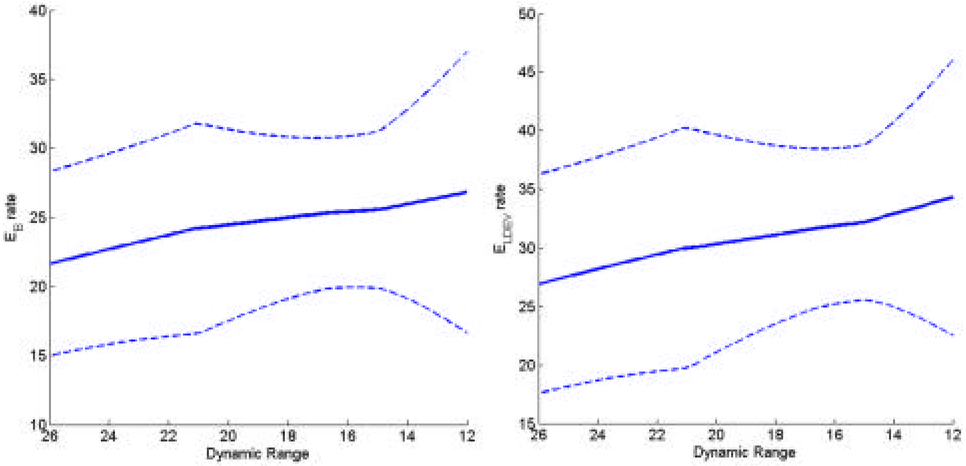


Figure 35 Error rates E_B (Basic) and E_{LDEV} (Layer DEViation) with decreasing dynamic range for all the data sets, with confidence interval (1.96 * standard deviation) marked by the dashed lines. For both error measures a slow rise in the error values can be observed, which guarantees robust performance with noisy data.

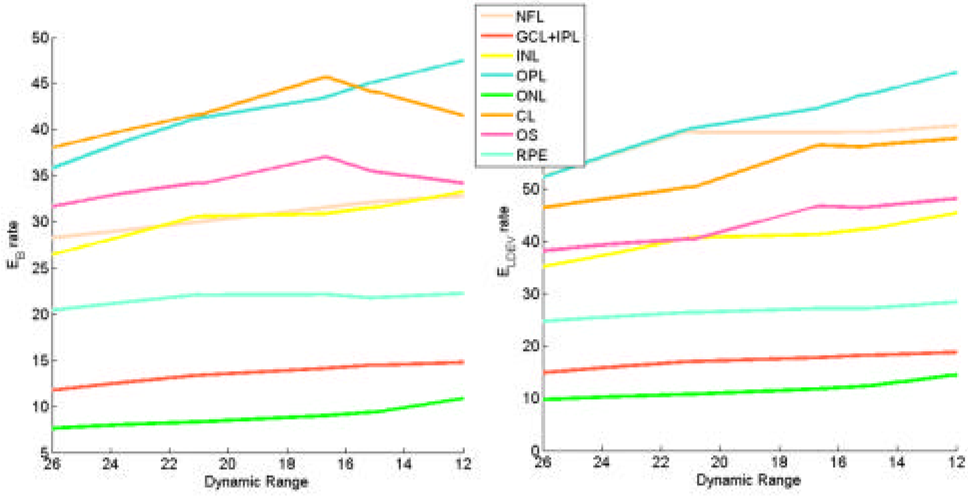


Figure 36 Error rates for the individual layers E_B (Basic) and E_{LDEV} (Layer DEViation) with decreasing dynamic range for all the data sets. For all the individual layers (nerve fibre layer (NFL), ganglion cell layer and inner plexiform layer (GCL+IPL), inner nuclear layer (INL), outer plexiform layer (OPL), outer nuclear layer (ONL), connecting cilia (CL), outer segment (OS), retinal pigment epithelium (RPE)) a slow rise in the error values can be observed. Thin layers inherently exhibit greater error values, as both errors are normalized by the layer area.

2.2.4.6 Conclusion

An algorithm for automatically segmenting all major retinal layers based on a novel statistical model has been developed. Two important novelties with respect to the standard AAM were introduced: a second term in the optimization function that penalizes large deviations from the three boundaries found by the adaptive thresholding algorithm and the second algorithm stage that refines the model fit for each A-scan independently, giving increased accuracy.

It has been thoroughly tested and evaluated against a manually segmented large dataset from a 800 nm OCT system. The algorithm proved highly robust in full foveal scans even in the presence of artefacts and added strong background noise that reduce dynamic range down to 12dB. It is the first time that a large, representative data set (466 B-scans from 17 eyes) has been used for evaluation of an OCT segmentation algorithm. Manual segmentations of the large data-set were used as ground truth, rather than the frequently used error computed between the results of the algorithm on inter-visit measurements, as it is susceptible to underdetermine the real error value, likely ignoring systematic error of the algorithm. Apart from the basic error measure that counts the number of the misclassified pixels, a second error measure was used to penalize large deviations from the ground truth.

It can be concluded that the algorithm successfully demonstrated reliable performance under conditions which prove extremely challenging for the pre-existing methods. It was used as a basis for the development of the model based choroid segmentation algorithm described in 2.3.2, proving that indeed it is well suited to the weak and missing data conditions. It would be also possible to extend the proposed algorithm to segmentation of pathological cases, as well as segmentation of ONH (optic nerve head) scans which contain discontinuous boundaries. In cases when the stack registration is very precise, the initial ILM and RPE boundary finding step could be replaced by the algorithm proposed by Fabritius et al. (Fabritius et al., 2009b) that relies on full 3D information present in the stack, since it is very efficient. Clinically, fully automated segmentation of all major layers is essential in making medically useful the possibilities given by the method of high resolution, high speed OCT of large portions of the human retina at microscopic detail.

2.3 Choroid segmentation

For segmentation of choroid affected often by extreme pathologies, it was necessary to develop a more robust top and bottom boundary detection algorithm. subchapter 2.3.1 presents how it works and why it is applicable to a large problem area outside OCT. In subchapter 2.3.2 choroid segmentation is explained using a statistical model similar to the one developed for the foveal segmentation (2.2.4)

High-speed 3D OCT-imaging at 1060 nm was performed with less than 2.5 mW at the cornea, well below the maximum power limit for 10-second exposure. Three-dimensional OCT volumes were acquired with 15 to 20 µm transverse resolution, approximately 7 µm axial resolution, and 512 voxels per depth scan (A-scan). Scans across a 36° by 36° ?eld with 512 by 512 A-scans at 47,000 A-scans/s were centred on the fovea and resulted in to 70 frames/s (B-scans/s). The spectrometer utilized a line camera with 1024 px (Goodrich SU-LDH2) clocked at 92 kHz linerate, which led to 512 voxel in each depth scan between the zero delay and the maximal depth. The signal-loss of -14dB along the depth scan was counteracted by placing the closer, less attenuated side of the tissue towards the zero delay. For registration and noise reduction ImageJ software was used. The OCT volume was averaged in both transverse directions within a ?eld of approximately 1°, to remove speckle and increase sensitivity. Axial choroidal thickness (ChT) was defined as the distance between the centre of the signal peaks originating from the RPE/Bruch's membrane/choriocapillaris (RBC) complex and the choroidal–sclera interface (CSI).

The pre-processing stage is performed for both the training step and the segmentation of the unseen data. Once the variation parameters have been learned from the manually segmented training data, they can be used to drive the model to perform segmentation of unseen data. The actual segmentation process is essentially an optimization run that adjusts the model parameters in order to minimize the objective function that defines the difference between the model and a given unseen image that is to be segmented. The overview is shown in Fig. 37.

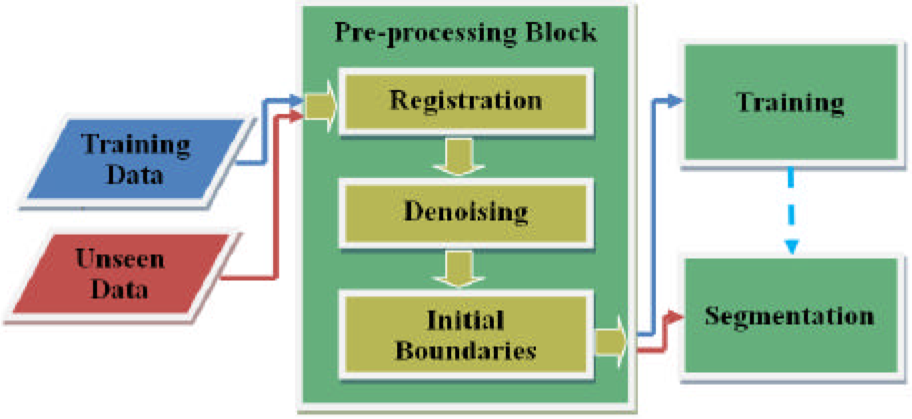


Fig. 37 Algorithm overview: both the training and unseen data pass through the pre-processing block. The registration is used only for averaging in z-direction as a despeckling procedure, while the segmentation is independent for each B-scan. A statistical model is subsequently constructed that captures the variance in the training data, which can be then used (dotted line) to segment unseen data.

2.3.1 Choroid initial boundary detection

To successfully segment internal limiting membrane (ILM) and RBC boundaries of choroidal data several challenges have to be overcome. RBC defines the top choroidal boundary and is used as initialization for the statistical model that subsequently finds the CSI (choroidal sclera interface) boundary. ILM is relatively easy to find and is of clinical use, thus it is segmented as well.

The RBC is usually almost at the same position as RPE, but in cases of RPE detachment it is often not. The OCT data obtained was from a 1060 OCT system by scanning over wide angle, which causes low signal strength on the image edges, especially towards the stack start and end, where the retina is narrow (Fig. 38 a), the right side of the image). Another problem were various pathologies that present different challenges; the boundaries can be discontinuous, also their shape is highly unpredictable and of highly changing curvature, so it is impossible to efficiently parameterize them, using a small set of parameters, with polynomials or other models that can be used for fitting. The third problem is that the data itself is often coming from different OCT systems and pre-processing pipelines; obviously it is highly beneficial if the correct segmentation can be achieved without retuning the algorithm parameters, thus a robust approach is required.

To meet all these goals an algorithm has been devised that consists of several basic steps, however special care was taken to keep the number and importance of free parameters as small as possible. This makes the algorithm intuitive and general while avoids over specializing.

First non-signal areas at the edges of an input image (Fig. 38 a)) are found by performing Canny edge filtering using large sigma (32) and high threshold (0.35). The parameters were heuristically chosen to extract only the major edges in an image. Columns on the sides where there is no edge detected are considered to be non-signal areas. The first column with an edge detected is considered to be the start of the signal area; this is determined both from the left and the right side (Fig. 38 b)).

A modified Dijkstra's algorithm is used to find a minimal path from the left to the right edge of an image, representing the RPE boundary, as it is almost always the brightest and longest one (in the sense that it extends furthest to the sides of the image). It is important to note that this algorithm has been used only to find the RPE which is a high intensity boundary, unlike some previous approaches which used path finding algorithms to find all the retinal boundaries, making the segmentation sensitive to any changes in the input data.

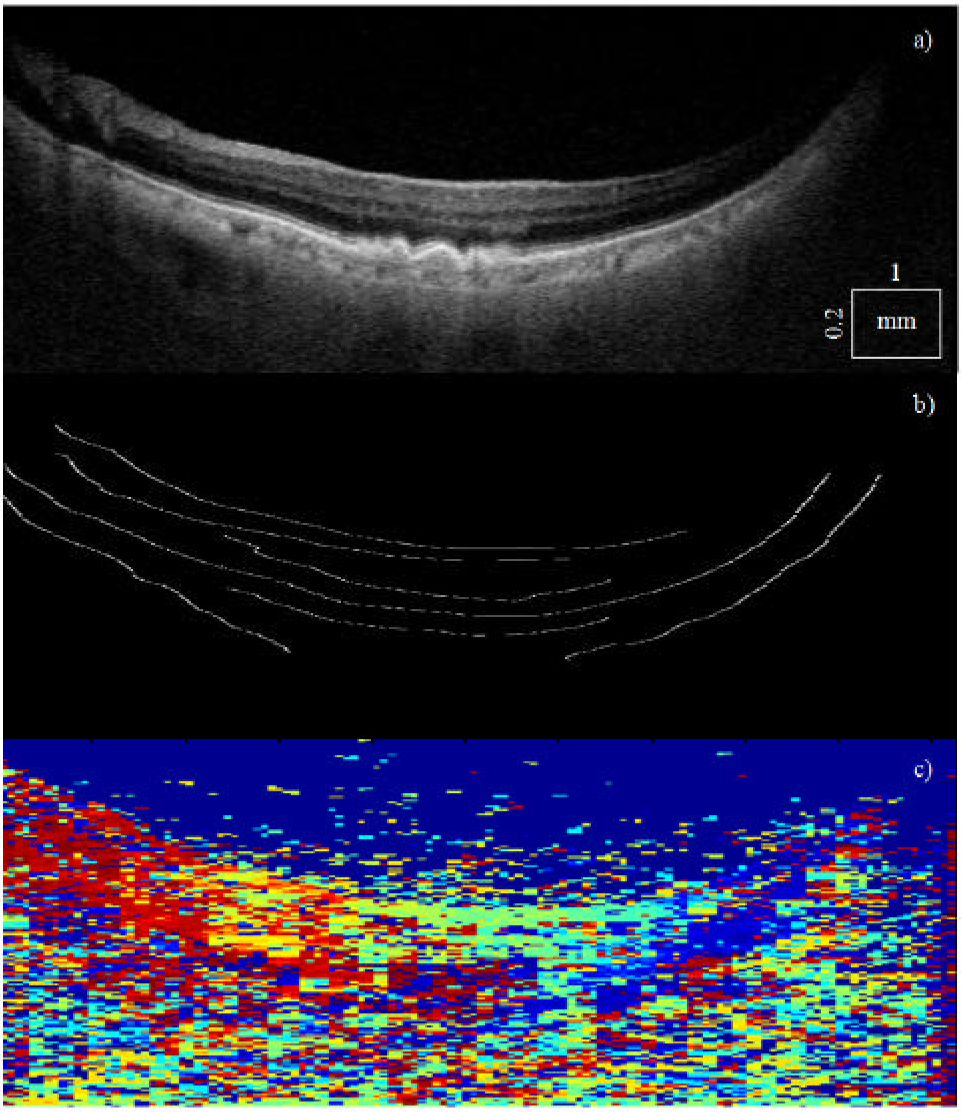


Fig. 38 Pre-processing steps for finding ILM and RBC: a) is the original image, b) is obtained after Canny edge filtering, c) are colour coded dominant wavelet domain orientations (from dark red "descending" (-90deg) angle, to dark blue "ascending" (90deg))

A graph is constructed, where each node represents the inverted pixel intensity in an image (i.e. the dark pixels become bright and bright pixels dark). Since the path to be found needs to be minimal, traversing bright pixels will mean going through low intensity values in the inverted image. A diagonal block connectivity matrix **A** consists of $m*n$ rows, where m is height and n is width of an image **I**. Thus, every row represents connectivity of certain pixel to pixels in the next column of the original image. Considering that the boundary cannot track backwards or skip columns, matrix **A** has only m columns, as each pixel is connected only to the next column of the original image. In this case it was assumed that plausible steps from column to column do not exceed 10 pixels, thus *jmp* was set to 10, Eq. (13). Vector **c** represents connectivity of a particular node (pixel) to other nodes and it is defined within the *jmp* range, with the rest of pixels set to 0 (no connection).

$$\mathbf{A} = \begin{pmatrix} \mathbf{c}_{11} & & & & 0 \\ \vdots & \ddots & & & \\ 0 & \mathbf{c}_{1m} & & & 0 \\ \vdots & & \ddots & & \\ 0 & & \mathbf{c}_{n1} & & 0 \\ \vdots & & & \ddots & \\ 0 & & & & \mathbf{c}_{nm} \end{pmatrix}$$

$$\mathbf{c}_{ij} = 1 - \mathbf{I}(x, j+1), \max(0, i - jmp) < x < \min(n, i + jmp) \qquad (13)$$

To increase robustness of this approach it was expanded in two ways. Firstly, derivatives were added to prevent unrealistic boundary jumps caused perhaps by just slightly more favourable intensity information.

Second, a complex dual-tree wavelet decomposition, being multi-scale, orientation sensitive and translation invariant, making it more efficient than standard edge enhancement methods, such as Sobel, Prewitt, Roberts or Canny, was used to obtain the edge orientation information (Fig. 38 c). Smooth dominant directions were calculated from -90 (dark red) to 90 (dark blue) degrees, which was applied in the modified algorithm to obtain the third term expressing how well the path is aligned with the real edges in an image. This prevents path drifting over high intensity areas that possibly have completely different orientation, thus producing a non-realistic path.

Unlike working with static weights and having a complicated workaround the problem that each node can be part of multiple paths and thus cannot be assigned an optimal weight using static values (Perkon et al., 2010), the problem has been solved by making an elegant extension to the optimal path algorithm itself, while preserving the same time complexity.

The Eq. (14) shows the modified algorithm distance function.

sc is the current node index, corresponding to any pixel in the original image, *xc* and *yc* are x and y coordinates obtained of the node *sc*, **y** are the neighbouring column's connected nodes (pixels), *yp* is *sc's* parent's y value (inherently always known from the path history), $\mathbf{y}''$ is the second derivative of the potential path, while $\overline{\mathbf{y}'}$ is the derivative mean used as an approximation to obtain the angles **j** of the potential path. Weights $\mathbf{ew_1}$ and $\mathbf{ew_2}$ indicate how close to each of the two neighbouring out of 6 total discrete orientations possible path is. **IW** is the tensor containing directional wavelet coefficients (amplitudes) for each of the six directions, that can be index using the potential path directions **j**. *n* is the number of wavelet levels used, for this application it was set to 4.

$$d(sc) = w_1\mathbf{A}(sc,\mathbf{y}) + w_2\left|\mathbf{y}''\right| + w_3\left(1 - \frac{\mathbf{ew_1} \cdot \sqrt[n]{\mathbf{IW}(yc,xc,\mathbf{j}_1)} + \mathbf{ew_2} \cdot \sqrt[n]{\mathbf{IW}(yc,xc,\mathbf{j_2})}}{\sum \sqrt[n]{\mathbf{IW}(yc,xc,f_i)}}\right)$$

$$xc = \left\lfloor \frac{sc-1}{M} \right\rfloor$$

$$yc \equiv sc\text{-}1 \pmod{M}$$

$$\mathbf{y} = \mathbf{A}(sc,i) > 0$$

$$\mathbf{y}'' = (\mathbf{y} - yc - (yc - yp)) = (\mathbf{y} - 2 * yc + yp)$$

$$\overline{\mathbf{y}'} = \frac{\mathbf{y} - yc + (yc - yp)}{2} = \frac{\mathbf{y} - yp}{2}$$

$$\mathbf{ew_1} = \frac{\mathbf{j_2} - \mathbf{j}}{\mathbf{j_2} - \mathbf{j_1}}$$

$$\mathbf{ew_2} = \mathbf{1} - \mathbf{ew_1} \qquad (14)$$

The weights *w* of the distance measure's three terms have been set heuristically to 1, 0.08 and 0.2 respectively; the algorithm is robust enough that the exact weight values are not very important. For some other application another values might be more appropriate. Having the distance measure defined between each pair of nodes in the input image, the task of finding the optimal path is simply that of finding a minimum sequence of distances from end to end.

After the initial estimate of RPE boundary has been found (Fig. 39 a)), a convex hull is fitted to the points that comprise the RPE boundary (Fig. 39 b)). A formal definition of the convex hull for a set of points X is the minimal area convex set containing X.

Fitting the convex hull allows for an efficient estimation of which regions should be interpolated over. The convex edges are always left intact, while concave regions (intuitively dents) are considered for interpolation. Following the efficient representation of the problem allowed for a simple statistical measure to be developed. In general, any segment which has a high boundary curvature, or is significantly shifted from the estimate of the convex shape is likely to be caused by pathological detachment of the RPE (i.e. drusen). Thus, an assumption can be made that the position of the RPE does not mark the start of the choroid, while in most cases a cubic interpolation gives a good estimate of the choroid. While the RBC estimate produced by the interpolation over these non convex regions is quite close to the boundary determined by experts, it is not always completely correct. For the simplicity and rarity of these occurrences, the algorithm was found to be performing sufficiently well even with simple interpolation. However, a more complex approach can be devised. After finding the cubic spline estimate of the start of the choroid region, a neighbourhood of the spline can be considered for a second run of the path finding algorithm. This will remove the area where the detached RPE is located, allowing for a more precise boundary estimate. That boundary is often very weak, and in some cases it is even marked by a black edge rather than a white one. Thus, for that run of the path finding algorithm, a more complex edge detector would have to be considered, instead of simply using intensity values. A straight forward approach would be to use a trained texture descriptor to obtain edge probabilities. If encoded as an intensity image, the same shortest path procedure could be subsequently applied.

Once the RBC boundary has been found, the neighbourhood and the image areas below it are removed (current position-50 pixels, to the bottom of the image). In the next step this image is used to find the real ILM position. Columns in which no edge was found signify gaps in the ILM, as the removal of the image signal around and below the RPE will result in no edge detection in the whole column (ILM and RPE being the brightest boundaries). The final ILM boundary is then found using piecewise cubic spline interpolation.

At the non-signal areas of the image, RBC is set as straight, while ILM is set to the value of the RPE (Fig. 40). Additionally, the program returns the non-signal area positions, so that the statistical model can use it later to find the choroidal-scleral interface, as at the non-signal areas of the image that boundary is set to the RBC value as well and is excluded from the optimization run (thus noise at the sides does not confuse the optimization function).

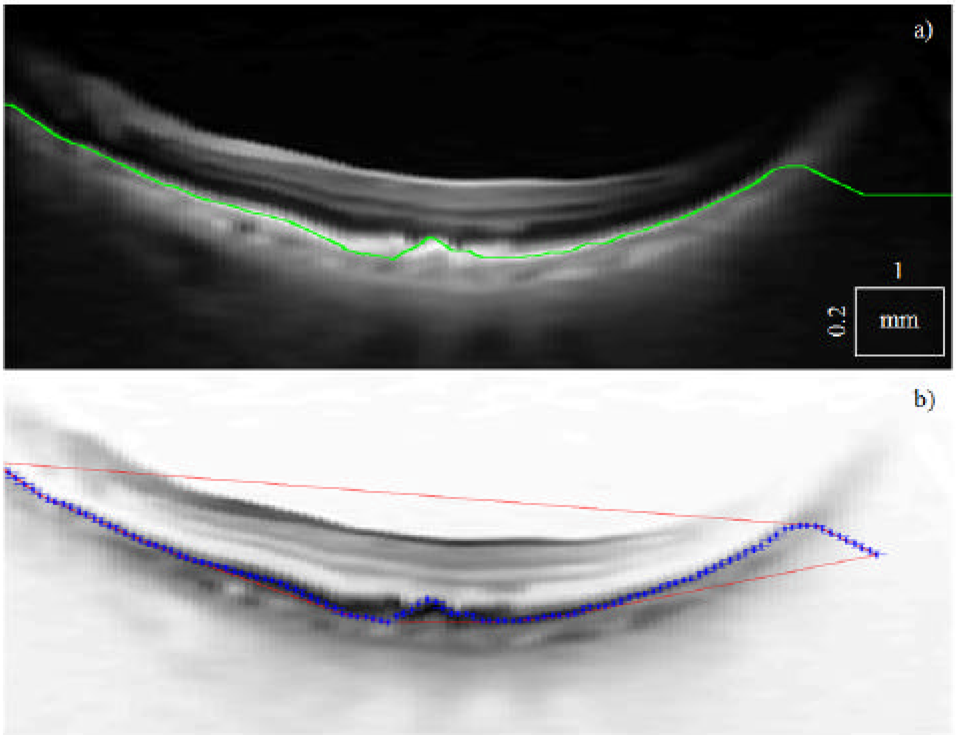


Fig. 39 RPE estimate (green line and blue dots) and convex hull (red line)

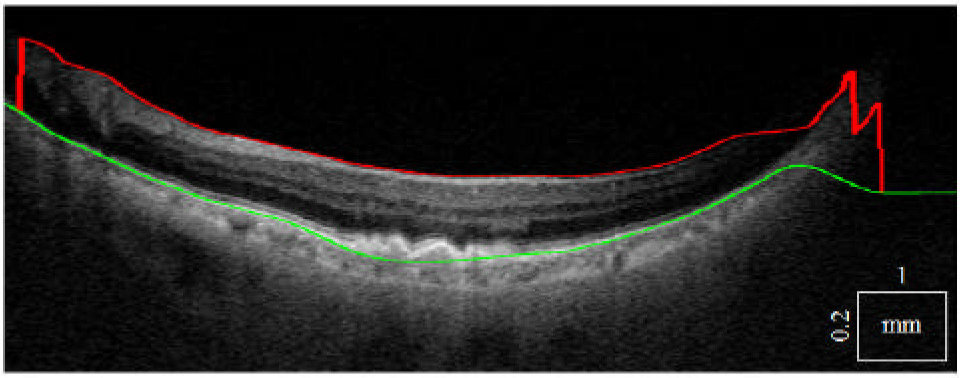


Fig. 40 The final result: ILM (red) and RBC (green)

2.3.2 Model Based Approach to Choroid Segmentation

Choroidal segmentation is an ideal application for the proposed statistical model, as the choroid is notoriously difficult to detect using standard approaches that have been applied so far to the foveal segmentation. There are three main reasons for that:

- the choroid is usually deep underneath the retina and thus OCT signals can degrade significantly, especially if the choroid is thick due to multiple scattering
- the CSI, its lower boundary is, unlike that of foveal layers, often very weak or in some locations invisible
- the texture is inconsistent comprising of large and small vessels

Essentially the choroid is more defined by its various vessel features as a textured region, rather than an area clearly separated by a strong boundary. All these factors make it an ideal application for a statistical model. The statistical model already assumes certain shape information and is thus far likely to perform poorly due to, at segments missing boundary and weak signal. Secondly, it considers the area bounded by the lower boundary being optimized, unlike many other segmentation methods that only try to track or in other ways segment the boundary itself without any knowledge about the area that that boundary defines.

The statistical model relies on initial boundary algorithm that defines the ILM and RBC and provides the offsets for the non-signal area of an image. The task for the model is then to find CSI boundary (the layer between the RBC and CSI). ILM is not used by the model, but is clinically useful to compare produced choroid thickness maps to the foveal shape.

The pre-processing and statistical model construction were done similarly to the procedure described in (Kajic et al., 2010), thus only algorithm elements that differ will be described in more detail. There are two distinctions. The first one is that the stacks are not registered by finding the position of the foveal pit, as it would be too unreliable to determine the exact location utilizing the same method as for the normal retinal segmentation. Dealing with pathologies means that there is no foveal pit, that part can be a bulge (i.e. AMD). Thus the stacks are simply used as they are, which does not cause any problems since the input to the shape training model is simpler, consisting of only one boundary (RBC), unlike with the foveal segmentation which had many more. The second difference is that the texture information is not included in the statistical model; texture is however used in form of a blob detector, but how that information is used by the objective function is fixed.

Given the RBC position the model is fitted to the image area below the RBC. A simple objective function is given: it tries to optimize the ratio of area covered by blobs (choroid being an area containing many vessels) versus the total area. Blobs are computed using a maximally stable extremal regions method (MSER (Matas et al., 2004)). An extremal region Q is a region for which all the elements are either larger (maximum intensity region) or smaller (minimum intensity region) than the neighbouring elements (pixels).

The concept more simply can be explained by thresholding. All the pixels below a given threshold can be considered black and all those below or equal white. In a progressive sequence of thresholded images the first one would be completely white, then black spots corresponding to local intensity minima would appear, then grow larger and eventually merge, until the whole image would be black. The set of all connected components in the sequence is the set of all extremal regions. A formal definition of maximally stable extremal region is given below:

Let Q1,..,Qi - 1,Qi,... be a sequence of nested extremal regions $Q_i \subset Q_{i+1}$. Extremal region Qi is maximally stable if and only if $q(i) = |Q_{i+\Delta} \setminus Q_{i-\Delta}| / |Q_i|$ has a local minimum at i. Δ is a parameter of the method. Intuitively it can be understood as a region that stays stable through a relatively large number of thresholds, i.e. it is a blob with a strong boundary, making it suitable for vessel cross-section detection.

Over a large range of thresholds, the local binarization is stable in certain regions, and has the properties listed below:

- Invariance to affine transformation of image intensities
- Covariance to adjacency preserving (continuous) transformation on the image domain
- Stability, as only regions whose support is nearly the same over a range of thresholds are selected.

- Multi-scale detection without any smoothing involved, both fine and large structure is detected

For this application the last two properties are important. The MSER implementation used is part of VLFeat, cross-platform open source collection of vision algorithms (Vedaldi and Fulkerson, 2008). It bundles a MATLAB toolbox, a portable C library and a number of command line utilities. Several parameter that the MSER function takes will be explained:

- Delta: already mentioned in the MSER definition
- MaxVariation: defines the maximum variation (absolute stability score) of the regions
- MinDiversity: the minimum diversity of the region. When the relative area variation of two nested regions is below this threshold, then only the most stable one is selected
- MinArea and MaxArea: these parameters define the minimum and maximum allowed areas of detected regions as percentage of the total area of an image

For this application the parameters were heuristically chosen as follows: 1, 0.25, 0.7, 10/imageArea, 500/imageArea. Given the RBC boundary and a MSER filtered image $F_{MSER}(\mathbf{I})$, the optimizer, using the objective function, looks for the minimal value by changing the shape of the lower boundary. The objective function is defined as:

$$f(\mathbf{s}) = -\frac{G_{CH}(S(\mathbf{s}), F_{MSER}(\mathbf{I}))}{\mathbf{A}_{CH} + \mathbf{A}_{PCH}} \tag{15}$$

It maximizes the ratio of the choroidal area covered in blobs against the total area of the choroid (A_{CH}) and the post choroidal region (A_{PCH}), taken to be an area from the choroid boundary plus 50 pixels, thus it is a constant. For the model initialization the mean of the model was used, which, having a robust objective function, proved to be sufficient. The A-scan optimization seeks black to white transition, given the fact that the choroid is usually darker than the post-choroidal region (though this boundary is often very weak or non-existent):

$$f_A(\mathbf{s}_A) = G_{DW}(S(\mathbf{s}_A),\mathbf{I}) - G_{AW}(S(\mathbf{s}_A),\mathbf{I}) \quad (16)$$

G_{DW} gives the average intensity of the choroidal region, CSI boundary minus 50 pixels, while G_{AW} of the post choroidal region, CSI plus 50 pixels. The final boundary is then filtered using a 1D median filter of size 11 pixels, with the size set heuristically.

2.3.2.1 Results

For evaluation, automatic segmentation results were compared to manual segmentation performed by an experienced examiner. A basic error measure was used, computed for a boundary *i* separately (in this case there is only one layer and two boundaries: RBC and CSI) and from these, error measures were computed for an entire B-scan or for an individual layer, Eq. (9). E_B (Basic) is the basic error measure that defines the number of misclassified pixels.

$$^{i}E_B = \sum_{j=1}^{j=w} \left| yAut_{ij} - yRef_{ij} \right| \quad (17)$$

To compute error for layer *k* separately, only the two boundaries that define a layer are added (offsets) and divided by the sum of the layer area as given by the automatic segmentation (A_A) and the layer area as given by the reference segmentation (A_R), Eq. (12). This is used, instead of summing up across all boundary errors, to normalize for double counting of misclassified pixels, as each layer is bounded by two boundaries.

$$E = \frac{\sum_{i=k}^{i=k+1} {}^{i}E}{A_A + A_R}, 0 < k < b \quad (18)$$

For clinical application, it would be useful for the operator to decide whether the obtained results are reliable. As part of future work, a confidence measure could be introduced, based on the values returned by the objective function after the optimization step. Large values are proportional to the low confidence in the boundary positions determined by the model fitting.

In Table 4 results are presented for the 7 diabetes type 1 eyes. In general, diabetes type 1 choroid is close to the choroidal shape and structure of healthy eyes. Error rates are even lower than what was obtained with retinal segmentation (Kajic et al., 2010). In Fig. 41 a typical B-scan is shown from a diabetes type 1 patient, before and after the A-scan optimization. Notice the good performance even in the central region where the signal is weak.

Table 4 Error values on 7 eyes with diabetes type 1, expressed as percentage of misclassified pixel relative to the layer area. The total average value is computed as an average of all the patient error values, and is independent of the number of B-scans per stack

Error Type	A (100 B-scans)	B (100 B-scans)	C (100 B-scans)	D (98 B-scans)	E (100 B-scans)	F (93 B-scans)	G (100 B-scans)	Total
E_B	14.1	9.2	19.4	15.8	13.4	15.7	16.1	14.8

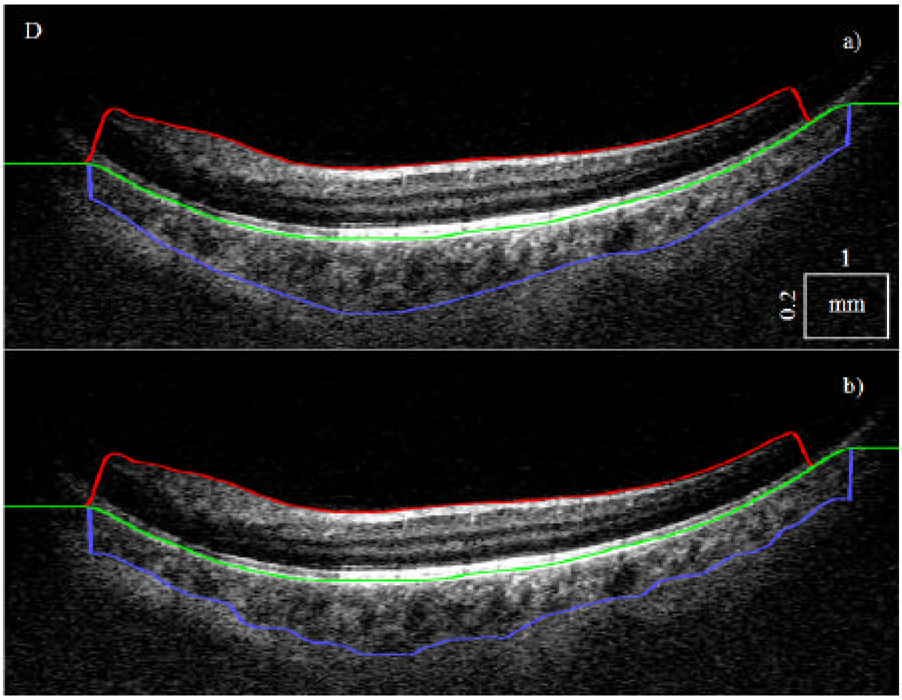


Fig. 41 A slice from a typical diabetes type 1 eye (patient D), before (a)) and after (b)) the A-scan optimization. ILM (red), RBC (green), CSI (purple)

In Table 5, 5 different pathologies are evaluated. The results are, surprisingly, even better than, close to normal, diabetes type 1 eyes (Table 4), especially considering the difficulty of segmenting such varied cases. One likely reason for that is the fact that the five pathologies were pre-processed slightly differently, with an average filtering of neighbouring B-scans performed after the registration (to remove speckle). Another reason is that the manual segmentation was performed in the B-scan range of 20 to 100 (every fourth, 20 B-scans in total), and not 15 to 114 (diabetes type 1 patients, 100 B-scans in total), thus the error rates that are larger in slices (B-scans) close to the start and end of stack, where the signal is lower, contributed to the larger error value of diabetes type 1 data segmentations.

Table 5 Error values on 5 eyes with various pathologies, about 20 slices manually segmented per stack, expressed as percentage of misclassified pixel relative to the layer area. The total average value is computed as an average of all the patient error values, and is independent of the number of B-scans per stack

Error Type	H (20 B-scans)	I (20 B-scans)	J (100 B-scans)	K (20 B-scans)	L (20 B-scans)	Total
E_B	5.6	12.8	10.1	13.0	15.3	11.3

Case H is an eye with a central serous choriorethinopathy. A lesion is present that affects the outer retina above the RBC complex. A deep, difficult to delineate, choroid is another characteristic of the pathology. This eye does not present a serious challenge for RBC segmentation (only slight RPE thinning in the middle), but determining CSI is difficult as the signal becomes very weak due to the thickness of the choroid. In

Fig. 42 a representative B-scan is shown after the final A-scan optimization.

Fig. 42 The algorithm performs well despite the thinning of the RPE in the middle and deep, low signal choroid

Case I is a terminal wet age related macular degeneration (AMD). The RBC complex has an irregular shape and thickness; however the dynamic programming approach that was introduced, performs well. In Fig. 43 a representative B-scan is shown after the final A-scan optimization.

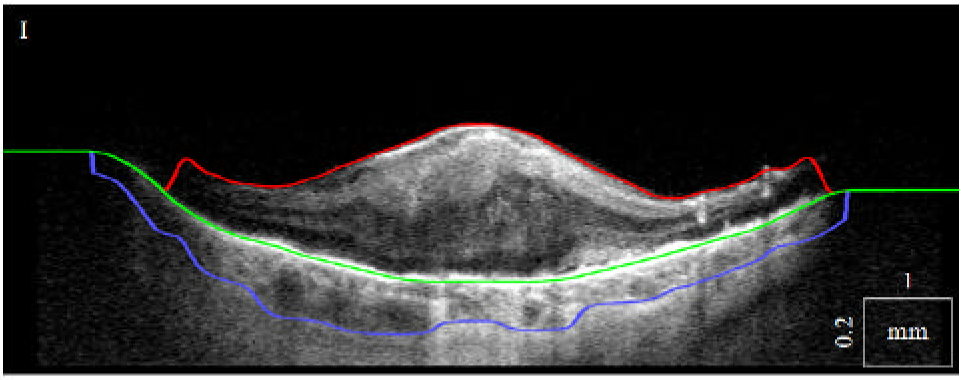


Fig. 43 The choroid is very thin and has a strong variation in contrast

Case J is an eye with RPE atrophy as a result of dry age related AMD with a strong signal in the choroid due to reduced melanin concentration in the RPE. The choroid is extremely thin. The bright area in the choroid presents a challenge for RBC detection as well as for the statistical model (determining CSI), as the choroidal structure is significantly different from a normal choroid. However, as can be observed, the model is general enough to perform well even in that case. In Fig. 44 a representative B-scan is shown after the final A-scan optimization.

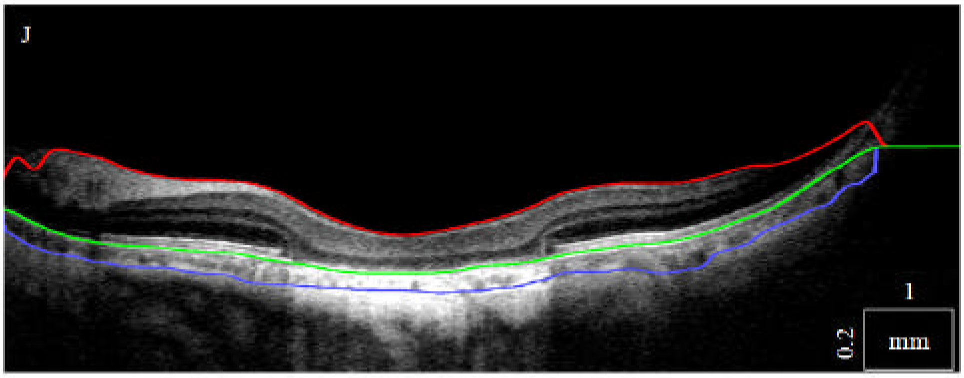


Fig. 44 RPE/RBC and CSI are found despite the extremely thin choroid with a strong signal in the middle as a result of the RPE atrophy

Case K shows an eye with deposits underneath the RPE (drusen). The RBC complex has an unusual inverted shape. A simple parametric approach for representing RBC, such as low order polynomial, would fail in this case. The proposed algorithm still produces reliable results. In Fig. 45 a representative B-scan is shown after the final A-scan optimization.

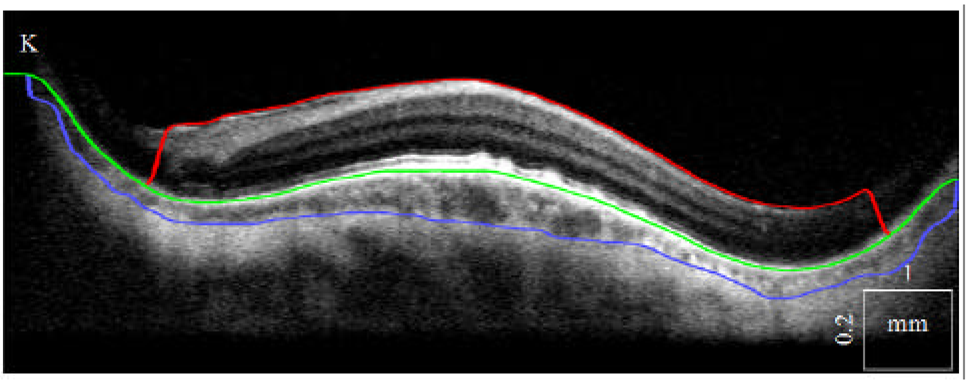


Fig. 45 RPE and RBC have an uncommon "inverted" shape

Case L is another AMD with a very difficult RBC to delineate. In Fig. 46 a representative B-scan is shown after the final A-scan optimization.

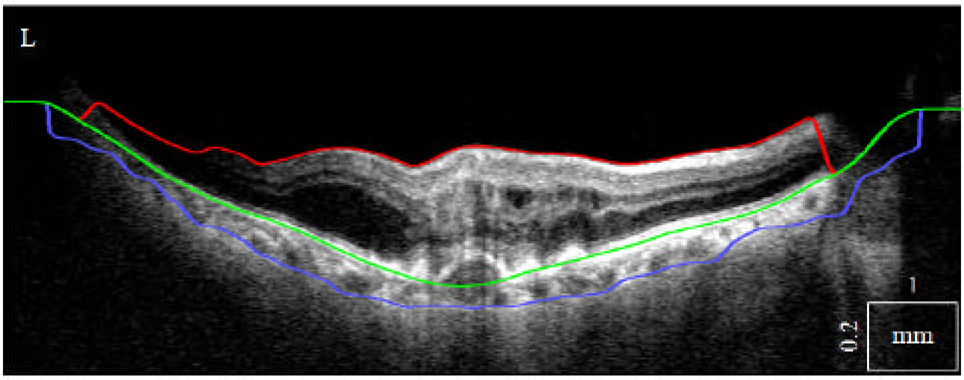


Fig. 46 The algorithm correctly interpolated over the drusen, finding a close approximation of the RBC

In (Kajic et al., 2010) we collected large, manually segmented datasets via Amazon Mechanical Turk (AMT) service. Since the segmentation of choroid, and especially pathologies, required too much expertise to use AMT, there is no inter-observer variability measure. By comparing the obtained error rates to the foveal inter-observer variability it became obvious that the error rates are within that range. The average of the total error values from Table 4 and Table 5 is 13%.

In Fig. 47 six thickness maps are presented. For brevity, only one diabetes type 1 eye is shown, as the other diabetes type 1 eyes are rather similar. However, all 5 pathologies have been included. Due to almost non-existing signal at the start and end of the stacks (volumes) sometimes a few of the first and last B-scans were left out.

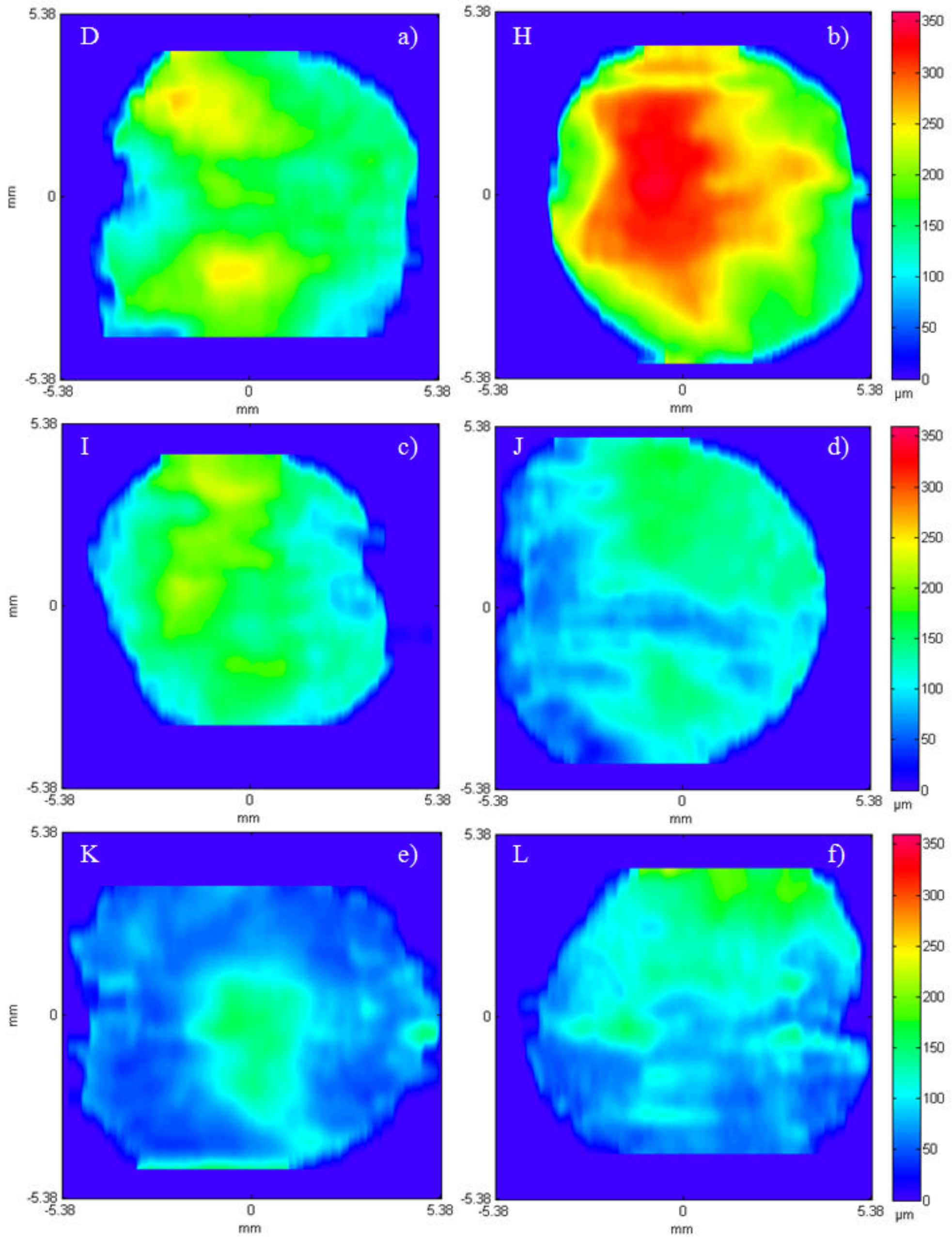


Fig. 47 Choroidal Thickness maps of a diabetes type 1 patient (D, subfigure a)) and 5 pathologies (H-L, subfigures b) to f))

The model based approach uses the variation obtained from the training set and imposes those constraints when segmenting an unseen image. This ensures that the segmentation will likely be close to the ground truth and less sensitive to noise. However, it is extremely important to have a large, representative training set that includes all possible variation. Overall the novel algorithm segments choroid accurately as compared to the human expert segmentation, with the error rate 13%. In case that a precise and robust stack registration algorithm is used in the pre-processing stage, the algorithm could be implemented in full 3D and would be even more robust and accurate.

The current processing time is about 30 seconds per slice, using non-optimized MATLAB implementation. For this application, close to real time processing is not essential as the results obtained are used for subsequent scientific analysis to obtain trends in various choroidal conditions. For clinical use, however, using optimized C implementation and a trivially parallelized approach, this could be brought down to approximately same time for the whole stack, using commodity hardware such as a GPU card.

2.3.2.2 Conclusion

An algorithm for automatic segmentation of the choroid based on a statistical model has been developed. Before applying the statistical model, the top choroidal boundary (RBC) has to be determined, as it used to initialize the model. A novel method was developed for that purpose according to high robustness requirements in order to work well in cases of severe pathologies and signal degradation. It is a shortest path dynamic programming method that was extended to use the path derivative information and edge orientation with the same time complexity. It is also applicable to a general class of computer vision problems.

The choroidal automatic segmentation algorithm has been thoroughly tested and evaluated against a large, manually segmented dataset obtained from a 1060nm OCT system and proved highly robust in wide-field scans even in the presence of artefacts and pathologies. It is the first time that a large, representative choroidal data set has been automatically segmented. The algorithm successfully demonstrated reliable performance under difficult conditions. Clinically, fully automated segmentation of the choroid is essential in making medically useful the possibilities given by the method of high resolution, high speed OCT.

2.4 Conclusions

In this chapter various approaches to segmentation have been presented, some for completeness even though they have been surpassed by the more efficient ones. Thus it is important to emphasize which are the current state of the art both from medical and computer science perspectives.

Detecting top and bottom boundaries using level sets and adaptive thresholding approach worked sufficiently well for many cases, however the method developed for choroidal top and bottom boundary detection outperforms both, while its potential use in a broad range of applications outside the OCT makes it a significant contribution from the computer science point of view as well. It is based on dynamic programming shortest path algorithm that incorporates path derivative and edge orientation criteria (based on directional wavelet analysis), implemented efficiently to preserve the same time complexity O(E) as the standard shortest path approach. The fundamental difference between this approach and others is that the algorithm works with limited path history, instead of only using static weights per edge to determine the distance function. Thus, there is no need for complex and slow workarounds, as when trying to incorporate path information with static weights only.

For the segmentation of inner retinal layers, simpler methods using pure texture classification and adaptive template matching have been proposed, which led subsequently to the development of a unified statistical model that utilizes both shape and texture. From the computer science perspective, two important general novelties with respect to the standard Active Appearance Model (AAM) were introduced: a second term in the optimization function that penalizes large deviations from the three boundaries found by the adaptive thresholding algorithm and the second algorithm stage that refines the model fit for each A-scan independently, giving increased accuracy. Thus, it can be concluded that the algorithm successfully demonstrated reliable performance under conditions which prove extremely challenging for the pre-existing methods. It was used as a basis for the development of the model based choroid segmentation algorithm, proving that indeed it is well suited to the weak and missing data conditions.

3 Pre-Apoptotic Monitoring

Here, a novel ligand-free method for assessment of neuronal viability *in vitro and ex vivo* is described. Using ultrahigh resolution optical coherence tomography (UHR-OCT) to detect changes in the light scattering properties of early stage apoptotic cells, it is shown that an increase in proportion of apoptotic cells *in vitro* can be accurately quantified. Moreover, the data indicates a similar increase in neuronal scatter in retinal explants following axotomy, suggesting that UHR-OCT can be a novel non-invasive technique for the *in vivo* assessment of neuronal health.

Texture inside the pre-apoptotic regions was analyzed by projecting the data into the multidimensional feature space and performing training and analysis in the feature space using Gaussian Mixture Models (GMMs). Analysis was performed on cell cultures grown on glass coverslip (in vitro) by training and distinguishing two groups: healthy and apoptotic. Following the extremely encouraging results (clear separation of the two groups with the distance between them far exceeding the confidence interval) of this proof of principle experiment, the second experiment focused on more demanding pre apoptotic analysis of the retinal explants.

3.1 OCT system

Imaging was performed using a spectrometer-based frequency-domain OCT using a commercially available Ti:Sapphire laser light source (Integral OCT, Femtolasers GmbH, Austria) with a bandwidth of 140 nm a schematic is shown Figure 48 A scanning microscope was used as the sample probe with the OCT scan lens (10x and 5x) (Thorlabs, Newton, New Jersey, USA) with 18 mm and 36 mm working distance, giving an axial resolution of ~4 µm and a transversal resolution of ~6 µm and ~8 µm respectively. The light source was connected to a 90/10 beam splitter (Ipitek Inc, Carlsbad, CA, USA) after which the bandwidth of the interferometer remained at 130 nm. The scanning microscope was connected to the 10% arm of the beam splitter and samples to be imaged were exposed to ~1.4 mW of power only during requisition. Images were recorded using a 12 bit 2048 pixel silicon CCD-camera (AVIVA M2 CL2014-BAO, Atmel, CA, USA) at a line rate of 20 kHz with 50 ms exposure time. Sample and reference arm fibres (SM 650, Fibrecore Ltd, UK) were looped through polarization control paddles to adjust for maximal fringe visibility. A gold coated hollow corner cube reflector mounted on a translation-stage was used in the reference arm and a reflective neutral-density filter-wheel allowed adjustment of the power such that the intensity on the camera was close below the saturation point.

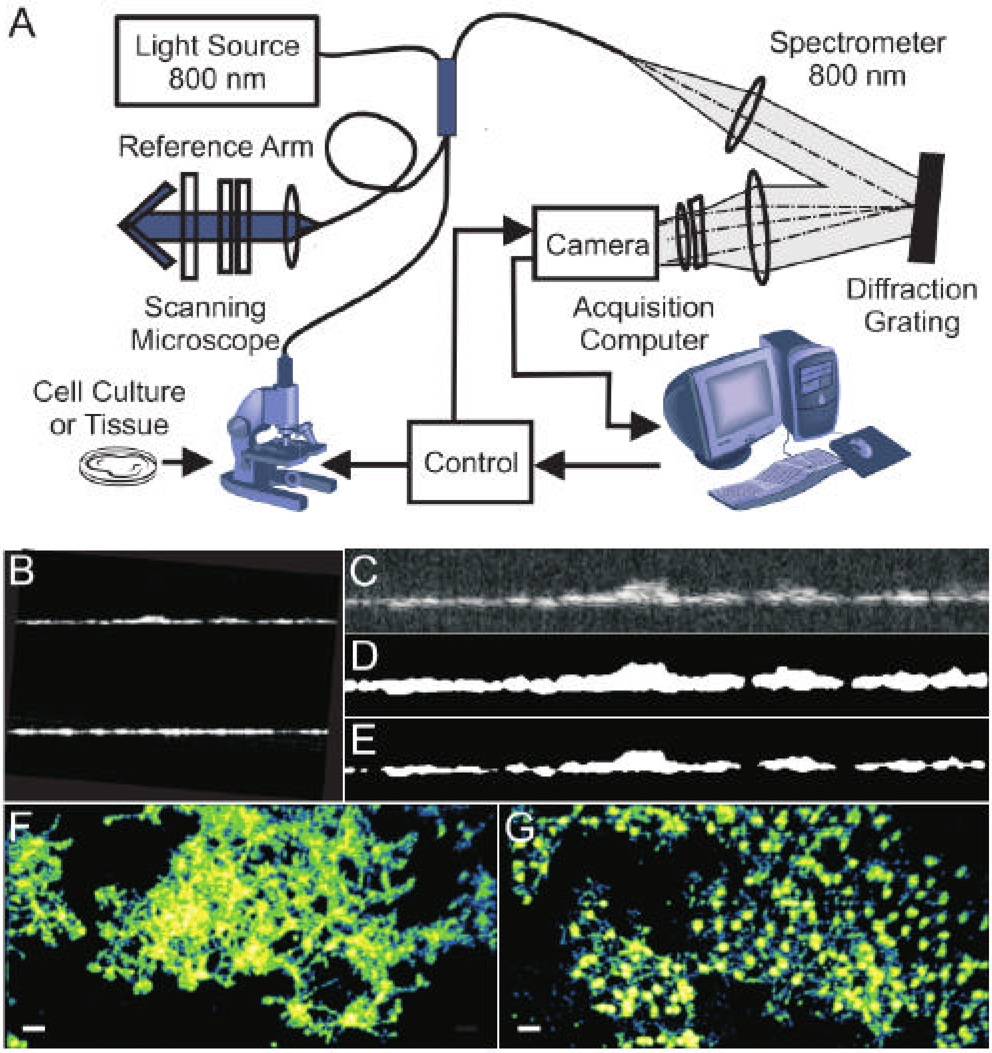


Figure 48 Cell imaging setup and pre processing. **In A the OCT system is shown. B and C (zoomed it) shows the input, the initial mask is shown in D, while the final mask is obtained after morphological operations, E. An example of control cells is shown in F, 24h after the treatment, they contract, G.**

Three different experimental sets of OCT images have been collected and analysed. The first experiment aimed to detect a difference between healthy (control) and apoptotic cell cultures 24 hours after treatment with staurosporine. The second experiment focused on identifying changes during the early stages of apoptosis *in vitro* where each cell culture was imaged at 6 different time points (10 to 60 mins) following treatment with staurosporine. Finally, in the third experiment, *ex vivo* explants where imaged during the first hour of the apoptotic process induced by axotomy.

3.1.1 Identifying the optical signature of apoptotic RGC-5 cells 24h after treatment with staurosporine.

For the first experiment, data were collected from 20 apoptotic and 20 control RGC-5 coverslip cultures whereby each coverslip was imaged at 10 different locations. The quality of all images was then visually inspected and some images were discarded based on technical imperfections. The remaining 120 control and 120 apoptotic 3D images (also referred to as images or image stacks) were selected for analyses. After initial image pre-processing, a fully automated approach was used to locate the cells within the image (the region of interest, (ROI)).

A set of parameters/features that described each ROI in compact form was then defined. In general, the choice of parameters is subjective. In this study, 65 interdependent texture features were defined based on 3D distribution of pixel intensity, using either individually or in combination averages, granulometry measures, and spatial wavelet analyses (as detailed in 8.2.3). A standard feature analysis technique was applied to distinguish between images of healthy versus apoptotic cells. To identify the generalised distribution of features for each type of data (data class) a GMM was used that also allowed for subgroups (clusters) within each data class. All 65 features were used for the analysis and up to four clusters were allowed for each class. As the basic learning process the exact number of required clusters was then established during the fitting procedure based on Akaike Information Criteria (AIC), that estimates the trade off between accuracy and complexity of the model. Once all the parameters of the model were determined, during the validation stage, a new image-stack was introduced and analysed as a new data point in the multidimensional feature space. The Mahalanobis distance was computed between the new point and the respective centre of each of the earlier defined clusters. The new point was then classified, e.g. assigned to a class separated by the smallest distance.

The cross-validation (“leave-one-out”) technique was used to evaluate the statistical accuracy of the classification. Data were divided into subsets, containing 10 to 20 image stacks according to the imaging session (to minimise variability due to the manual adjustment of parameters such as light polarisation during acquisition). During each step of the procedure, one subset was used for validation (each of the images in the subset were classified) while the rest of the data was used for finding the model parameters. The procedure was repeated for each of the subsets and the classification rate defined as a ratio of correct classifications over the total number of classifications.

3.1.2 Identifying the optical signature of early apoptosis in RGC-5 cells

The second experiment was set up to determine whether it is possible to identify optical differences in RGC-5 cells at early stages of apoptosis. For this analysis, data was collected from 34 separate RGC-5 cultures undergoing apoptosis, restricting the analyses to the first hour after application of staurosporine. Some of the images were discarded after initial visual inspection due to poor quality and the feature analysis was performed on 28 cultures, imaged every 10 min from 10 to 60 min post staurosporine treatment (168 image stacks in total, divided into 6 subsets according to the imaging session).

3.1.3 Identifying the optical signature of early apoptosis in explants

The final experiment was designed to identify the early signs of apoptosis in *ex vivo* retinal explants during the first hour post axotomy. Major changes were anticipated in the Inner Plexiform Layer (IPL), where dendritic terminals of RGCs are located, this was based on observations that RGC dendrites undergo a significant amount of remodelling and pruning prior to shrinkage and loss of the cell body (Weber, Kaufman and Hubbard, 1998).

Calcein/PI staining was carried out to identify the presence of live/dead cells within the murine retinal explants.

ROI was identified by a human expert, manually masking several regions within the retina, using a simple graphic interface. The borders of each ROI were then automatically adjusted (eroded) as stated in the methods section. 13 retinas were analysed and 10 ROI sized 128 x 64 x 64 pixels within the IPL of each retina with a depth of 64 slices were collected. This produced 130 image-stacks for each time point that were further divided into 5 subsets according to the timing of image acquisition.

Since feature analysis reliably discriminates healthy and apoptotic RGC5 cells, and that subtle early changes can be identified, the same procedure was then applied to the ROI within intact retinal explants undergoing apoptosis following axotomy (Quigley et al., 1995) .

3.2 OCT image analyses

OCT images were acquired as a stack of 512 slices sized 512 by 1024 pixels (1492 µm by 746µm). Prior to detailed analyses, all OCT images were subjected to standard image processing techniques. In particular, background noise was subtracted from the image; the position of the coverslip holding cell culture was detected by examining the direction of high frequency component in two dimensional frequency space, and all possible tilts were removed by rotating the image; the line scan artefacts were removed using FFT filtering. All image processing was performed in Matlab (Mathworks) using standard functions from image analysis and statistics toolbox.

3.2.1 Masking the Region Of Interest (ROI) for in vitro cell culture

For each 3D image, the intensity values of pixels were plotted as a histogram. The derivative of this distribution determined a heuristic threshold for the selection of pixel intensity values. This approach effectively removed the majority of pixels with a low signal to noise ratio attributed to background noise. All input pixels (Figure 48 B and C) above the threshold were classified as within ROI (Figure 48 D). This approach, however, did not guarantee continuity of ROI, because regions of low signal to noise intensity could also occur within ROI boundaries, producing “holes”. To resolve this issue, basic mathematical morphology techniques were used (first morphological dilation combined with mask erosion, then mask erosion alone; the sphere radius was chosen to be 5 pixels for texture filters and subsequently for morphological operators, that was the estimated bio-physical scale of interest, as it relates to the mitochondria size) to gradually adjust the ROI boundaries, eliminate “holes” and low intensity regions around the border, and produce a homogeneous

3D globule shaped ROI (Figure 48 E). The extension of erosion was set empirically to ensure that any background signals were removed. Finally, any brightness variations within the image (largely dependent on the depth of focus) were corrected using histogram equalization inside the ROI. After the background was removed, it was possible to plot a clear projection of control 0h (Figure 48 F) and apoptotic 24h cells (Figure 48 G).

3.2.2 Masking the Region of Interest for explants

In contrast to homogeneous cell cultures, such as RGC-5 cells, retinal explants are complex multilayer structures. Shadows produced by the adjacent layer and imaging artefacts make automatic ROI detection challenging. Thus, a semi-automatic approach was used with a human expert masking several ROI within the IPL, with a good SNR, using a custom designed graphical user interface (GUI). The borders of each ROI then were automatically adjusted (eroded) similarly as described above.

3.2.3 Feature selection

Each ROI was parameterized using 65 features. First, three standard image filters were applied to calculate the local entropy of the image, local range, and local standard deviation of pixel intensity. For each of these three matrices, four average parameters were calculated across the ROI: mean, median, entropy and standard deviation, hence producing the first 12 features. The feature values were further normalised according to the grey scale resolution of the original image. For example, for the 8 bit image, the entropy value was divided by 8 and standard deviation multiplied by 2.

Next, three co-occurrence matrices were defined over the image with an offset of 5 pixels along one of three (x, y, and z) coordinates respectively. The value of the offset was chosen similar to the size of a single cell. For each of the three matrices four average values were calculated across the ROI, such as contrast, correlation coefficient, energy and homogeneity. This operation provided us with the next 12 features.

Using the threshold that was previously established for the background signals while defining the ROI, the gray scale image was transformed into a binary black and white, and performed simple granulometry analyses taking into account only the size of continuous patches of image, referred to as particles. This histogram, specifying the number of particles of different sizes, provided us with further 12 features (particles up to size 12); the size of the histogram bin was increased (doubled), and another 6 features were extracted. By repeating the same procedure once again, 3 more features were extracted, totalling the number of granulometry features to 21.

To estimate the distribution of spatial frequencies in the image stack, directional wavelet analysis (Kingsbury, 1999) along three axes was performed: x, y, and 45 degree between x and y, concentrating on the three smallest scales of wavelets. For each scale and direction, the mean and standard deviation of coefficient values across the ROI were calculated, adding another 18 features to the feature space.

Finally, 3rd and 4th moments of pixel intensity distribution were calculated across the ROI adding skewness and kurtosis to the feature space.

The resulting feature space is thus comprised of 65 features with high levels of dependence between one another.

3.2.4 Feature Analyses

The core idea behind feature analysis is that objects within images can be easily recognized in a feature space with a smaller dimensionality than the original image, using a GMM to identify feature representation for each of the classes of data (apoptotic cells vs control). The GMM allows for the presence of subpopulations within a class, without requiring information about identity of these subpopulations. It allows up to four subpopulations in GMM (in our methodology) with the exact number being subject to the fitting procedure (training). Taking into account the normal distribution of the data, the best fit was determined in accordance with Akaike Information Criteria (AIC) using the Expectation Maximization (EM):

$$AIC = 2*k - 2*\ln(L) \tag{19}$$

where k is the number of subpopulations, and L is the likelihood function characterising the fit of the model with a given set of parameters to the training data.

After defining model parameters during the training stage, the validation stage consisted of presenting a “new” image to be classified using the model parameters. The image is defined as a set of features and its class identity is known but not revealed. To classify the image, the Mahalanobis distance is computed between a new point in the multidimensional feature space X (representing the new image) and the centre of a cluster Y, representing all the images in a given class:

$$d(\mathbf{x},\mathbf{y})=\sqrt{(\mathbf{x}-\mathbf{y})^T S^{-1}(\mathbf{x}-\mathbf{y})} \tag{20}$$

where S is the covariance matrix of the cluster.

The new image is assigned to a class separated by the minimal distance. The cross-validation technique (“leave-one-out”) was used to assess how accurately the predictive model differentiates between classes. On each step of cross-validation, all data excluding one of the data subsets were used for finding model parameters. The remaining subset was used for model validation, e.g. all images of this subset were classified (correctly or incorrectly) to one of the classes learned from the training set. Multiple rounds of cross-validation were performed, and the percentage of correctly classified images was estimated.

To visualise multidimensional data, Sammon projection along the x-axis with the largest variance was used. As Mahalanobis distance is only defined for two points within the same class/cluster, the distance between two points from two different clusters was calculated as follows:

$$d(\mathbf{x},\mathbf{y})=\frac{2}{\frac{1}{d_1}+\frac{1}{d_2}} \tag{21}$$

where d_1 is Mahalanobis distance computed using the covariance matrix of the first cluster and d_2 is Mahalanobis distance computed using the covariance matrix of the second cluster. Harmonic mean was used to mitigate the impact of large outliers. The Centre of Gravity (COG) plots were produced by finding the mean of each cluster; the confidence intervals were computed, assuming normal data distribution.

3.2.5 Feature Reduction

Feature reduction helps to further decrease dimensionality and speed up the learning process. The feature reduction was performed using sequential forward search and an additional cross-validation (“leave-one-out”) procedure over each training set. In practice, for each step of the analysis one data subset was chosen for validation (classification) and the rest of the data were used for finding model parameters (training) as before. However, now the training set was also subdivided, with one data subset left out for validation of feature selection and the rest of the data used to find model parameters as a function of smaller feature set. The procedure was repeated for each subset from the training set and the multiple sets of “optimal” features were produced. Due to the small number of subsets, “union” was performed on all the relevant features, producing a reduced feature set that was not necessary the optimal. It should be noted that classification was performed on the set that was different and completely independent from the training set and the feature selection, all the images were classified in this remaining data subset using reduced feature set. The whole procedure was repeated with different choices of validation data and classification rate was calculated as before.

Giving the binary nature of classification results, the classification error p can be described by the Bernoulli distribution. Bernoulli distribution belongs to a class of exponential distributions, thus the optimal number of features k can be estimated using Bayesian Information Criterion (BIC). This is an asymptotic result derived under the assumption that the data distribution is in the exponential family:

$$BIC = k * \ln(n) - 2 * \ln L \tag{22}$$

To use the measure in practice, it is necessary to estimate the likelihood L. While standard BIC applications usually assume normal distribution of variable errors/disturbances, in this case it is a Bernoulli distribution again, as the classification error p is either 0 or 1 but so are the error disturbances.

We derived the following measure:

$$BIC = n * \ln(p * (1 - p)) + k * \ln(n) \tag{23}$$

where *n* is the number of data sets. During the iterative procedure, all features could then be sorted by their classification contribution and the resulting set could be formed by adding classification features one by one, starting from the most relevant feature, while evaluating the BIC after each addition. The minimal value of the BIC corresponds to the most appropriate data set.

3.3 Results

3.3.1 Identifying the optical signature of apoptotic RGC-5 cells 24h after treatment with staurosporine.

In the first experiment, two distinct classes were identified representing healthy and apoptotic cells respectively. The classification rate was estimated at 95%. To visualise the data, low dimensional projection plots were produced using Sammon projection methods (Sammon, 1969) allowing the conversion of distances from multidimensional to two dimensional space. The resulting distance is dimensionless, and the scale is given in arbitrary units (Figure 49 a), b)). The results were correlated with the tunnel labelling (Figure 49 c)).

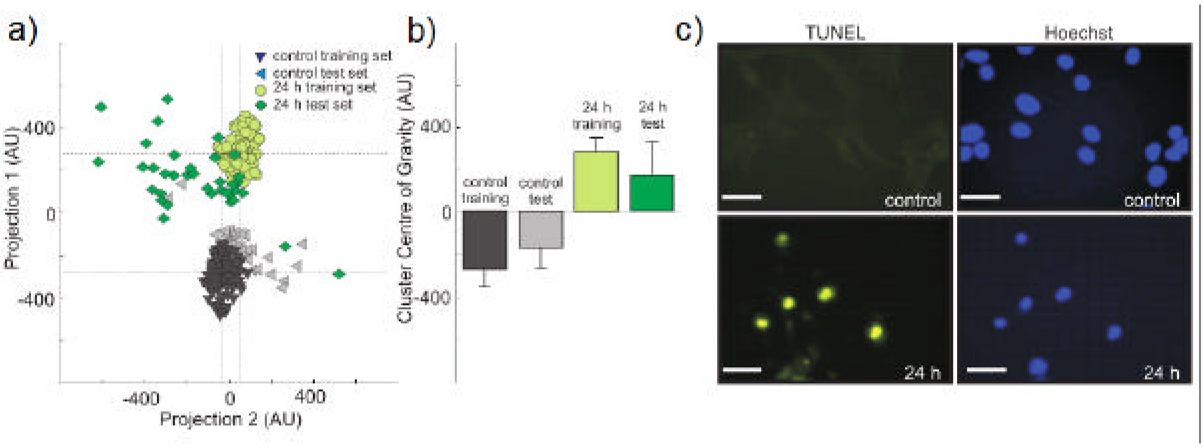


Figure 49 a) and b) show a clear separation of the control and 24h datasets, after the projection into a 2D feature space. Tunnel labelling in c) correlates with the analysis results.

3.3.2 Identifying the optical signature of early apoptosis in RGC-5 cells

The ROIs were automatically detected for all the images as described in the previous section. Due to the increased complexity of the classification task and the smaller data set, the number of independent features had to be reduced. By disregarding the most irrelevant and redundant features, dimensionality was reduced and the learning process was accelerated. Feature reduction was performed using a sequential forward search and an additional cross-validation (“leave-one-out”) procedure for each training set (see details in 8.2.5). Due to the relatively small number of subsets, “union” was performed on all the relevant features. The resulting set of features was then used for validation, which was completely independent in regard to both training and feature reduction. The “union” of the features, however, is not ideal, especially for a large data set. On average 17 best features for each of the validation procedures were used; however, a much smaller number of features could often suffice, although the specific set of highly discriminative features varies. The classifications during the validation stage were, hence, performed with a slightly different set of features for each data subset. As before, the classification was repeated multiple times in order to calculate the classification rate.

There is a progressive right shift in the centre of gravity (COG) of clusters representing cells in the feature space (Figure 50 a) and b)) from the control position towards the apoptotic position. The most dramatic shift occurred during the first 30 mins after the administration of staurosporine. The COG then remained in the same position between 30 and 40 min and shifted again between 40 and 50 min before reaching a plateau.

The classification rate in differentiating 10-20 min from 30-60 min data was 85%. These changes correlated with the MitoTracker Red mitochondrial labelling (Figure 50 c)), which showed a change in the mitochondria morphology post 20 mins of apoptosis induction.

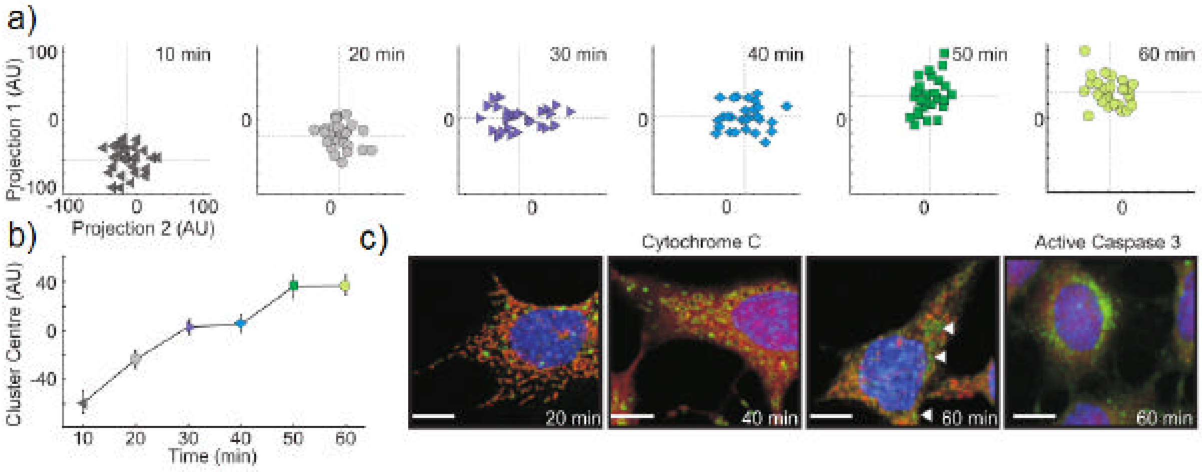


Figure 50 In a) and b), a progressive right shift in the centre of gravity (COG), of clusters representing cells in the feature space, can be observed, from the control position towards the apoptotic position. These changes correlated with the MitoTracker Red mitochondrial labelling, c), which showed a change in the mitochondria morphology post 20 mins of apoptosis induction.

3.3.3 Identifying the optical signature of early apoptosis in explants

Positive Calcein (green) staining with little or no negative PI (red) staining at 23 and 53 min post axotomy indicated that the retinal explants were alive with little or no dead cells within the RGC layer of the retina (Figure 51 a)). Haematoxylin and Eosin histology indicated that the retinal morphology during retinal explant culture did not change.

Similar to the analysis of RGC-5 cultures the centre of gravity clusters representing ROIs in *ex vivo* explants progressively shifted to the right over time (Figure 51 b)). The shift was, however, less pronounced than for the RGC-5 cultures, most likely due to the higher degree of variability of cells within the retinal tissues. Nevertheless, the difference between 15-23 min and 33-53 min post axotomy timepoints was detected in the areas of the IPL with the accuracy of 68%. Visual evaluation of the tissue confirmed that the retina remained transparent throughout the experiment, indicating that the changes observed were not secondary to opacification of the retinal tissue.

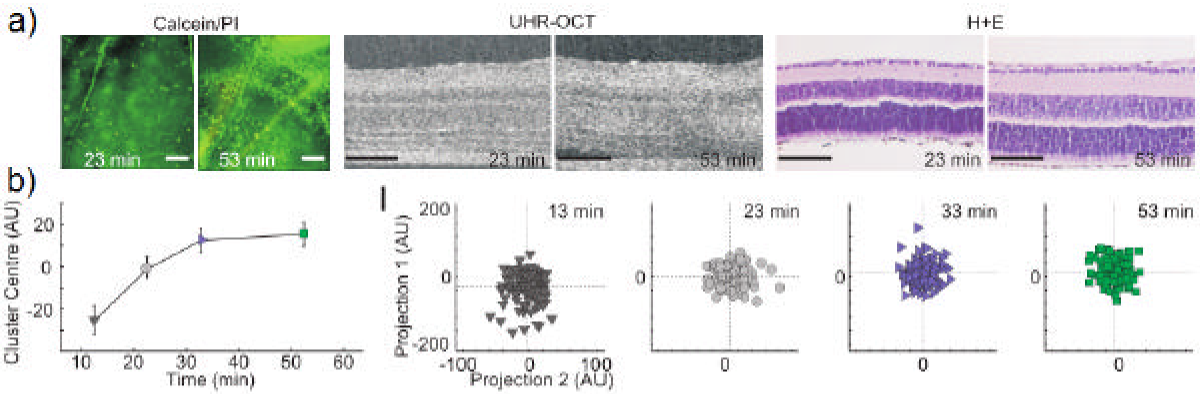


Figure 51 For the explant timeseries, a) shows Positive Calcein (green) staining with little or no negative PI (red) staining, at 23 and 53 min post axotomy, indicating that the retinal explants were alive with little or no dead cells within the RGC layer of the retina. Haematoxylin and Eosin histology indicated that the retinal morphology during retinal explant culture did not change. In b), a shift of COG can be observed, although not as prominent as in the case of the cell timeseries.

3.4 Conclusions

It was shown that texture analysis of retinal tissue imaged by UHR-OCT can identify optical changes associated with cellular events leading to apoptosis. Importantly, these changes were observed in both cell culture and explant systems. The results indicate that UHR-OCT can detect changes in the optical signatures of cells and explants very early after the insult to initiate cell death. In RGC-5, *in vitro* cultures changes were seen as early as 10 mins and in retinal *ex vivo* explants as early as 20 mins, suggesting that UHR-OCT has the potential for *in vivo* diagnostics.

The UHR-OCT derived changes closely match observations from immuno-histochemical analysis reporting morphological changes in the mitochondrial morphology. Whilst it could not be conclusively verified that the optical changes detected, solely derived from mitochondrial morphological changes, recent studies, using a Fourier processing microscope setup, also concluded that the mitochondria generate optical scatter changes during apoptosis (Pasternack, Zheng and Boustany, 2010). The value of this imaging technique in scoring neuronal viability/health is not contingent on knowing which organelles are generating the optic signature.

Binary discrimination of viable and dead cells was evaluated by histochemical staining. Whilst this demonstrated a high degree of accuracy, the conclusion was that the greater value of this imaging technique is in the derivation of quantitative indices of neuronal health. The shift in the centre of gravity of a duster representing cells in feature space correlated well with the anticipated time course of degeneration within treated cell cultures and axotomised retinal explants. It is possible to overlay these indices on the UHR-OCT images to provide tissue related maps of cellular health.

It was shown that this methodology can be translated from *in vitro* to *ex vivo* retinal explants where early apoptotic changes were detected within the IPL of the retina. The UHR-OCT system is able to image the full depth of the retinal explant and all layers of the retina can be identified, allowing for a pan-retinal analysis of apoptosis. The high axial resolution of OCT is dependent on the spectral bandwidth of the light source and is, effectively, not aperture limited, meaning it does not depend on the objective used in the OCT setup.

This study has demonstrated that UHR-OCT can be used to provide an index of neuronal viability *in vitro* and *ex vivo* and has several advantages over previous apoptosis detection methods, namely it is a non-toxic, ligand-free, non-invasive technique, that can be used in live cell imaging. UHR-OCT is often used in clinical practice to image tissues, therefore it is highly probable that with the development of detection technology, this methodology could be applied to *in vivo* studies. This is a major step forward towards identifying timely therapeutic interventions to reverse or prevent cell loss.

4 Summary

Overall, during the course of the PhD numerous results were obtained that enabled more efficient application of OCT in the existing applications (retinal imaging) and applications to novel fields (pre-apoptotic monitoring). Work on facial dynamics and behaviour transfer was a separate Computer Science department project that produced interesting results on behavioural transfer and general signal processing. Although application wise the main PhD focus and the facial dynamic work did not have any common areas, the methods used were often linked (e.g. machine learning in general). Importantly, some of the developed methods have general use outside the OCT field as well.

The registration results were useful, however more work is possible to improve the elastic (non-affine) registration results and ideally use the independent fundus image in the process as a reference which is free of motion artefacts compared to a fundus image generated from 3D OCT data. Level set method was tried for retinal boundary detection but proved to be too unreliable and slow. The fast adaptive thresholding method was thus developed that is fast and effective. Segmentation results using texture-based segmentation are reasonably accurate when segmenting data that is close to the training data, but behave poorly on the more varied data. The reason for this is the use of the geometrical features in the process that enforces geometry of the training set on target data. If these features are left out, then segmentation results are patchy and geometrically inconsistent, with various layers detected at large offsets from proper locations. Adaptive template matching results were more accurate than the texture based results but are also susceptible to large failures since all the A-scans are segmented independently.

A statistical model based approach was developed that unifies the texture and geometric based approaches under statistical constraints posed and extracted from training data. It performs very well, it was tested on a large independent data set and the results were published in an international journal (Kajic et al., 2010). Two novel methods were added to the standard model approach: the shape prior information extracted by thresholding that increases the overall accuracy (7.2.4.3), and the second 1D refinement step based on the back projection of the training data through the 2D model (7.2.4.3). Both methods can be used in general problems outside the OCT field as well. A similar statistical model was constructed to segment the choroid. It incorporates maximum stable extremal region (MSER) blob detector as the key texture feature.

The top and bottom fovea boundary (ILM and RPE) novel method was developed according to high robustness requirements in order to work well in cases of severe pathologies and signal degradation. It is a shortest path dynamic programming method extended to use the path derivative information and edge orientation with the same time complexity. It is also applicable to a general class of computer vision problems.

Texture analysis was successfully applied to pre-apoptotic monitoring to in vitro and ex vivo (retinal explant) data, justifying future work on in vivo analysis that would allow early diagnosis of some major neurodegenerative conditions. Effective texture features were constructed (e.g. point of view rotationally invariant directional 3D wavelet coefficients) and optimal number of feature criteria derived from Bayes information criteria (BIC) proposed.

The work on facial dynamics accomplished two goals. Using facial curvature information, it was possible to register 3D faces using facial key points (eye pits, nose etc.). The work on visual sequences of various subjects uttering the same sentences led to important results in trustworthiness analysis and additionally produced important improvements to dynamic time warping (DTW) general signal processing algorithm (real dynamic derivative computation and graph based DTW).

The results obtained were presented at various international and local conferences:

- texture based segmentation: oral presentation at SPIE 2008, San Jose, CA, USA
- adaptive template fitting poster presentations at ISIE and ARVO 2008, both Fort Lauderdale, FL, USA
- model based segmentation oral presentation at ECBO 2009, Munich, Germany
- pre-apoptotic monitoring oral presentation at SPIE 2010 San Jose, CA, USA
- facial dynamics poster presentation at Face and Gesture (FG) 2011, Santa Barbara, CA, USA
- various UK based conferences (BMVC London, Aberystwyth, Cardiff etc.).

and were published in international journals and peer reviewed conference proceedings:

- registration work (Povazay et al., 2009)
- model based foveal segmentation (Kajic et al., 2010)
- cell imaging apoptosis related work (Hofer et al., 2011b) and (Hofer et al., 2011a). Also a submission is underway for Nature Methods
- choroid segmentation related (Esmaeelpour et al., 2010) and (Esmaeelpour et al., 2011). An Optics Express submission is underway and another possible computer science journal publication
- facial dynamics related work (Benedikt et al., 2008) and (Aubrey et al., 2011), also another IEEE Signal processing submission is underway

5 Future Work

Based on progress with retinal and choroidal segmentation encourages future work on powerful methods for segmentation of 3D OCT retinal images using statistical models that combine shape and texture information learnt from a large set of training examples. Despite the success of the algorithms developed and presented in this work, there is still an important, and technically extremely challenging, subset of OCT data; severe pathologies, successful segmentation of which would enable therapy monitoring even in the later stages of retinal disease.

Problems arise due to variations in geometry, as layers do not always form a single, continuous region in a 2D slice, but can be fragmented creating many small blobs in a two-dimensional cross-section (e.g. RPE detachment), while even within the same tissue there are structural inhomogeneities due to the different orientations of the tissue. When segmenting layers, the presence of extraneous structures (e.g. blobs of fluid, shadows, vessels) interferes with the segmentation process as well.

These cases are the most challenging aspects of retinal image analysis, well beyond the current state of the art, and new approaches need to be developed. The methods available for segmenting non layer like structures (e.g. subretinal fluid, vessels) use simple thresholding techniques, and are not able to simultaneously perform layer segmentation (Ahlers et al., 2009).

A good solution to the given problems would continue in the spirit of machine learning and use a combination of a statistical model with iterative graph cuts. For the case of binary labelling (e.g. foreground/background), graph cuts provide an efficient (polynomial time) means for finding the optimal image segmentation (with respect to the given cost function) (Boykov, Veksler and Zabih, 2001). However, the choice of cost functions is limited (Kolmogorov and Zabih, 2004), and only approximate solutions are obtained for more than two labels (Boykov et al., 2001), but nevertheless graph cuts are an extremely powerful segmentation method, and have become very popular in the field of computer vision in the last few years.

As mentioned before, graph cuts have already been applied to segment OCT retina images (Garvin et al., 2009), but this work is limited in various respects. First, their images had very good SNR. Second, although the tissue imaged contained pathologies (anterior ischemic optic neuropathy), the structure was identical to normals except for a slight thickening of some layers. Third, they require manually specified heuristic constraints on the layers (smoothness, distance between layers, etc.). Thus, their graph cut approach would not directly extend to more varied pathologies.

Given the inherent difficulty of automatic segmentation it is worthwhile including all available prior information in the process. This has led to the recent trend to introduce spatial priors (the expected appearance of the modelled shape) within the graph cuts framework. One idea (Das et al., 2009) is to constrain the segmented foreground region to be compact, but they required user interaction and also used a simple (limited) definition of compactness. Another similar approach (Gulshan et al., 2010) used a star shape prior, again based on the user specifying one or several key points in the image. The above methods are not suitable for retinal image segmentation since they require a single, generic global shape descriptor, while we are interested in segmenting patchy areas with a certain distribution. Alternatively, it is possible to use a fixed shape template (Freedman and Zhang, 2005), but this is also too limited. The most relevant work for the proposed approach (Zhu-Jacquot and Zabih, 2007) combines a statistical shape and intensity model with graph cuts.

They use four energy terms:

- a single GMM to model the intensity probability distribution of the background
- the standard graph cut spatial homogeneity term
- a fitness between the statistical shape prior and the current segmentation
- the entropy of intensity distributions outside and inside the shape prior, which measures how well the shape prior fits the image and prevents an initial inaccurate segmentation from producing inaccurate shape priors

They iteratively update the GMM and shape prior parameters, followed by performing a graph cut, and repeat these steps until convergence. Another related work (Malcolm, Rathi and Tannenbaum, 2007) on combining a statistical shape prior with graph cuts was used in order to segment a connected shape (non-patchy).

The proposed approach differs from (Zhu-Jacquot and Zabih, 2007) since, given the layer fragmentation, it would have to drop the shape prior term and instead use, for each layer, multiple GMMs to model probability distributions of features which would be learnt from the training set. Fitting the model would involve updating the GMM parameters based on the graph cut segmentation at the previous iteration and would also be constrained by learning from the training set. This would allow a fragmented segmentation which would be nevertheless spatially constrained.

Another issue is that the existing manual labelling program used for the Mechanical Turk interface works with exact boundaries marking different layers. Since the new pathological data would have fragmented layers, a new method would need to be developed using region growing algorithms to ease access for well trained and expert human segmenters to identify the different layers, as the previous interface for users to identify layer boundaries would be unsuitable (i.e. too time consuming).. Therefore a new interface should be designed, based on a semi-automatic approach. The user approximately paints over the areas of interest, while the subsequent automatic stage performs the accurate segmentation based on brightness and some simple texture parameters. The user is then able to correct/refine these results by iterating the process.

As before, evaluation can be done both numerically against the manually segmented layers and clinically by cross-correlating disease staging with other classic methods of evaluation (i.e. fundus diagnostics).

It is important to mention that the developed method would be suitable for a general image segmentation problem and thus have a far wider area of application than just for the OCT field itself.

6 References

Ahlers C, Golbaz I, Einwallner E, Dunavölgyi R, Malamos P, Stock G, Pruente C et al. (2009) Identification of Optical Density Ratios in Subretinal Fluid as a Clinically Relevant Biomarker in Exudative Macular Disease. *Investigative Ophthalmology & Visual Science* 50: 3417-3424.

Ahlers C, Götzinger E, Pircher M, Golbaz I, Prager F, Schütze C, Baumann B et al. (2010) Imaging of the Retinal Pigment Epithelium in Age-Related Macular Degeneration Using Polarization-Sensitive Optical Coherence Tomography. *Investigative Ophthalmology & Visual Science* 51: 2149-2157.

Aubrey A J, Kajic V, Cingovska I, Rosin P L, and Marshall D (2011) Mapping and Manipulating Facial Dyanamics. In: *Int. Conf. on Automatic Face and Gesture Recognition*. Santa Barbara, CA, US.

Balu M, Liu G J, Chen Z P, Tromberg B J, and Potma E O (2010) Fiber delivered probe for efficient CARS imaging of tissues. *Optics Express* 18: 2380-2388.

Banaszek K, Radunsky A S, and Walmsley I A [eds.] (2004) *Blind dispersion compensation for optical coherence tomography*. Lasers and Electro-Optics, 2004. (CLEO). Conference on.

Beichel R, Bischof H, Leberl F, and Sonka M (2005) Robust active appearance models and their application to medical image analysis. *Medical Imaging, IEEE Transactions on* 24: 1151-1169.

Benedikt L, Kajic V, Marshall D, and Rosin P L (2008) Facial Dynamics in Biometric Identification. In: *British Machine Vision Conference*. London, UK.

Blankenberg F G, Katsikis P D, Tait J F, Davis R E, Naumovski L, Ohtsuki K, Kopiwoda S et al. (1998) In vivo detection and imaging of phosphatidylserine expression during programmed cell death. *Proceedings of the National Academy of Sciences of the United States of America* 95: 6349-6354.

Boppart S A, Bouma B E, Pitris C, Tearney G J, Fujimoto J G, and Brezinski M E (1997) Forward-imaging instruments for optical coherence tomography. *Opt. Lett.* 22: 1618.

Boykov Y, Veksler O, and Zabih R (2001) Fast Approximate Energy Minimization via Graph Cuts. *pami* 23: 1222-1239.

Broyden C G (1970) The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations. *IMA Journal of Applied Mathematics* 6: 76-90.

Cabrera Fernández D, Salinas H M, and Puliafito C A (2005) Automated detection of retinal layer structures on optical coherence tomography images. *Opt. Express* 13: 10200.

Cao L, and Fei-Fei L [eds.] (2007) *Spatially Coherent Latent Topic Model for Concurrent Segmentation and Classification of Objects and Scenes*. IEEE 11th International Conference on Computer Vision.

Cauchy A (1847) *Méthode générale pour la résolution des systèmes d'équations simultanées*.

Cereghetti G M, and Scorrano L (2006) The many shapes of mitochondrial death. *Oncogene* 25: 4717-4724.

Chen C H, Pau L F, and Wang P S P (1993) *Handbook of pattern recognition and computer vision*. World Scientific Publishing Co., Inc.

Chinn S R, Swanson E A, and Fujimoto J G (1997) Optical coherence tomography using a frequency-tunable optical source. *Opt. Lett.* 22: 340.

Chiu S J, Li X T, Nicholas P, Toth C A, Izatt J A, and Farsiu S (2010) Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation. *Opt. Express* 18: 19413-19428.

Cootes T F, and C.J.Taylor (2004) *Statistical Models of Appearance for Computer Vision*.

Cootes T F, Twining C J, Babalola K O, and Taylor C J (2008) Diffeomorphic statistical shape models. *Image Vision Comput.* 26: 326-332.

Cordeiro M F, Guo L, Luong V, Harding G, Wang W, Jones H E, Moss S E et al. (2004) Real-time imaging of single nerve cell apoptosis in retinal neurodegeneration. *Proceedings of the National Academy of Sciences of the United States of America* 101: 13352-13356.

Cortes C, and Vapnik V (1995) Support-Vector Networks. *Machine Learning* 20: 273.

Cristinacce D, and Cootes T F (2007) Boosted Regression Active Shape Models. In: *British Machine Vision Conference*.

Das P, Veksler O, Zavadsky V, and Boykov Y (2009) Semiautomatic segmentation with compact shape prior. *Image Vision Comput.* 27: 206-219.

De Souza-Ramalho P, Pereira Da Silva A, Pego P, and Bicho M P (2009) Diabetic retinopathy - Apoptosis and BRB breakdown prevention in early diabetic retinopathy. *Acta Ophthalmologica* 87: 0-0.

Debuc D C, Salinas H M, Ranganathan S, Tátrai E, Gao W, Shen M, Wang J et al. (2010) Improving image segmentation performance and quantitative analysis via a computer-aided grading methodology for optical coherence tomography retinal image analysis. *Journal of Biomedical Optics* 15.

Deng B, Jawerth B, Peters G, and Sweldens W (1993) Wavelet probing for compression based segmentation. In: Laine A F [ed.] *Mathematical Imaging: Wavelet Applications in Signal and Image Processing*. pp. 266--276.

Drexler W, and Fujimoto J G (2008) *Optical Coherence Tomography: Technology and Applications*. Springer.

Dunaief J L, Dentchev T, Ying G-S, and Milam A H (2002) The Role of Apoptosis in Age-Related Macular Degeneration. *Arch Ophthalmol* 120: 1435-1442.

Efros A A, and Freeman W T [eds.] (2001) *Image quilting for texture synthesis and transfer*. Proceedings of the 28th annual conference on Computer graphics and interactive techniques. ACM.

Esmaeelpour M, Povazay B, Hermann B, Hofer B, Kajic V, Hale S, North R V et al. (2011) Mapping choroidal and retinal thickness variation in Type 2 diabetes using 3D-1060nm-OCT. *Invest Ophthalmol Vis Sci*.

Esmaeelpour M, Povazay B, Hermann B, Hofer B, Kajic V, Kapoor K, Sheen N J et al. (2010) Three-dimensional 1060-nm OCT: choroidal thickness maps in normal subjects and improved posterior segment visualization in cataract patients. *Invest Ophthalmol Vis Sci* 51: 5260-5266.

Fabritius T, Makita S, Hong Y, Myllyl R, and Yasuno Y (2009a) Automated retinal shadow compensation of optical coherence tomography images. *Journal of Biomedical Optics* 14: 010503.

Fabritius T, Makita S, Miura M, Myllyla R, and Yasuno Y (2009b) Automated segmentation of the macula by optical coherence tomography. *Optics Express* 17: 15659-15669.

Fan-Yin T (2005) An Intelligent System Approach to Higher-Dimensional Classification of Volume Data. *IEEE Transactions on Visualization and Computer Graphics* 11: 273-284.

Fercher A, Hitzenberger C, Sticker M, Zawadzki R, Karamata B, and Lasser T (2001) Numerical dispersion compensation for Partial Coherence Interferometry and Optical Coherence Tomography. *Opt. Express* 9: 610.

Fercher A F, Mengedoht K, and Werner W (1988) Eye-length measurement by interferometry with partially coherent light. *Opt. Lett.* 13: 186.

Frank S, Gaume B, Bergmann-Leitner E S, Leitner W W, Robert E G, Catez F, Smith C L et al. (2001) The role of dynamin-related protein 1, a mediator of mitochondrial fission, in apoptosis. *Developmental Cell* 1: 515-525.

Freedman D, and Zhang T (2005) Interactive Graph Cut Based Segmentation with Shape Priors. *CVPR*: 755-762.

Garvin M M K, Abramoff M M D, Kardon R R, Russell S S R, Wu X X, and Sonka M M (2009) Intraretinal Layer Segmentation of Macular Optical Coherence Tomography Images Using Optimal 3-D Graph Search. *IEEE Transactions on Medical Imaging* 27: 1495-1505.

Gilboa G, Sochen N, and Zeevi Y Y (2004) Image Enhancement and Denoising by Complex Diffusion Processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26: 1020-1036.

Glocker B, Komodakis N, Tziritas G, Navab N, and N.Paragios (2008) Dense Image Registration through MRFs and Efficient Linear Programming. *Medical Image Analysis*.

Gossage K W, Smith C M, Kanter E M, Hariri L P, Stone A L, Rodriguez J J, Williams S K et al. (2006) Texture analysis of speckle in optical coherence tomography images of tissue phantoms. *Physics in Medicine and Biology* 51: 1563.

Gourley P L, Hendricks J K, McDonald A E, Copeland R G, Barrett K E, Gourley C R, Singh K K et al. (2005) Mitochondrial correlation microscopy and nanolaser spectroscopy - New tools for biophotonic detection of cancer in single cells. *Technology in Cancer Research & Treatment* 4: 585-592.

Green D R, and Reed J C (1998) Mitochondria and apoptosis. *Science* 281: 1309-1312.

Grulkowski I, Gora M, Szkulmowski M, Gorczynska I, Szlag D, Marcos S, Kowalczyk A et al. (2009) Anterior segment imaging with Spectral OCT system using a high-speed CMOS camera. *Opt. Express* 17: 4842-4858.

Gulshan V, Rother C, Criminisi A, Blake A, and Zisserman A (2010) Geodesic Star Convexity for Interactive Image Segmentation. *CVPR*.

Guyon I, and Elisseeff A (2003) An introduction to variable and feature selection. *J. Mach. Learn. Res.* 3: 1157-1182.

Hazaveh K (2003) Image Processing Techniques for Noise Removal, Enhancement and Segmentation of Cartilage OCT Images – A Review and Critique. In: *2nd International Conference for Upcoming Engineers*.

Hazaveh K, and Raahemifar K (2003) Local discriminant basis algorithm-a review of theory and applications in signal processing. *Circuits and Systems, 2003. ISCAS '03. Proceedings of the 2003 International Symposium* 4: 341-344.

Hee M R, Puliafito C A, Duker J S, Reichel E, Coker J G, Wilkins J R, Schuman J S et al. (1998) Topography of diabetic macular edema with optical coherence tomography. *PubMed* 105: 360-370.

Herz P, Chen Y, Aguirre A, Fujimoto J, Mashimo H, Schmitt J, Koski A et al. (2004) Ultrahigh resolution optical biopsy with endoscopic optical coherence tomography. *Opt. Express* 12: 3532.

Hofer B, Povaz ay B, Hermann B, Rey S M, Kajic V, Tumlinson A, Powell K et al. (2011a) Artefact reduction for cell migration visualization using spectral domain optical coherence tomography. *J Biophotonics* 4: 355-367.

Hofer B, Považay B, Unterhuber A, Wang L, Hermann B, Rey S, Matz G et al. (2011b) Fast dispersion encoded full range OCT for retinal imaging at 800 nm and 1060 nm. 18: 4898-4919.

Huang D, Wang J, Lin C P, Puliafito C A, and Fujimoto J G (1991) Micron-resolution ranging of cornea and interior chamber by optical reflectometry. *Lasers Surg. Med.*

Huber R, Wojtkowski M, Taira K, Fujimoto J, and Hsu K (2005) Amplified, frequency swept lasers for frequency domain reflectometry and OCT imaging: design and scaling principles. *Opt. Express* 13: 3513.

Ibanez L, Schroeder W, Ng L, and Cates J (2005) *The ITK Software Guide, Second Edition*.

Jan-Mark G, and Arnold W M S (2005) A Six-Stimulus Theory for Stochastic Texture. *Int. J. Comput. Vision* 62: 7-16.

Jiao S, Knighton R, Huang X, Gregori G, and Puliafito C (2005) Simultaneous acquisition of sectional and fundus ophthalmic images with spectral-domain optical coherence tomography. *Opt. Express* 13: 444.

Jiao S, Wu C, Knighton R W, Gregori G, and Puliafito C A (2006) Registration of high-density cross sectional images to the fundus image in spectral-domain ophthalmic optical coherence tomography. *Opt. Express* 14: 3368.

Joachim W (1999) Coherence-Enhancing Diffusion Filtering. *Int. J. Comput. Vision* 31: 111-127.

Kajic V, Povazay B, Hermann B, Hofer B, Marshall D, Rosin P L, and Drexler W (2010) Robust segmentation of intraretinal layers in the normal human fovea using a novel statistical model based on texture and shape analysis. *Opt Express* 18: 14730-14744.

Karbowski M, and Youle R J (2003) Dynamics of mitochondrial morphology in healthy cells and during apoptosis. *Cell Death Differ* 10: 870-880.

Karmarkar N (1984) A new polynomial-time algorithm for linear programming. In: *Proceedings of the sixteenth annual ACM symposium on Theory of computing*. ACM.

Kingsbury N (1999) Image processing with complex wavelets. *Philosophical Transactions of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences* 357: 2543-2560.

Kirk W G, Cynthia M S, Elizabeth M K, Lida P H, Alice L S, Jeffrey J R, Stuart K W et al. (2006) Texture analysis of speckle in optical coherence tomography images of tissue phantoms. *Physics in Medicine and Biology*: 1563.

Kohli P, and Torr P (2007) Dynamic Graph Cuts for Efficient Inference in Markov Random Fields. *IEEE Trans. Pattern Anal. Mach. Intell.* 29: 2079-2088.

Kolmogorov V, and Zabih R (2004) What Energy Functions Can Be Minimized via Graph Cuts? *pami* 26: 147-159.

Komodakis N, Tziritas G, and Paragios N (2008) Performance vs Computational Efficiency for Optimizing Single and Dynamic MRFs: Setting the State of the Art with Primal Dual Strategies. *Computer Vision and Image Understanding*.

Koozekanani D, Boyer K, and Roberts C (2001) Retinal thickness measurements from optical coherence tomography using a Markov boundary model. *Medical Imaging, IEEE Transactions on* 20: 900.

Kulkarni M D, Thomas C W, and Izatt J A (1997) Image enhancement in optical coherence tomography using deconvolution. *IEE Electronics Letters* 33: 1365-1367.

Leitgeb R, Hitzenberger C, and Fercher A (2003) Performance of fourier domain vs. time domain optical coherence tomography. *Opt. Express* 11: 889.

Lemaur G, Drouiche K, and DeConinck J (2003) Highly regular wavelets for the detection of clustered microcalcifications in mammograms. *Medical Imaging, IEEE Transactions on* 22: 393.

Levenberg K (1944) A method for the solution of certain non-linear problems in least squares. *Quarterly Journal of Applied Mathmatics* 2: 164-168.

Lexer F, Hitzenberger C K, Fercher A F, and Kulhavy M (1997) Wavelength-tuning interferometry of intraocular distances. *Appl. Opt.* 36: 6548.

Li L, Socher R, and Fei L (2009) Towards total scene understanding: Classification, annotation and segmentation in an automatic framework. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on* 0: 2036-2043.

Malcolm J, Rathi Y, and Tannenbaum A (2007) Graph Cut Segmentation with Nonlinear Shape Priors. *icip*: 365-368.

Marks D L, Oldenburg A L, Reynolds J J, and Boppart S A (2003) Digital Algorithm for Dispersion Correction in Optical Coherence Tomography for Homogeneous and Stratified Media. *Appl. Opt.* 42: 204.

Matas J, Chum O, Urban M, and Pajdla T (2004) Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing* 22: 761-767.

Mattson M P (2000) Apoptosis in neurodegenerative disorders. *Nat Rev Mol Cell Biol* 1: 120-129.

Mayer M A, Hornegger J, Mardin C Y, and Tornow R P (2010) Retinal Nerve Fiber Layer Segmentation on FD-OCT Scans of Normal Subjects and Glaucoma Patients. *Biomed. Opt. Express* 1: 1358-1383.

McGowan D (2006) Pruning processes. *Nat Rev Neurosci* 7: 685-685.

Mishra A, Wong A, Clausi D A, and Fieguth P W (2010a) Quasi-random nonlinear scale space. *Pattern Recognition Letters* In Press, Corrected Proof.

Mishra A K, Fieguth P W, and Clausi D A (2010b) Decoupled Active Contour (DAC) for Boundary Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33: 310-324.

Molnár J, Chetverikov D, Cabrera DeBuc D, Gao W, and Somfai G (2011) Layer extraction in rodent retinal images acquired by optical coherence tomography. *Machine Vision and Applications*: 1-11.

Mourant J R, Freyer J P, Hielscher A H, Eick A A, Shen D, and Johnson T M (1998) Mechanisms of light scattering from biological cells relevant to noninvasive optical-tissue diagnostics. *Applied Optics* 37: 3586-3593.

Mujat M, Chan R, Cense B, Park B, Joo C, Akkin T, Chen T et al. (2005) Retinal nerve fiber layer thickness map determined from optical coherence tomography images. *Opt. Express* 13: 9480.

Nocedal J, and Wright S J (1999) *Numerical Optimization*. Springer.

Ojala T, and Pietikäinen M (1997) Unsupervised texture segmentation using feature distributions. In: Del B [ed.] *Image Analysis and Processing*. Vol. 1310. Springer Berlin / Heidelberg, pp. 311-318.

Palade G E (1953) An electrone micrpscope study of the mitochondrial structure. *Journal of Histochemistry & Cytochemistry* 1: 188-211.

Pasternack R M, Zheng J Y, and Boustany N N (2010) Optical scatter changes at the onset of apoptosis are spatially associated with mitochondria. *Journal of Biomedical Optics* 15.

Perkon I, Košir A, Tasic J, and Diamond M (2010) Whisker detection as a shortest path problem. In: *Visual Observation and Analysis of Animal and Insect Behavior*.

Perona P, and Malik J (1990) Scale-Space and Edge Detection Using Anisotropic Diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.* 12: 629-639.

Podoleanu A, Charalambous I, Plesea L, Dogariu A, and Rosen R (2004) Correction of distortions in optical coherence tomography imaging of the eye. *Physics in Medicine and Biology* 49: 1277.

Povazay B, Bizheva K, Unterhuber A, Hermann B, Sattmann H, Fercher A F, Drexler W et al. (2002) Submicrometer axial resolution optical coherence tomography. *Opt. Lett.* 27: 1800.

Povazay B, Hermann B, Hofer B, Kajic V, Simpson E, Bridgford T, and Drexler W (2009) Wide-field optical coherence tomography of the choroid in vivo. *Invest Ophthalmol Vis Sci* 50: 1856-1863.

Prilloff S, Fan J Y, Henrich-Noack P, and Sabel B A (2010) In vivo confocal neuroimaging (ICON): non-invasive, functional imaging of the mammalian CNS with cellular resolution. *European Journal of Neuroscience* 31: 521-528.

Pshenichny B N, and Danilin Y M (1983) *Numerical Methods in External Problems*. Mir Publishers, Moscow.

Puliafito C A, Hee M R, Lin C P, Reichel E, Schuman J S, Duker J S, Izatt J A et al. (1995) Imaging of macular diseases with optical coherence tomography. *Ophthalmology* 102: 217.

Quigley H A, Nickells R W, Kerrigan L A, Pease M E, Thibault D J, and Zack D J (1995) Retinal ganglion-cell death in experimental glaucoma and after axotomy occurs by apoptosis. *Investigative Ophthalmology & Visual Science* 36: 774-786.

Reichstein D, Ren L Z, Filippopoulos T, Mittag T, and Danias J (2007) Apoptotic retinal ganglion cell death in the DBA/2 mouse model of glaucoma. *Experimental Eye Research* 84: 13-21.

Robertson J D, Orrenius S, and Zhivotovsky B (2000) Nuclear events in apoptosis. *Journal of Structural Biology* 129: 346-358.

Rossant F, Ghorbel I, Bloch I, Paques M, and Tick S (2009) Automated segmentation of retinal layers in OCT imaging and derived ophthalmic measures. In: *Proceedings of the Sixth IEEE international conference on Symposium on Biomedical Imaging: From Nano to Macro*. Boston, Massachusetts, USA, 1700225: IEEE Press, pp. 1370-1373.

Rottey S, Loose D, Vakaet L, Lahorte C, Vermeersch H, Van Belle S, and Van De Wiele C V (2007) Tc-99m-HYNIC Annexin-V imaging of tumors and its relationship to response to radiotherapy and/or chemotherapy. *Quarterly Journal of Nuclear Medicine and Molecular Imaging* 51: 182-188.

Rottey S, Slegers G, Van Belle S, Goethals I, and Van de Wiele C (2006) Sequential Tc-99m-hydrazinonicotinamide-annexin V imaging for predicting response to chemotherapy. *Journal of Nuclear Medicine* 47: 1813-1818.

Sammon J W (1969) A nonlinear mapping for data structure analysis. *Ieee Transactions on Computers* C 18: 401-&.

Sarunic M V, Yazdanpanah A, Gibson E, Xu J, Bai Y, Lee S, Saragovi H U et al. (2010) Longitudinal study of retinal degeneration in a rat using spectral domain optical coherence tomography. *Opt. Express* 18: 23435-23441.

Schmitt J M (1998) Restoration of Optical Coherence Images of Living Tissue Using the CLEAN Algorithm. *Journal of Biomedical Optics* 3: 66.

Schmitt J M, Xiang S H, and Yung K M (1999) Speckle in Optical Coherence Tomography. *Journal of Biomedical Optics* 4: 95.

Schmitz-Valckenberg S, Guo L, Maass A, Cheung W, Vugler A, Moss S E, Munro P M G et al. (2008) Real-time in vivo imaging of retinal cell apoptosis after laser exposure. *Investigative Ophthalmology & Visual Science* 49: 2773-2780.

Scholz M (2005) Non-linear PCA: a missing data approach. *Bioinformatics* 21: 3887-3895.

Scholz M, Fraunholz M, and Selbig J (2007) Nonlinear Principal Component Analysis: Neural Network Models and Applications.*Principal Manifolds for Data Visualization and Dimension Reduction*. pp. 44-67.

Selesnick I W, Baraniuk R G, and Kingsbury N G (2005) The Dual-Tree Complex Wavelet Transform. *IEEE Signal Processing Magazine* 22: 123-151.

Sorzano C O S, Thevenaz P, and Unser M (2005) Elastic registration of biological images using vector-spline regularization. *Biomedical Engineering, IEEE Transactions* 52: 652.

Spall J C [ed.] (1997) *Accelerated second-order stochastic optimization using only function measurements*. Decision and Control, 1997., Proceedings of the 36th IEEE Conference on.

Swanson E A, Izatt J A, Hee M R, Huang D, Lin C P, Schuman J S, Puliafito C A et al. (1993) In vivo retinal imaging by optical coherence tomography. *Opt. Lett.* 18: 1864.

T.F.Cootes, Edwards G J, and C.J.Taylor (2001) Active Appearance Models. *IEEE PAMI* 23: 681-685.

Terreno E, Castelli D D, and Aime S (2010) Encoding the frequency dependence in MRI contrast media: the emerging class of CEST agents. *Contrast Media & Molecular Imaging* 5: 78-98.

Thevenaz P, and Unser M [eds.] (1996) *A pyramid approach to sub-pixel image fusion based on mutual information*. Image Processing, 1996. Proceedings., International Conference on Image Processing.

Thornberry N A (1997) The caspase family of cysteine proteases. *British Medical Bulletin* 53: 478-490.

Thornberry N A (1998) Caspases: key mediators of apoptosis. *Chemistry & Biology* 5: R97-R103.

Timo O, Matti P, inen, Topi M, and enp (2002) Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* 24: 971-987.

Tolliver D, Koutis Y, Ishikawa H, Schuman J S, and Miller G L [eds.] (2008) *Unassisted Segmentation of Multiple Retinal Layers via Spectral Rounding*. ARVO.

van der Meer F J, Faber D J, Aalders M C G, Poot A A, Vermes I, and van Leeuwen T G (2010) Apoptosis- and necrosis-induced changes in light attenuation measured by optical coherence tomography. *Lasers in Medical Science* 25: 259-267.

Vedaldi A, and Fulkerson B (2008) *VLFeat: An Open and Portable Library of Computer Vision Algorithms* [Online]. Available at: http://www.vlfeat.org/ [Accessed:

Vercauteren T, Perchant A, Malandain G, Pennec X, and Ayache N (2006) Robust mosaicing with correction of motion distortions and tissue deformations for in vivo fibered microscopy. *Medical Image Analysis* 10: 673.

Vermeer K A, van der Schoot J, Lemij H G, and de Boer J F (2011) Automated segmentation by pixel classification of retinal layers in ophthalmic OCT images. *Biomed. Opt. Express* 2: 1743-1756.

Vermes I, Haanen C, and Reutelingsperger C (2000) Flow cytometry of apoptotic cell death. *Journal of Immunological Methods* 243: 167-190.

Vermes I, Haanen C, Steffensnakken H, and Reutelingsperger C (1995) A novel assay for apoptosis - flow cytometric detection of phosphatidylserine expression on early apoptotic cells using fluorescein-labeled annexin-v. *Journal of Immunological Methods* 184: 39-51.

Viola P, and Jones M (2001) Rapid object detection using a boosted cascade of simple features. *CVPR*.

Wang I J, and Spall J C [eds.] (1998) *A constrained simultaneous perturbation stochastic approximation algorithm based on penalty functions*. Intelligent Control (ISIC), 1998. Held jointly with IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA), Intelligent Systems and Semiotics (ISAS), Proceedings.

Weber A J, Kaufman P L, and Hubbard W C (1998) Morphology of single ganglion cells in the glaucomatous primate retina. *Investigative Ophthalmology & Visual Science* 39: 2304-2320.

Weissman J, Hancewicz T, and Kaplan P (2004) Optical coherence tomography of skin for measurement of epidermal thickness by shapelet-based image analysis. *Opt. Express* 12: 5760.

Westphal V, Rollins A, Radhakrishnan S, and Izatt J (2002) Correction of geometric and refractive image distortions in optical coherence tomography applying Fermat?s principle. *Opt. Express* 10: 397.

Wong A, Mishra A, Bizheva K, and Clausi D A (2010) General Bayesian estimation for speckle noise reduction in optical coherence tomography retinal imagery. *Optics Express* 18: 8338-8352.

Wong E, Kumar V, Howman-Giles R B, and Vanderheyden J L (2008) Imaging of Therapy-Induced Apoptosis Using Tc-99m-HYNIC-Annexin V in Thymoma Tumor-Bearing Mice. *Cancer Biotherapy and Radiopharmaceuticals* 23: 715-725.

Wright A J, Poland S P, Girkin J M, Freudiger C W, Evans C L, and Xie X S (2007) Adaptive optics for enhanced signal in CARS microscopy. *Optics Express* 15: 18209-18219.

Yang Q, Reisman C A, Wang Z, Fukuma Y, Hangai M, Yoshimura N, Tomidokoro A et al. (2010) Automated layer segmentation of macular OCT images using dual-scale gradient information. *Opt. Express* 18: 21293-21307.

Yazdanpanah A, Hamarneh G, Smith B, and Sarunic M (2009) Intra-retinal Layer Segmentation in Optical Coherence Tomography Using an Active Contour Approach. In: *Proceedings of the 12th International Conference on Medical Image Computing and Computer-Assisted Intervention: Part II*. London, UK, 1691261: Springer-Verlag, pp. 649-656.

Yoo B, and Pagel M D (2008) An overview of responsive MRI contrast agents for molecular imaging. *Frontiers in Bioscience* 13: 1733-1752.

Yu K, Ji L, Wang L, and Xue P (2001) How to optimize OCT image. *Opt. Express* 9: 24.

Yun S H, Tearney G, de Boer J, and Bouma B (2004) Motion artifacts in optical coherence tomography with frequency-domain ranging. *Opt. Express* 12: 2977.

Yung K M, Lee S L, and Schmitt J M (1999) Phase-Domain Processing of Optical Coherence Tomography Images. *Journal of Biomedical Optics* 4: 125.

Zawadzki R J, Choi S S, Jones S M, Oliver S S, and Werner J S (2007) Adaptive optics-optical coherence tomography: optimizing visualization of microscopic retinal structures in three dimensions. *J. Opt. Soc. Am. A* 24: 1373.

Zhu-Jacquot J, and Zabih R (2007) Graph Cuts Segmentation with Statistical Shape Priors for Medical Images. *IEEE Conf. on Signal-Image Technologies and Internet-Based System*: 631-635.

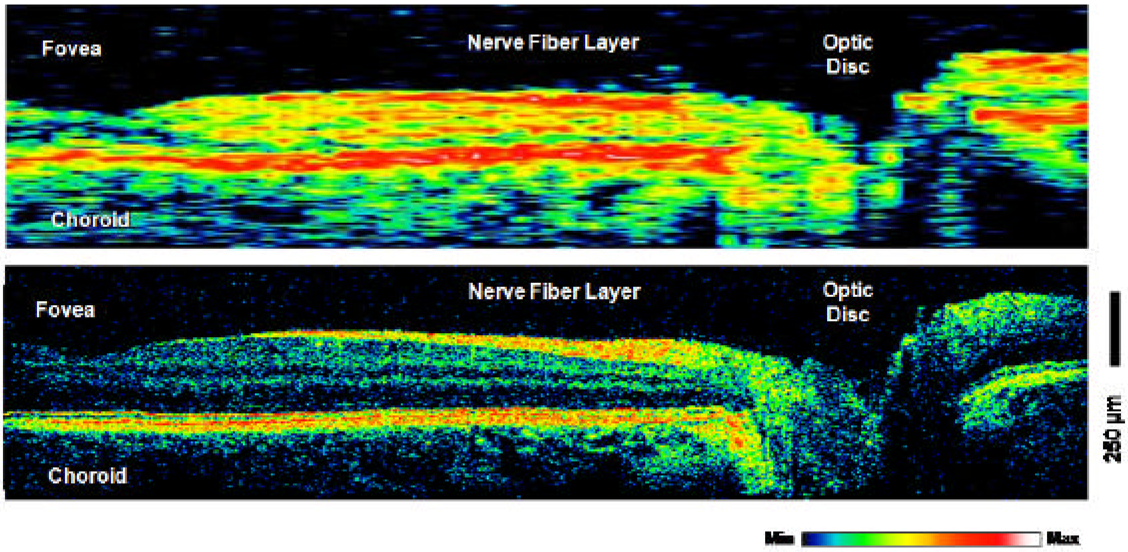


Figure 1 Comparison of standard (top) and ultrahigh (bottom) OCT (Drexler and Fujimoto, 2008)

Zvyagin A V, Blazkiewicz P, and Vintrou J (2005) Image reconstruction in full-field Fourier-domain optical coherence tomography. *Journal of Optics A: Pure and Applied Optics* 7: 350.

.

To understand the importance of OCT as an imaging technique it is necessary to compare it with other imaging methods such as ultrasound and confocal microscopy. Similar as in OCT, where resolution depends on the wavelength of employed light, in ultrasound resolution depends on the wavelength of the used sound waves. Standard clinical ultrasound frequencies are in a range of 3-40 MHz and result in spatial resolutions of approximately 0.1 – 1 mm, significantly lower than that of OCT. However, the main advantage of ultrasound is the property of sound waves that allows them to easily penetrate deep into biological tissue, enabling imaging at several centimetre depths in the investigated organ. It should also be mentioned that high-resolution ultrasound has been developed which uses high-frequency (100 MHz) waves to allow for axial resolutions of up to 15µm. However, high-frequency ultrasound is strongly attenuated in biological tissue which limits imagining depths to only a few millimetres. Also, ultrasound is a contact method and therefore risk of infection is present (Drexler and Fujimoto, 2008).

Confocal microscopy is an optical imaging technique that uses point illumination and a pinhole in an optically conjugate plane in front of the detector to eliminate out-of-focus information and thus provides much better image quality with resolutions of about 1 µm

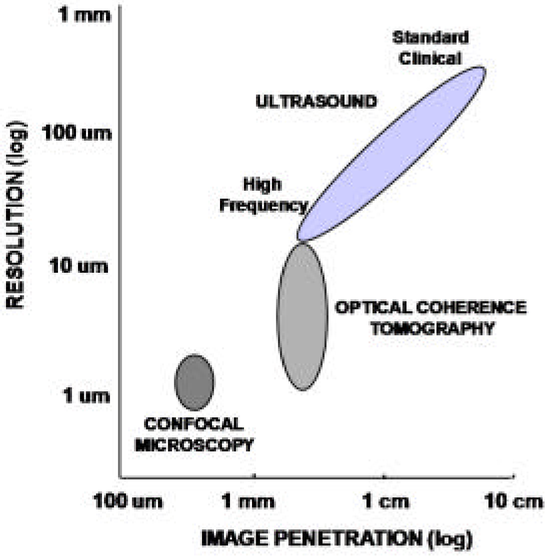


Figure 2 Comparison of OCT to ultrasound and confocal microscopy (Drexler and Fujimoto, 2008)

(Drexler and Fujimoto, 2008). However, since it is a direct method of optical measurement, image signal and contrast are degraded by optical scattering, which allows for effective imaging depths of only a few hundred microns, while OCT can achieve high axial resolution independent of the beam focusing and spot size. The transverse resolution of confocal microscopy is determined by the diffraction limited spot size of the focused optical beam, and defines the OCT transverse resolution as well. From the specifications of these two techniques, confocal microscopy with high resolution and poor penetration and ultrasound with low resolution and very high penetration depth, it is clear that a gap exists, as shown in Figure 2, that is filled by OCT. Offering penetration depths of a few millimetres and high resolution, OCT enables the visualization of tissue structure at the level of architectural morphology. It is best suited for ophthalmology since the eye is almost transparent, therefore enabling easy optical access. OCT can also be used in other medical fields since it is an optical method that can be easily integrated into various investigative instruments, which are well known to the physician: endoscopes, laparoscopes, catheters, or needles which enable access to internal organs (Boppart et al., 1997; Herz et al., 2004). Principle of operation of all OCT systems is based on interferometry of the light beam coming from the sample arm of the system and the beam from the reference arm. There are three different types of OCT systems: Time domain OCT and two types of Frequency domain OCT (Fercher, Mengedoht and Werner, 1988; Huang et al., 1991; Chinn, Swanson and Fujimoto, 1997; Lexer et al., 1997; Leitgeb, Hitzenberger and Fercher, 2003; Huber et al., 2005). The latter one can be divided into Spectral domain (spatially encoded frequency domain) and Swept source (time encoded frequency domain) OCT. Also, instead of using a single beam for scanning it is possible to increase the acquisition speed by using Full-field OCT and acquiring the whole volume at once (Zvyagin, Blazkiewicz and Vintrou, 2005). Full-field illumination of the sample is performed and it is imaged with a charge-coupled device (CCD) or complementary metal–oxide–semiconductor (CMOS) camera, hence eliminating the electromechanical lateral scan. However, this introduces signal cross talk between the pixels and reduces the image quality.

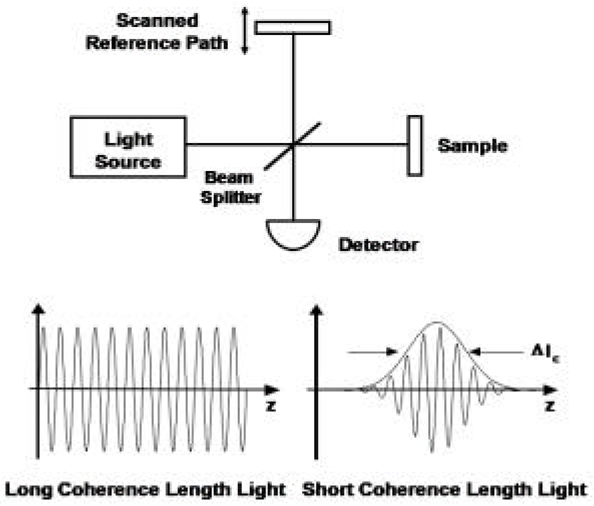


Figure 3 Basic Time domain OCT setup (Drexler and Fujimoto, 2008), signal is obtained by mirror scan (z axis)

Time domain OCT, shown in Figure 3, is the original version of OCT and uses a depth scanning mirror in a Michelson type interferometer (Huang et al., 1991). The beam from the light source is split into two parts: the sample and the reference arm. The investigated specimen is placed in the sample arm, while the reference arm is terminated by the movable mirror that determines the actual data acquisition speed in time domain OCT. Beams from both arms interfere and this interference intensity is measured at the detector. Depth scanning is performed by moving the reference arm mirror. This will change the relative path length and actually change the system in a way that it becomes sensitive to the light reflected from the different depth inside the sample. This is achieved by using broad bandwidth light source that will provide a short coherence length light, so the detected signal will be localised to a certain depth (coherence gating). If narrow linewidth light sources are used (corresponding to a long coherence length), than it would not be possible to determine from where the signal is originating due to the fact that constructive and destructive interference is detected along the entire depth scan.

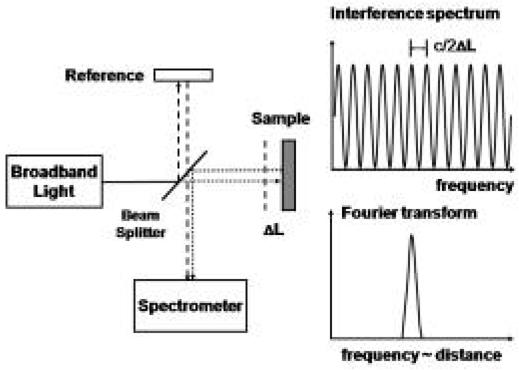


Figure 4 Basic Spectral Fourier domain OCT setup (Drexler and Fujimoto, 2008)

Frequency domain OCT systems have a stationary reference arm and rely on the property of the acquired spectra that contains higher frequency components for the scattering surfaces further away from the interferometer, i.e. at greater depths (Lexer et al., 1997). By applying an inverse Fourier transform on the detected spectra, it is possible to retrieve distances and amplitudes of the back reflected signal components. An important advantage of Frequency domain systems is significantly higher signal detection efficacy (since all of the depth signals are measured at once) resulting in a data acquisition speed or sensitivity increase as compared to time domain OCT systems (Leitgeb et al., 2003). Volumetric imaging in a variety of medical fields is therefore possible with significantly less motion artefacts.

Spectral domain system, as depicted in Figure 4, use a broad bandwidth light source and the interference beam is passed through a grating and then to a line scan camera that forms a spectrometer (Lexer et al., 1997). The light from the sample beam is reflected at different tissue depths, while the second beam is reflected from the fixed reference mirror. When the two combine the interference signal which is produced will have a spectral modulation as a function of frequency since the oscillation frequency depends on the path difference between the sample and the reference arm. Different echo delays that produce different frequency modulations are measured by rescaling the spectrometer output from wavelength to frequency and then Fourier transform is performed on the interference signal to retrieve spatial information.

Swept source domain imaging uses a narrow linewidth light source while the frequency is swept in time over a broad bandwidth (Chinn et al., 1997; Huber et al., 2005). Figure 5 shows a basic swept source OCT system setup. The narrow linewidth frequency swept beam is split by a beam splitter; one beam illuminates the sample while the other beam is back reflected from the static mirror. The path length difference between the structure in the tissue and reference mirror will be detected as a modulation in intensity at a frequency which is proportional to the spatial offset of the structure, since the frequency of the light is swept as a function of time. These echo delays are transformed into the spatial domain by applying Fourier transform.

Since this thesis primarily focuses on the segmentation of retinal OCT data, applications of OCT to the field of ophthalmology have to be presented in more detail. Because of its optical transparency, applying OCT in the eye is obvious and has so far been the most successful application of OCT (Swanson et al., 1993). In ophthalmology, OCT enables non-invasive imaging of the anterior eye segment (Grulkowski et al., 2009) and the posterior one (retina and choroid) despite its low optical backscattering (Puliafito et al., 1995). In the retina, OCT not only detects highly scattering structures like the nerve fibre layer (NFL) and the retinal pigment epithelium (RPE), which contains melanin, and the highly vascular choroid but it also differentiates all cellular layers: ganglion cell layer, inner plexiform layer, inner nuclear layer, outer plexiform layer, outer nuclear layer and inner and outer segment of the photoreceptive layer (Povazay et al., 2002). OCT can be successfully applied to imaging of various pathologies: macular edema, macular holes, age-related macular degeneration, central serous chorioretinopathy, choroidal neovascularisation and others (Puliafito et al., 1995). However, to improve the data quality and to obtain quantitative measurements for clinical diagnosis, computer

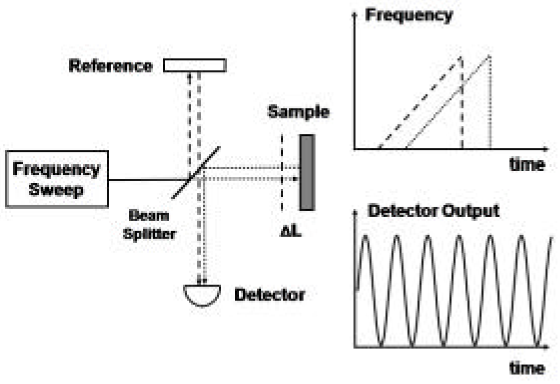


Figure 5 Basic Swept source Fourier domain OCT setup (Drexler and Fujimoto, 2008)

processing algorithms are extremely important. Axial motion is usually not large but transverse eye motion still remains difficult to correct. Layer segmentation is another important issue. Most commercial OCT systems already come with built-in algorithms for retinal thickness measurement and NFL thickness measurement. These algorithms are still very limited and prone to failure in cases of pathologies since they rely on fixed tolerance levels applied to the expected boundary positions and edge/peak intensities (Debuc et al., 2010). Thickness measurements of specific intraretinal layers can be effectively used for diagnosis and monitoring of various pathologies such as glaucoma, macular edema (Hee et al., 1998) or age related macular degeneration (Ahlers et al., 2010). Furthermore it is important to use a large normative database for early detection of pathologies to take anatomical variation in normals into account. To effectively display this thickness information different methods have been developed, such as thickness maps, which is a useful display method that shows from top down different layer thicknesses in colour coding (Hee et al., 1998).

6.1 OCT Signal Degradation

Noise is an important issue in OCT imaging and successful denoising is an important prerequisite for successful image registration and segmentation. There are several causes of signal degradation in OCT imaging such as light dispersion, optical speckle, refraction, quantization and motion artefacts.

6.1.1 Light Dispersion

Dispersion is caused by the property of light that different wavelengths travel at different speeds through a medium. It is a significant limiting factor to the enhancement of OCT depth resolution by employing light sources with broader spectral bandwidth, since dispersion of the sample increases the coherence length of the signal field. The standard technique to avoid this consequence is to balance the dispersion of the sample by placing a dispersive material in the reference arm. However, it is difficult to implement a corresponding depth dependent dispersion balancing technique since the impact of dispersion is depth dependent.

Most of the developed software dispersion compensation techniques rely on the prior knowledge of dispersive properties of the measured sample, such as water and proteins (Fercher et al., 2001; Marks et al., 2003). The method for blind dispersion compensation does not assume any specific dispersion distribution. It computes a generalized autoconvolution function from the mutual coherence function of the signal and reference beam (Banaszek, Radunsky and Walmsley, 2004). A weighting factor is used to suppress unwanted artefacts in the autocorrelation function.

6.1.2 Light Refraction

Several software methods have been developed that deal with refraction and other geometric distortions. Westphal et al. presented a method for correction of nonlinear axial scan patterns, non-telecentric (diverging or converging) scanning and refraction in layered media based on Fermat's principle (Westphal et al., 2002). First two algorithms are implemented on the raw data, while the third one works on the final image and thus constitutes a post-processing phase. Nonlinear axial scanning refers to almost sinusoidal motion in the reference arm of time domain systems and can be corrected by a backward transformation that takes this motion assumption into account. Geometric image distortions due to non-telecentric scanning are corrected by applying a backward transformation based on defining the coordinate systems for the raw and target images in terms of the scan geometry. Goal of the refraction correction is to transform optical paths into physical paths. This method is based on the assumption of having values of refraction indices for all the sample layers and relies on the segmented layer boundaries, obtained manually in this case. If forward transform was to be obtained it would be possible to simply apply Snell's law. But since Snell's law becomes complicated in raw image coordinates it is simpler to use Fermat's principle in the distortion-free image target coordinates, which states that optical path between source and target should be minimal. Podoleanu et al. based their approach for refraction correction on a different procedure, using the refraction law only (Podoleanu et al., 2004).

6.1.3 Optical Speckle

Speckle is the most important feature obscuring physical phenomena that occurs in OCT and is clearly visible in the obtained images giving them a grainy look and often obscuring the finer details. Speckle results from the coherent addition of multiple waves of different phases and was first studied in more detail by early researchers working with lasers, although it appears in all types of narrowband detection systems such as radar, ultrasound and radio astronomy. Laser light reflected back from a rough surface produces an image containing a random pattern of bright and dark spots in no obvious relation to the texture of the surface, changing whenever the surface moves slightly. Light can be backscattered, forward scattered or even out-of-field multiply scattered. Scatterers of different sizes and distributions in the sample contribute to the observed speckle pattern. Speckle pattern is also influenced by the motion of the sample and the aperture of the detector. If two or more scatterers in the sample volume backscatter waves that reach some point on the detector out of phase within an interval of time less than the coherence time of the source speckle will occur. Speckle will appear in cases that the backscattered waves from any pair of point scatterers are separated by an optical distance close to an odd multiple of one half of the wavelength, provided that the optical distance does not exceed the coherence length of the source in the medium (Schmitt, Xiang and Yung, 1999). Axial speckle size is correlated with coherence length and transverse speckle size is correlated with spot size of the lightsource. Since there is a large number of subwavelength-diameter scatterers in the biological tissue it is very likely that many pairs or clusters of particles will generate speckle. To simply illustrate noise produced by the multiple backscattering speckle effect one A-scan may be considered with a dense distribution of small scatterers. By convolving this distribution with theoretical point-spread function (PSF) of an OCT scanner, the coherent A-scan signal is produced. The incoherent signal is produced by convolving the scatterers' distribution with the envelope of the PSF. By looking at the envelope of the coherent signal it is obvious that it tracks original distribution of scatterers poorly. This is shown in Figure 6. The same speckle effect occurs in 2D and is visible as a superimposed pattern of dark and bright spots over the original signal.

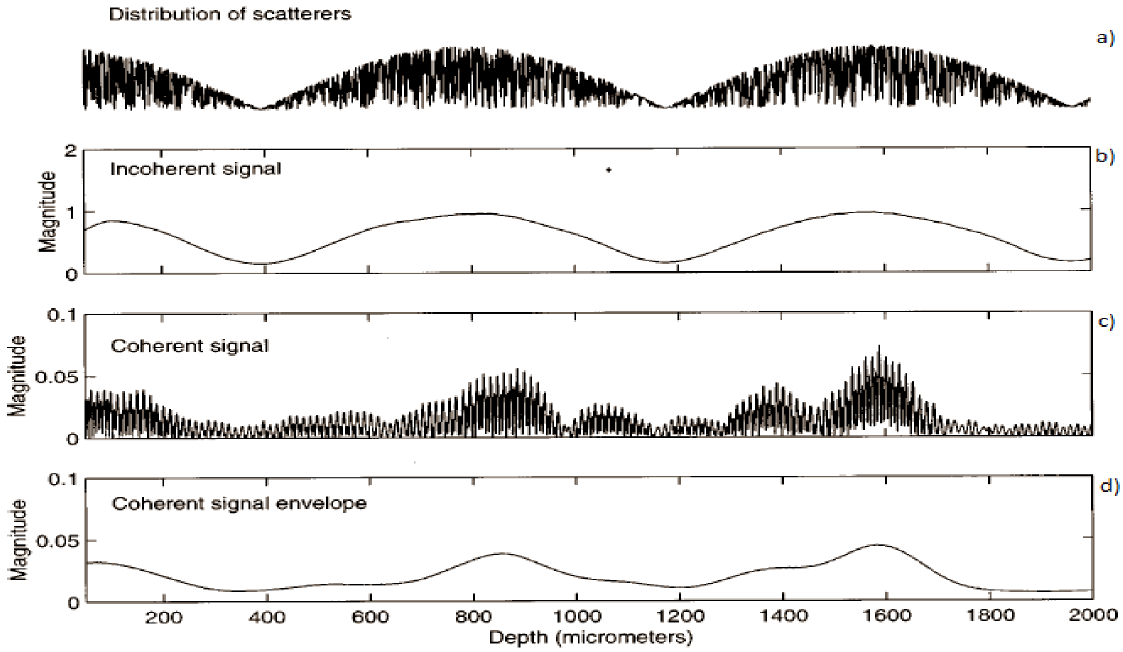


Figure 6 Simulation of the speckle noise created by the coherent detection process in 1D (Schmitt et al., 1999). a) shows the distribution of scatterers, with the incoherent signal "imagining" giving a close approximation to the original distribution. However, coherent light, such as used in the OCT, produces a bad approximation c) with the envelope shown in d)

Even though speckle is usually considered to be a corrupting factor in OCT images, it can also provide some insight into the tissue composition. Most of the algorithms are focused on the idea of treating speckle as noise and removing it as much as possible without losing too much of the original information. Several different algorithms have been applied to the speckle removal in OCT images. Standard pre-processing techniques are mean and median filtering. Median filtering has an important edge preserving property. Wiener filter might be used to reduce the noise in the signal if a reasonably accurate assumption about the spectral content of the signal and the noise can be made in advance. Usually the requirement is that the filter has to be physically realisable resulting in the casual solution. Performance criteria used is minimum mean-square error. Common use of Wiener filter is in Wiener deconvolution. It will suppress frequencies that have poor signal to noise ratio. Even though the exact noise and signal spectra are not known in advance, reasonable spectra assumptions can often be made. For instance, in the image processing applications it can usually be assumed that most of the signal is contained in the lower frequencies while the noise has usually a flat frequency spectrum. Kulkarni et al. built a linear shift invariant model describing coherent specimen-light interactions in OCT (Kulkarni, Thomas and Izatt, 1997).

Based on this model, an iterative deconvolution algorithm was demonstrated for enhancing the sharpness of biological structures in OCT images. An example of algorithm developed for another purpose and then applied to the OCT field is CLEAN (Schmitt, 1998). An iterative point-deconvolution algorithm originally intended for use in radio astronomy, it was successfully used for restoring signal in OCT images and allowing observation of tissue morphology previously hidden under noise. Its performance degrades reasonably as the number of irresolvable scatterers increases. Derived deconvolution kernel is based on the theoretical point-spread function of an OCT scanner and depends on the properties of the imaging system and the scattering properties of the medium. It incorporates a modification based on inverse Wiener filter designed to reduce ripple artefacts in the images of tissue containing densely packed scatterers. The sticks method adapted from ultrasound imaging is based on fitting short linear segments to the boundaries of the vessels for example. Rotating kernel transformation has been applied to coronary images, based on matching edges in the image to a set of binary-valued two-dimensional square kernels. Undecimated wavelet filter transformation uses the logarithm of the magnitude to remove the uncorrelated speckle features at multiple scales. Logarithm scale is used since the multiplicative noise becomes additive noise in the logarithm scale. Another method that has been developed for this purpose simultaneously constrains the image magnitude to match a blurred version of the image in a least square sense. The constraint used is relative entropy called the I-divergence. Despeckling result of this method is shown in Figure 7.

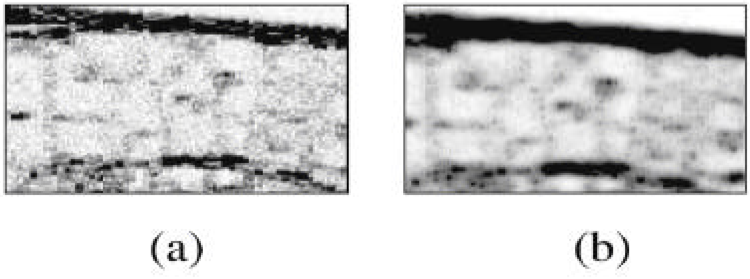


Figure 7 Comparison of the original and the despeckled image using I-divergence method (Drexler and Fujimoto, 2008)

Since in OCT images are usually formed from the envelope of the measured interference signal, computation of the absolute magnitude of the signal is a nonlinear process that destroys phase information. A speckle-reduction technique that works in the complex domain, zero-adjustment-procedure (ZAP), has been introduced (Yung, Lee and Schmitt, 1999). However, ZAP tends not only to reduce speckle but also to blur sharp edges between the image features. Anisotropic diffusion filtering in scale-space with the diffusion coefficient chosen in a way to encourage intraregion rather than interregion smoothing was presented (Perona and Malik, 1990). Scale-space is a kind of multiresolution image representation with certain feature preserving properties. Gilboa et al. demonstrated further improvements in denoising by extending the analysis from real to complex domain and combining the diffusion equation with the free Schrödinger equation (Gilboa, Sochen and Zeevi, 2004). Additional property is that the imaginary value can be used as a robust edge filter. These methods have been further extended by making the diffusion process truly anisotropic, in a sense that the diffusivity varies not only with the edge location but with the orientation as well (Joachim, 1999). The main advantage of this approach is that the diffusion rather acts along coherent structures and thus provides an efficient denoising method since noise has no preferred orientation. Fernandez et al. combined the complex diffusion and coherence-enhanced diffusion filtering and successfully applied it to the OCT data pre-processing as a step before the peak based image segmentation (Cabrera Fernández, Salinas and Puliafito, 2005).

6.1.4 Quantization

Quantization refers to mapping raw data to a series of fixed values. There are three reasons for signal quantization (Yu et al., 2001). The first reason relates to image compression, transmission and digital storage. Another purpose is to adapt the data for the human vision system and in this case subjective visual impression is more important than the absolute distortion. The third reason is for pixel level transformation. Four commonly used methods are equal-interval, equal-probability, minimum-variance and histogram hyperbolization. Equal-interval is a simple linear transformation from the original signal range to the new range. Equal-probability, or often referred to as histogram equalization, creates intervals, or bins, of variable sizes in a way that all bins contain equal number of pixels. Minimum-variance quantization creates the intervals so that the weighted sum of the variances of the intervals is minimized. Histogram hyperbolization adapts the image to the nonlinearity of the human vision system. The logarithm scale is usually used in OCT field as a mapping from the complex domain raw data to the final image. However, some detail will inevitably be lost in this process and more attention should be paid to the choice of the appropriate method. The simplest one is Direct Logarithm (DL) method. Truncation Logarithm (TL) method considers both the dynamic range determination and the noise present in the signal. TL method uses a fixed threshold to reduce the noise but it also degrades image quality. The main reason for choosing a logarithm-based method is to compress the dynamic range in respect to the exponential law that governs the attenuation of light in scattering materials. However, logarithm-based methods often suffer from poor contrast and loss of details in the image. Three quantization methods presented in (Yu et al., 2001) are Minimum Distortion (MD) and Truncated MD (TMD), Information Expansion (IE) and finally Maximum Entropy. MD method is based on the minimum distortion principle. MD method becomes equivalent to minimum-variance method when mean-square measure is used as a measure of distortion. TMD method can be used on data that contains large peaks. IE method is similar to histogram equalization except that it works on the raw data. The advantage lies in the fact that histogram equalization reduces the image entropy while entropy of the raw data remains unchanged during equalization. ME method is focused on preserving the hidden information in the raw data. A standard measure from the information theory that measures mutual dependability of two variables, mutual information, is used as a maximizing variable. When mutual information of data before and after quantization is maximal, loss of the information is minimal. Since the quantization function is deterministic, maximization of mutual information is equal to the maximization of the image entropy after the quantization and hence the name of the method.

6.2 Apoptosis and Pre-Apoptotic Monitoring

Apoptosis or “programmed cell death” is a highly organised form of cell death, that although essential for the maintenance of normal tissue homeostasis, has been implicated in a number of neurodegenerative diseases including Parkinson’s disease, Alzheimer’s disease and spinal muscular atrophy (Mattson, 2000), as well as in retinal neuronal pathologies including glaucoma, diabetic retinopathy (De Souza-Ramalho et al., 2009) and age-related macular degeneration (Dunaief et al., 2002). The quantification of cellular health is essential for benchmarking responses to experimental or therapeutic interventions.

The most recent advances in *in vivo* monitoring of cells and biological structures include development of confocal scanning laser ophthalmoscopy (cSLO) (Cordeiro et al., 2004; Schmitz-Valckenberg et al., 2008) and *in vivo* confocal neuroimaging (ICON) (Prilloff et al., 2010). These methods that still rely either on external ligands, for example the use of annexin V in oncology (Rottey et al., 2006; Rottey et al., 2007; Wong et al., 2008) and retinal research (Cordeiro et al., 2004; Reichstein et al., 2007; Schmitz-Valckenberg et al., 2008; Prilloff et al., 2010), or on the development of new contrast agents for the use in magnetic resonance imaging (MRI) (Yoo and Pagel, 2008; Terreno, Castelli and Aime, 2010). However, development of a non-invasive, non-toxic technique is key to the advancement of *in vivo* cell and tissue imaging.

A number of optical methods were developed to stage cells as they progress through apoptosis based on detailed knowledge of subcellular changes (Palade, 1953; Cereghetti and Scorrano, 2006). These ligand-based techniques are routinely used for histochemical staging of apoptosis including phosphatidylserine localisation in the outer plasma membrane (Vermes et al., 1995; Blankenberg et al., 1998), mitochondrial dysfunction(Green and Reed, 1998), activation of caspases (Thornberry, 1997, 1998), DNA fragmentation (Robertson, Orrenius and Zhivotovsky, 2000) and loss of membrane integrity. Yet while these technologies are robust for *in vitro* analyses, their use in *in vivo* studies can be challenging due to their potential toxicity, requirements for ligand delivery, and bias towards events that coincide with the commitment of a cell to death.

In view of these limitations the development of *in vivo* ligand-free approaches are of special interest. For example, apoptotic cells can be distinguished from healthy cells by flow cytometry as indicated by changes in forward scatter and side scatter properties (Vermes, Haanen and Reutelingsperger, 2000). Coherent anti-Stokes Raman scattering spectroscopy (CARS) also show promise but is limited by the need for a high numerical aperture and low signal to noise ratio (Wright et al., 2007; Balu et al., 2010). Currently, these methods are widely used for the analysis of cell cultures but are of limited use for monitoring of cellular events in living tissue.

Apoptosis is a cellular event involving many organelles including the mitochondria, endoplasmic reticulum, golgi and the nucleus, all of which can generate optical signal changes detectable by UHR-OCT. The mitochondria are key initiators of the apoptosis pathway. During apoptosis the mitochondrial membrane becomes “leaky” allowing the release of a number of pro apoptotic molecules including cytochrome c, leading to the activation of the caspase cascade and ultimately cell death (Frank et al., 2001; Cereghetti and Scorrano, 2006). Prior to this mitochondrial membrane dysfunction, neurons are subjected to prolonged periods of compromise during which dendrite pruning (McGowan, 2006) and remodelling occur prior to loss of the cell body. During this period of compromise, before the breakdown of the mitochondrial membrane, the mitochondria change morphology where they cease to be part of a continuous network, round off and become discrete (Karbowski and Youle, 2003). The dimensions of these discrete organelles are close to the wavelength of the light used in the 800 nm UHR-OCT system so the scattering coefficient will be maximal (Palade, 1953). This, in combination with the fact that mitochondria have a higher refractive index than the cytosol due to the high concentrations of proteins within them makes them ideal targets (Mourant et al., 1998) for the pre-apoptotic detection. OCT has been previously used to detect in vivo changes in the neuronal population of mouse cortex, thought to arise from light scattering by the mitochondria (Mourant et al., 1998), and other light scattering techniques have detected mitochondrial changes in in-vitro cells undergoing malignant change (Gourley et al., 2005).

The requirement for use of external ligands is overcome by recent developments in UHR-OCT suggesting that it may finally be possible to study localized light scattering within tissues and therefore map indices of cellular health to tissue structures. In particular, it has been shown that UHR-OCT is able to detect differences in light attenuation between apoptotic versus necrotic cells (van der Meer et al., 2010).

6.3 Machine Learning and Texture Analysis

Since the algorithms that are going to be described later in this thesis depend on the supplied training data to tune their parameters it is necessary to give a basic overview of the machine learning algorithms in general. Machine learning algorithms can be divided in two major classes of methods: supervised and unsupervised learning.

Supervised learning is a technique for learning a function based on training data, which consists of pairs of input objects, and desired outputs. The output of the function can be a continuous value, or a discrete class label. The task of the supervised learner is to predict the value of the function for any valid input object after having seen a number of training examples.

Unsupervised learning is a type of machine learning where manual labels of inputs are not given. The most common type of unsupervised learning is clustering. This classifies objects into different groups, so that the data in each subset share some common trait, usually proximity according to some defined distance measure. The task of classifying the data set into *k* clusters is often referred to as *k*-clustering.

It is difficult to give a comprehensive definition of what texture is. One approach is that a region in an image has a constant texture if a set of local statistics or other local properties of the picture function are constant, slowly varying, or approximately periodic (Chen, Pau and Wang, 1993). The goal of texture based classification is to produce a classification map of the input image where each uniform textured region is identified with the texture class it belongs to. Image texture is defined as a function of the spatial variation in pixel intensities. One immediate application of image texture is the recognition of homogeneous image regions using texture properties. In general, from each image, texture features are produced by applying various filters (mean, co-occurrence matrix, Laws', Haralick, Gabor, wavelets, etc.) of different window sizes and mapping every pixel from the original image into the created *n*-dimensional feature space. Another useful class of texture features are local binary patterns (LBP) that are especially suited for cases when the texture pattern is of interest and smaller intensity variations are to be ignored (Timo et al., 2002). LBP divides an image into cells. For each cell it looks at every pixel in that cell and assigns 0 or 1 to its neighbours, depending on whether they have a lower or higher intensity. In case of a typical 8-neighbourhood, an 8 bit number is produced. These are aggregated into a histogram, representing texture properties of a given cell.

Different methods are then used to separate classes in the feature space and using that information to classify new images pixelwise. Texture analysis is an important method that captures local information very well but it lacks global knowledge of the segmented structure which makes it sensitive to variations in the local appearance, such as introduced by speckle. Also, in case when structures that need to be segmented do not have a defining unique texture, as is the case with OCT layers, using contextual spatial information becomes necessary. It is important to mention that it is difficult to decide what feature set is optimal to use since the number of possible combinations of different feature types and window sizes is huge and it is computationally costly to use unnecessarily large set of features for the segmentation of new images. Thus, it is useful to perform a feature selection step in the training phase that will keep the best features from a possibly very large set of initial features. In this way the training phase of the algorithm will be lengthy, but once the optimal set of features has been found, segmentation of new images will proceed much faster. One method that can be used to perform this task, and was chosen in this work, is sequential forward search (Guyon and Elisseeff, 2003). It belongs to the class of supervised learning methods and works by iteratively expanding the optimal set by one starting from zero. This algorithm does not guarantee to give optimal set since choosing some of the features might give poor segmentation improvements individually, but they might work very well in combination. In most cases however, sequential forward search will give a good trade-off between computational cost of the training phase and the obtained reduced feature set which is usually close to the optimum.

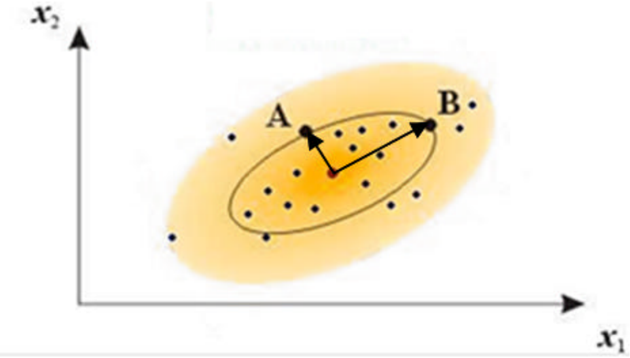


Figure 8 Mahalanobis distance in 2D. Points A and B have the same Mahalanobis distance from the centre

To perform the actual separation of different classes in feature space several methods can be used. The most common choice is to use Mahalanobis distance, neural networks or support vector machines. These methods will be described in more detail.

Mahalanobis distance differs from Euclidean distance in that it takes into account correlations of the data set. It can be intuitively described as a hyperellipsoid (ellipsoid in *n*-dimensional space) that best represents the set's probability distribution and is estimated by building the covariance matrix of the samples. The Mahalanobis distance is simply the distance of the test point from the centre of mass of class distribution divided by the width of the hyperellipsoid in the direction of the test point. An illustration of the Mahalanobis distance in 2D case is shown in Figure 8.

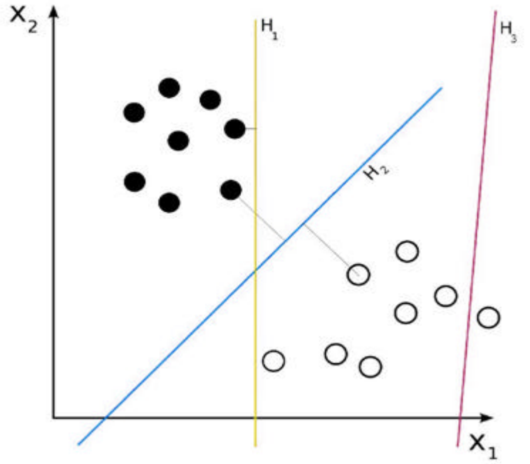


Figure 9 SVM classification. Three possible division lines are shown: H1, H2 and H3. SVM produces H3 as the optimal separation.

A Neural network (NN) is a mathematical model or computational model based on biological neural networks. It consists of an interconnected group of artificial neurons and processes information where each connection between neurons has a weight, with the weights modulating the value across the connection. The training phase is performed to modify the weights until the network implements a desired function. Once training has completed, the network can be applied to data that was not part of the training set.

A Support vector machine (SVM) is a classification technique that not only separates data into different classes, but also leaves a maximal margin between those classes. Because of their maximum margin separation, SVMs can obtain better classification results, which minimize the risk of misclassification for data that is not shown in the training set. The maximal margin principle is shown on Figure 9 with the line H2 representing the maximal margin division. A useful technique that can be applied if the desired separation surface is of higher order (for example second or third order polynomial) maps the problem to a higher dimensional space and uses a hyperplane to mark the boundary. This hyperplane will correspond to the higher order surface in the original space. How to construct a soft margin hyperplane in cases when training data cannot be separated without errors is demonstrated in (Cortes and Vapnik, 1995).

With NNs, training is done repeatedly to obtain iteratively better results for classification and it is difficult to determine when to stop training or when the best network is generated. On the other hand, SVMs handle the entire training data set simultaneously and provide the optimal solution when the training is complete (Fan-Yin, 2005).

A large number of texture based segmentation methods have been published. Some of the representative and relevant approaches will be covered.

A six-stimulus basis for stochastic texture perception was presented in (Jan-Mark and Arnold, 2005). Fragmentation of the scene by a chaotic process (adding new objects to the scene randomly) causes the spatial scene statistics to conform to a Weibull-distribution, while the parameters of the Weibull distribution characterize the spatial structure of uniform stochastic textures of many different origins completely. The perceptual significance of the Weibull parameters was reported with perceptual parameters being: regularity, coarseness, contrast, roughness, and directionality.

Viola et al. achieved high speed texture segmentation using effective image representation (integral image), feature selection algorithm that filters thousands of simple features based on AdaBoost and a cascade approach to ignore the areas of the image which are not likely to contain objects of interest early in the process (Viola and Jones, 2001).

Gossage et al. presented good texture segmentation based results on mouse and fat tissue, as well as healthy and diseased mouse lung tissue (Gossage et al., 2006). They used energy, entropy, correlation, local homogeneity and inertia features. However, it is important to note that the speckle pattern differences in the differentiated tissue were visible by naked eye which is not the case with retinal layers.

Texture analysis has been successfully applied to resolve collagen tissue with and without endothelial cells and shows that tissues that contain scatterers of different sizes and distributions can be differentiated with reasonable accuracy using speckle information only (Kirk et al., 2006). However, not all retinal layers contain enough speckle information for texture analysis based segmentation to be successful alone. Results obtained with texture analysis so far are presented later on.

An interesting approach to the problem of segmentation has been presented in (Ojala and Pietikäinen, 1997). By splitting an image into small consistent regions and then merging them into larger areas, a higher level of spatial consistency has been achieved. Even though they performed the analysis on images with much more apparent difference in texture than is the case with OCT data, it is still a highly relevant approach.

6.4 Numerical Optimization Algorithms

Optimization algorithms are used to solve problems that are either impossible or impractical to solve analytically. They are essential building blocks of registration and nontrivial segmentation methods and thus have to be studied in detail. In this work they are used for minimizing a distance function in image registration algorithms and for minimizing an objective function that defines success in fitting an adaptive template or model to data.

An important optimization example is the least square problem which is used to model numerical data obtained from observations by adjusting the parameters of a model to get an optimal fit to the data. The best fit is that instance of the model for which the sum of the squared residuals has the minimal value, the residual being the difference between an observed value and the value given by the model. All of the optimization methods share the same principle of iteratively minimizing (or maximizing) the objective function. Since both data registration and segmentation approaches often fall into this domain it is important to give an overview of some of the many existing optimization methods. The number of individual optimization methods is huge so it is important to give a review of the different categories and which might be of interest to the practical applications presented in this work.

6.4.1 Classification of Optimization Methods

Objective functions can differ significantly in its nature. They can be linear, nonlinear or convex. A convex set is a set for which any line segment connecting any two points in the set lies in that set. A convex function is a function whose domain is a convex set and for any two points in this domain, its graph lies below the line segment. In respect to smoothness, functions can be differentiable or nondifferentiable. If the function is both smooth and convex, then its graph is bowl shaped. If the function satisfies these conditions, optimization problem becomes much easier. Most of the algorithms guarantee convergence to a stationary point in finite time but for convex, smooth functions it is also certain that this local optimum is global optimum as well.

Considering the reach of the optimization in the domain space, algorithms can be defined as local or global. Most of the algorithms focus on finding a local minimum, a point for which all the neighbouring points hold higher values. For the neighbourhood to be defined, the domain space has to be topological. Global algorithms try to find a global solution, a minimum in the whole domain space of the objective function. In the most general case it is impossible to guarantee that the global solution has been found. It is also very difficult to be certain that the minimum that has been found really is a global one. However, global optimization methods often demand certain type of objective function in order to make the search reasonably effective. In a special case, when the objective function is convex, local solution is equal to the global solution (Nocedal and Wright, 1999).

One important difference that can be made between optimization methods is whether they put constraints on the variables or not. Unconstrained optimization deal with problems that require no constraints on the variables or the constraints on the variables arise naturally and are usually safe to dsregard. Another common case when unconstrained problems appear is when constrained problems are reduced to unconstrained problems by replacing the constraints with penalizing terms that are added to the objective function. In general, unconstrained problems are easier to solve. Constrained problems arise when dealing with models that contain explicit constraints on the variables. The set of constraints can either be equalities or inequalities that again can be simple bounds on the variables or complex inequalities among the variables often represented by a matrix. Set of constraints which are active in respect to the current point is called Active set. At each iteration, the active set class of constrained optimization algorithms searches for and removes inactive constraints and then approximately solves the equality problem defined by the active set.

Optimization methods can also be separated based on the domain of values that the variables can hold, continuous or discrete. For example, in some problems variables can hold only integer values. If the discrete domain fact is ignored and the problem is optimized by allowing variables to hold real values and then rounding up is performed at the end, solution is by no means guaranteed to be optimal. This problem is properly handled by adding a new constraint to the set of constraints and forcing the variable set to hold discrete values (Nocedal and Wright, 1999). In general, continuous problems are easier to solve because the smoothness of the function allows for use of objective and constraint information at some point to deduce information about the point neighbourhood as well.

Depending on the certainty of the model definition, optimization algorithms can be deterministic or stochastic. Deterministic algorithms are applied to completely defined models, when no random behaviour is expected. In many situations behaviour of the objective function cannot be completely predicted. However, probabilities can be usually applied to different scenarios. In many economic and financial planning models this situation occurs. In this problem domain best solution is to use stochastic optimization methods. Stochastic optimization algorithms can be also used in solving deterministic problems if difficult nonlinear, constrained objective function has to be optimized and some benefits are expected from the random behaviour of the optimization algorithm in hope of avoiding local minima. It is important to define certain concepts regarding stochastic optimization since they will be used later on when describing some stochastic algorithms in more detail. Monte Carlo methods are a broad term that refers to a class of computational algorithms that rely on repeated random sampling to compute their results and are usually used when it is infeasible or impossible to compute an exact result with a deterministic algorithm. Stochastic optimization methods are examples of Monte Carlo methods and as such are based on random walks. The algorithm moves around a marker in a multi-dimensional space, tending to move in directions that lead to a lower function, but occasionally moving against the gradient. A Markov chain is a stochastic process with the Markov property, which means that given the present state, future states are independent of the past states. In other words, the description of the present state fully captures all the information that may influence the future evolution of the process while the future states are reached through a probabilistic process instead of a deterministic one. At each step the system may change its state from the current state to another state through any of the applicable transitions, or remain in the same state, according to a certain probability distribution of the transition probabilities. An example of a Markov chain is a simple random walk where the state space is a set of vertices of a graph and the transition steps involve moving to any of the neighbours of the current vertex with equal probability. Markov chain Monte Carlo (MCMC) algorithms are used for sampling from probability distributions based on constructing a Markov chain that has the desired distribution as its equilibrium distribution. Equilibrium distribution is a vector for which holds that a probability of state S is equal to the sum of multiplies of probabilities of all states that S is connected to and their transition probabilities. The state of the chain after a large number of steps is then used as a sample from the desired distribution with the quality of the sample improving as a function of the number of steps.

Another distinction that can be made among the optimization algorithms is how large are the sets of variables that they can handle effectively. Some algorithms can be defined as medium-scale and are not effective when used with large variable sets, thousands or hundreds of thousands of variables. Large-scale algorithms should be used in such cases.

In respect to the iterative approach that is taken, optimization methods fall into Line search or Trust region classes. Line search algorithms determine first the direction of the step to be taken and will then determine the optimal size of the step. On the other hand, trust region algorithms first define an initial area of search and will collapse or expand it depending on how well this area approximates the model function, which is usually a quadratic.

Optimization algorithms use different strategies for finding the optimal solution. Some will rely only on the objective function value at a certain point; others will use first or second derivatives, while some will even use the information from previous iterations. However, it should be common to all the good optimization methods that they are robust, efficient and accurate (Nocedal and Wright, 1999). It is usually difficult to achieve all of those goals since they may conflict so a trade-off is usually sought. In the following some of the most popular optimization methods will be shortly described, as well as those, which were considered as potentially useful to this work.

6.4.2 Deterministic Optimization

Gradient descent is one of the simplest and most intuitive optimization methods (Cauchy, 1847). If the function is differentiable it will decrease fastest in the direction of the negative gradient, so to find a local minima, steps proportional to the negative of the gradient of the function, or to the approximate gradient, are taken at the current point. However there are two significant weaknesses of gradient descent. Firstly, it may take the algorithm many iterations before it converges towards a local minimum, in cases the curvature in different directions varies a lot. Secondly, finding the optimal step length per iteration can be time-consuming. Conversely, using a fixed step length can yield poor results.

Newton's method is based on General Newton’s method for finding roots of equations in one or more dimensions by iteratively finding roots of the tangent at the current point. It can be easily used to find local maxima or minima of functions by noticing that if any real number is a stationary point of a function, then that number is also the root of the derivative of the function, and therefore it is possible to find stationary points by applying General Newton's method to the derivative of the function. The geometric interpretation of Newton's method is that at each iteration function is approximated by a quadratic function around the point, and then a step is taken towards the minimum of that quadratic function. If the function happens to be a quadratic, then the exact minimum is found in only one step. Newton's method converges much faster towards a local minimum than the gradient descent. However, finding the inverse of the Hessian matrix required by the algorithm is an expensive operation.

Sequential Quadratic Programming (SQP) is a constrained optimization method that has the basic idea analogous to Newton's method for unconstrained minimization (Pshenichny and Danilin, 1983). At each step, a local model of the optimization problem is constructed and solved, yielding a step towards the solution of the original problem. However, instead of using Taylor expansion of the objective function as a quadratic approximation, Lagrangian is used instead, which gives better update results of both Lagrange multipliers and optimizing variable. Lagrange multipliers are used to force the optimization procedure to stay close to the given constraints.

Unlike Newton’s method, in Quasi-Newton class of methods the Hessian matrix of second derivatives of the function to be minimized does not need to be computed at any stage. Hessian is updated by analyzing successive gradient vectors instead. Quasi-Newton methods are a generalization of the secant method to find the root of the first derivative to multidimensional problems. The secant method is a root-finding (function zeros) algorithm that uses a succession of roots of secant lines to better approximate a root of the function. In a multi-dimensional space, the secant equation is under-determined and members of a class of quasi-Newton methods differ in how they constrain the solution. The Broyden-Fletcher-Goldfarb-Shanno (BFGS) method for solving unconstrained nonlinear optimization problems is one of the most successful members of this class (Broyden, 1970).

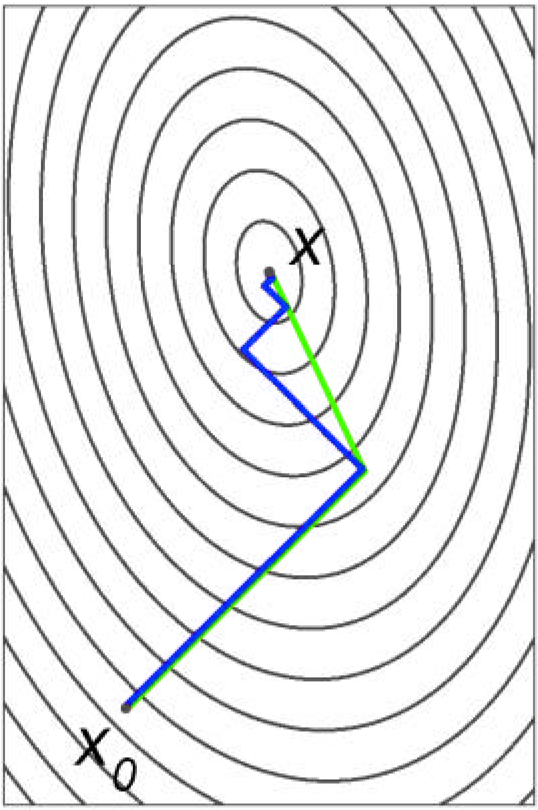
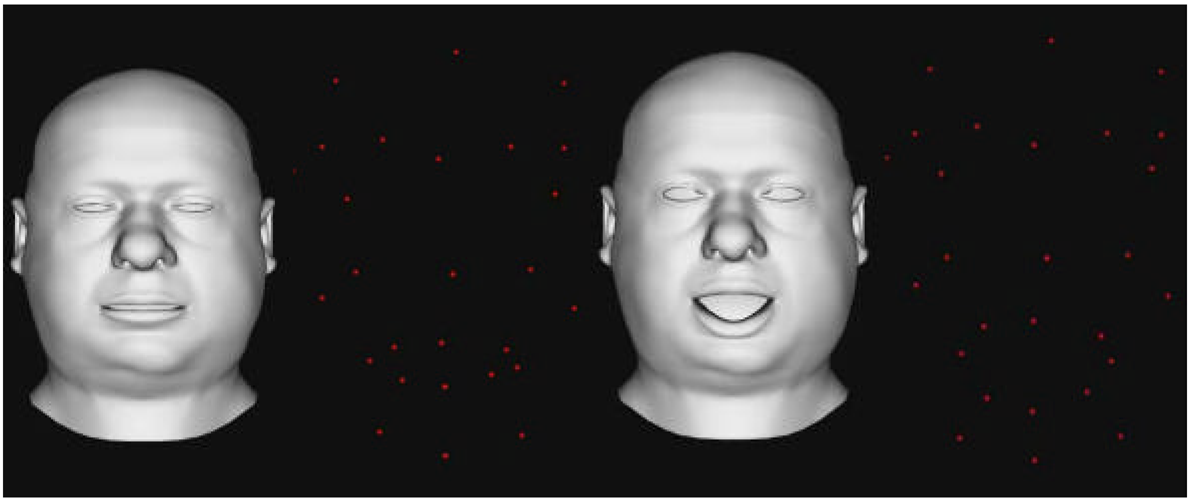


Figure 10 Conjugate gradient (green) compared to the steepest descent method (blue)

The Levenberg–Marquardt Algorithm (LMA) provides a numerical solution to the problem of minimizing a function, generally nonlinear, over a space of parameters of the function (Levenberg, 1944). These minimization problems arise often in least squares curve fitting and nonlinear programming. LMA interpolates between the Gauss-Newton algorithm (GNA), which is a variation of the Newton's method that can only be used to minimize a sum of squared function values, and the method of gradient descent. This shifting is achieved by using a dynamic damping parameter that sets LMA closer to gradient descent or GNA. LMA is more robust than GNA, which means that in many cases it finds a solution even if it starts far off the final minimum. On the other hand, for well-behaved functions and reasonable starting parameters, LMA tends to be a bit slower than GNA. This algorithm was used for stack registration, Chapter 7.1.

Conjugate gradient method is an algorithm for the numerical solution of particular systems of linear equations, namely those whose matrix is symmetric and positive definite, but it has broader use since quadratic can be used on the quadratic form as well. It can be also used to solve unconstrained optimization problems such as energy minimization. An example of the convergence rates of conjugate gradient and steepest descent is given in Figure 10 with the green line showing the steps of steepest descent method, while the red line shows the rapid conjugate gradient convergence. It is based on the notion of conjugate vectors, vectors that are orthogonal in respect to some matrix. This property allows an easy extraction of vector coefficients from the representation of the linear equation in the basis of conjugate vectors. If the conjugate vectors are chosen carefully, then usually not all of them are needed to obtain a good approximation to the solution. That is why it is useful to regard the conjugate gradient method as an iterative method. It also allows solving of large systems where the direct method would take too much time. This suggests taking the first basis vector to be the gradient of the function at the starting point, while the other vectors in the basis will be conjugate to the gradient. In other words, the conjugate gradient method can be considered as a steepest descent method with an extra constraint that all steps have to be conjugate, which leads to the faster convergence. One illustrative example where conjugate gradient method was applied on a real-life problem was minimizing the distance between motion capture data and the Dylan Thomas model to be driven by that data. This work was done as a part of my side project at Cardiff University Computer Science department and even though not directly related to this PhD work, it demonstrates well the importance of the optimization algorithms. In Figure 11 four images are shown captured from the Dylan Thomas video sequence. On the right is a set of points extracted from the positions of the markers on the actor's face (not shown). On the left is a Dylan Thomas expression generated to best match the marker's position. This is done by combining the set of static face expressions and performing the optimization to get the set of expression weights that will produce the minimum distance between the model and the actual marker positions.



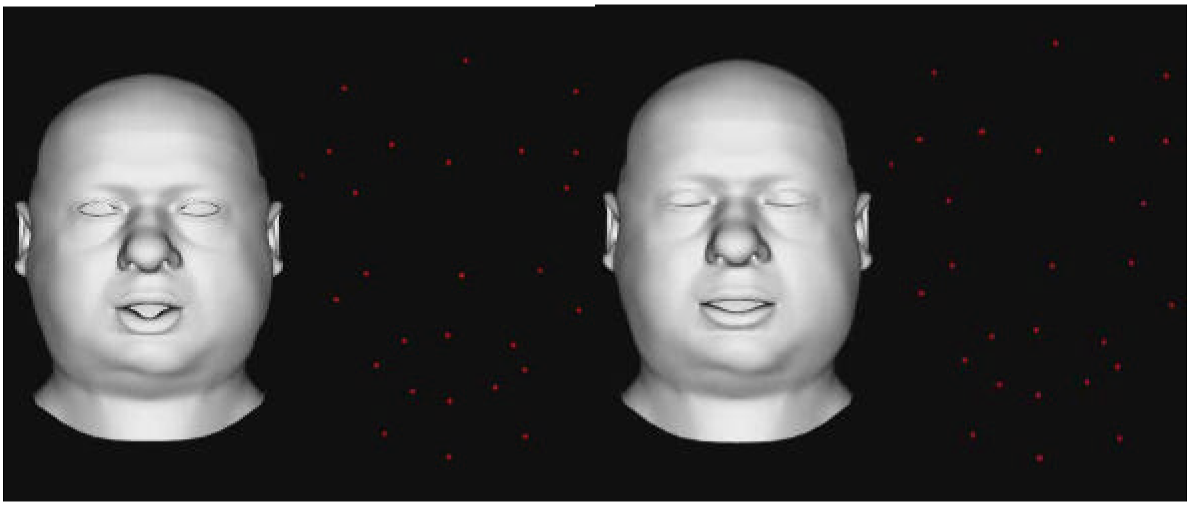


Figure 11 Conjugate gradient method applied to driving a Dylan Thomas model. Four different example expressions are produced by extracting the markers from an actor's face (right) an applying them to the model (left)

Nelder-Mead method or downhill simplex method uses the concept of a simplex, which is a polytope of N + 1 vertices in N dimensions; a line segment on a line, a triangle on a plane, a tetrahedron in three-dimensional space etc. The method approximately finds a local optimum to a problem with N variables when the objective function varies smoothly. As each run of the simulation is expensive, it is important to make a good decision when placing a starting point. The Nelder-Mead algorithm generates a new test position by extrapolating the behaviour of the objective function measured at each test point arranged as a simplex. At each iteration, the algorithm chooses to replace one of these test points with the new test point. The simplest option is to replace the worst point with a point reflected through the centroid of the remaining N points. If this point is better than the best current point, then it can be stretched out exponentially along that line. On the other hand, if that new point is not much better than the previous value, then the algorithm is stepping across a valley, so the simplex shrinks towards the best point.

The Simplex algorithm is a popular algorithm for numerically solving linear programming problems. Linear problem can be represented in geometric terms as a closed convex polytope, defined by intersecting a number of half-spaces in n-dimensional Euclidean space; each half-space is the area which lies on one side of a hyperplane. In this case it can be shown that the optimum value is attained on the boundary of the polytope. Methods for finding this optimum point on the polytope work in several ways: some attempt to improve a possible point by moving through the interior of the polytope (interior point methods), while others start and remain on the boundary when searching for an optimum. Simplex algorithm follows the latter methodology. The idea is to move along the facets of the polytope, point to point, in search for the optimum. However, unless the optimum occurs on an edge or face that is parallel to one of the hyperplanes, the optimum will be unique and occur at a vertex of the polytope. If the optimum is found on an edge or a face that is parallel to one of the hyperplanes, then the optimum is not unique and can be obtained at any point on that edge or face. Since the simplex algorithm is concerned only with finding a single optimal point, it is possible to look solely at moves skirting the edge of a simplex, ignoring its interior. Starting at some vertex of the polytope, at every iteration an adjacent vertex is chosen so that the value of the objective function does not decrease. If no such vertex exists, a solution to the problem has been found. However, often such an adjacent vertex is nonunique, and a pivot rule must be specified to determine which vertex to pick next. Various pivot rules exist to solve this problem but there is no need to go into more detail.

Interior point methods (or barrier methods) are a certain class of algorithms to solve linear and nonlinear convex optimization problems. These algorithms have been inspired by Karmarkar's algorithm, developed by Narendra Karmarkar for linear programming (Karmarkar, 1984). The basic principle of the method consists of using a self-concordant barrier function used to encode the convex set. Contrary to the simplex method described earlier, it reaches an optimal solution by traversing the interior of the feasible region. As mentioned earlier, any convex optimization problem can be transformed into minimizing (or maximizing) a linear function over a convex set.

The primal-dual schema is a discrete optimization technique that is well known in the combinatorial optimization field. It started as a very general technique for solving linear programming problems but it was realized that it could be a very powerful tool for deriving approximation algorithms to problems of linear integer programming as well. An interesting application of this method is Markov random field optimization problem that works on a discrete set of objects, all of which are vertices in a graph while the edges of this graph encode the objects' relationships. One label from the set of labels has to be assigned to each object and a price has to be paid when assigning a label to the object and to the pair of objects. The goal is to minimize the sum of those prices. The Fast-PD is an efficient approach for solving the primal-dual problem in Markov random field optimization by ensuring that the number of augmentations decreases significantly over time (Komodakis, Tziritas and Paragios, 2008).

6.4.3 Stochastic Optimization

So far only standard, deterministic and relatively simple optimization algorithms have been presented. In the following stochastic and global optimization algorithms will be reviewed, since it is often important to escape local minima when dealing with difficult objective functions. These algorithms are often imaginative and draw the fundamental principles of operation from the natural phenomena. An unfortunate drawback of many of these algorithms is that they tend to be relatively slow.

The Simultaneous Perturbation Stochastic Approximation (SPSA) algorithm is capable of an effective loss-function minimization when only, often noisy measurements, of the loss function are available. No measurements of the gradient of the loss function are assumed available. SPSA successfully addresses one of the major shortcomings of these finite-difference stochastic approximation algorithms that take derivative steps in each direction by significantly reducing the number of measurements required in many multivariate problems of practical interest. In theory, the execution speed up in respect to the standard stochastic approximation methods is equal to the number of the dimensions. SPSA has also been extended to the second-order approximation based on estimating both the loss function gradient and inverse Hessian matrix at each iteration. The aim of this approach is to emulate the acceleration properties associated with deterministic algorithms of Newton form, particularly in the terminal phase where the first-order SPSA algorithm slows down in its convergence. The second-order SPSA algorithm requires only five loss function measurements at each iteration, independent of the problem dimension (Spall, 1997).

Wang (Wang and Spall, 1998) presented a stochastic approximation algorithm based on the penalty function method and a simultaneous perturbation gradient estimate for solving stochastic optimization problems with general inequality constraints based on using a penalty function. A very general convergence result for the proposed algorithm has been presented, however some of the requirements on the problem are difficult to satisfy in practice.

Simulated Annealing (SA) is a global optimization method, designed for locating a good approximation to the global optimum of a given function in a large search space. It is often used when the search space is discrete. The name and inspiration come from the annealing technique in metallurgy, a technique involving heating and controlled cooling of a material to increase the size of its crystals and reduce their defects. The heat causes the atoms to become unstuck from their initial positions and wander randomly through states of higher energy, while the slow cooling provides more chance for finding configurations with lower internal energy than the initial one. By analogy with this physical process, each step of the SA algorithm replaces the current solution by a random near solution, chosen with a probability that depends on the difference between the corresponding function values and a global temperature parameter T that is gradually decreased during the process. The dependency is such that the current solution changes almost randomly when T is large, but behaves increasingly "downhill" as T approaches zero. The allowance for "uphill" moves saves the method from becoming stuck at local minima.

Quantum Annealing (QA) is a general method for finding the global minimum over a given set of candidate solutions by a process analogous to quantum fluctuations. It is used mainly for problems where the search space is discrete and with many local minima. In quantum annealing, a current state is randomly replaced by a randomly selected neighbour state if the latter has a lower energy (value of the objective function). The process is controlled by the tunnelling field strength, a parameter that determines the extent of the neighbourhood of states explored by the method. The tunnelling field starts high, so that the neighbourhood extends over the whole search space and is slowly reduced through the computation, until the neighbourhood shrinks to those few states that differ minimally from the current state.

Parallel Tempering, also known as replica exchange, is a simulation method aimed at improving the dynamic properties of Monte Carlo method simulations and Markov chain Monte Carlo sampling methods. For example energy updating simulation based on the temperature that runs as a single Monte Carlo simulation can suffer from critical slowing down when the system is highly correlated. This can be improved by running for example two parallel simulations with slightly different temperatures so that the generated distributions overlap. Another way to interpret this overlap is to say that system configurations sampled at one temperature are likely to appear during a simulation at another temperature. At a given Monte Carlo step, the global system can be updated by swapping the configuration of the two systems, or alternatively swapping the two temperatures. If temperatures and number of systems are carefully chosen it is possible to achieve an improvement in the mixing properties of a set of Monte Carlo simulations that exceeds the extra computational cost of running parallel simulations. Parallel tempering method can also be used as a super simulated annealing that does not need restart, since a system at high temperature can provide new local optimizers to a system at low temperature, allowing tunnelling between metastable states and improving convergence to a global optimum.

Tabu search enhances the performance of a search method by using memory structures. Once a potential solution has been determined, it is marked as tabu so that the algorithm would not visit that possibility repeatedly.

A Genetic Algorithm (GA) is a global, heuristic search technique used in computing to find exact or approximate solutions to optimization and search problems. Genetic algorithms are a particular class of evolutionary algorithms that use techniques inspired by evolutionary biology such as inheritance, mutation, selection, and crossover.

The Ant Colony Optimization (ACO) algorithm is a probabilistic technique for solving computational problems that can be reduced to finding good paths through graphs. They are inspired by the behaviour of ants when finding paths from the colony to food. In the real world, ants initially wander randomly and upon finding food return to their colony while laying down pheromone trails. If other ants find such a path, they are likely not to keep travelling at random, but to instead follow the trail, returning and reinforcing it if they eventually find food. ACO is basically a computer simulation of that behaviour.

Particle swarm optimization is a stochastic, population-based, evolutionary optimization algorithm based on the principle of social optimization. The communication structure or social network is defined, assigning neighbours to each individual to interact with. A population of individuals is defined as random guesses when the problem solution is initialized. These individuals are also known as the particles, hence the name particle swarm. The particles iteratively evaluate the fitness of the candidate solutions and remember the location where they had their best success. They also make this information available to their neighbours. Movements through the search space are guided by these successes, with the population usually converging to a problem solution better than that of a non-swarm approach using the same methods.

Differential Evolution (DE) is an optimization technique similar to GA but adapted to continuous problems. Floating-point instead of bit-string encoding and arithmetic vector operations instead of logical ones are used. These recasts have changed Genetic Annealing from a combinatorial into a continuous optimizer. The crucial idea behind DE is a scheme for generating trial parameter vectors by adding the weighted difference between two population vectors to a third vector.

Harmony search is a music phenomenon-inspired algorithm that was further developed by considering ensemble among music players. It conceptualizes a group of musicians together trying to find better state of harmony, where each player produces a sound based on one of the three operations: random selection, memory consideration, and pitch adjustment.

Memetic algorithm, similar to genetic algorithms, represents one of the recent growing areas of research in evolutionary computation. It is widely used as a synergy of evolutionary or any population-based approach with separate individual learning or local improvement procedures for problem search.

Bees algorithm is a population-based search algorithm that mimics the food foraging behaviour of swarms of honey bees. In its basic version, the algorithm performs the neighbourhood search combined with the random search and it can be used for both combinatorial and functional optimisation.

6.5 Statistical Model Building for Registration and Segmentation

Using a statistical model based on the training data is a potentially effective tool for both segmentation and registration (Cootes and C.J.Taylor, 2004). Its main advantage is that knowledge of the problem can be used to resolve the confusion caused by structural complexity, provide tolerance to noisy or missing data, and provide a means of labelling the recovered structures. The idea is to apply the knowledge of expected shapes of structures, their spatial relationships, and their texture appearance to restrict the automated system to plausible interpretations. To be useful, a model needs to be specific, capable of representing only legal examples of the modelled object. Statistical models can reproduce specific patterns of variability in shape and texture by analysing the variations in shape over the training set. It is difficult to achieve this whilst allowing for natural variability and thus it is essential what features will be supplied from the training set to the model building phase. The key step of the statistical model training phase is the dimensionality reduction on the large set of features of the training data set. The reason for dimensionality reduction is to reduce the computational cost of the optimization method that is actually used to apply the model to the real data. The idea behind this concept is to find statistical dependencies between the produced features and reduce the dimensionality of the space by selecting only a certain number of the most prominent directions in the data set, the most important eigenvectors. It is important to explain in more detail techniques that are used to perform this step.

6.5.1 Dimensionality Reduction

Principal component analysis (PCA) is the standard vector space transform technique used to reduce multidimensional data sets to lower dimensions for analysis. It works by calculating the eigenvalue decomposition of a data covariance matrix or singular value decomposition of a data matrix. Usually, a relatively small number of eigenvectors with greatest eigenvalues can describe the original data well. A technique similar to PCA is Local Discriminant Basis (LDB) algorithm, a powerful algorithmic framework that was developed as a technique for analyzing object classification problems. The main advantage of LDB over other similar techniques such as PCA is its lower computational cost (Hazaveh and Raahemifar, 2003). An adapted waveform framework called best basis algorithm had been developed prior to the development of LDB, mainly for signal compression problems. However, PCA and LDB work well on linear data but do not model well the non-linearities or discontinuities in the data. There are also non-linear dimensionality reduction methods, such as the constraint shape space point distribution model that automatically clusters the data into a combination of Gaussian distributions.

6.5.2 Model Types

Two main classes of statistical models will be briefly discussed. An Active Shape Model (ASM) deforms a shape model to describe the location of structures in a target image. It works by iteratively finding a new best match for each point of the model, updating the parameters to best fit the model to the new found points and repeating this cycle until convergence. These models can be combined with more complex texture feature models into efficient discriminative models (Cristinacce and Cootes, 2007). An Active Appearance Model (AAM) manipulates a model capable of synthesising new images of the object of interest by finding the model parameters which generate a synthetic image as close as possible to the target image (Cootes and C.J.Taylor, 2004).

Statistical shape models usually work by analysing a set of landmark points. However, it is often desirable to have smooth and invertible deformation fields, or diffeomorphisms. Unfortunately it is difficult to interpolate the deformation field between the points and ensure that the resulting mapping is diffeomorphic. One effective way of representing diffeomorphic deformation fields is by composing simple diffeomorphisms and using those to build a model (Cootes et al., 2008).

Statistical models have not been applied to the OCT segmentation problem yet, so that would be a novel approach to the problem. Additionally, constructing an OCT model is a novel, specific task and thus might yield not only specific but also general contribution to the statistical model building that could be related to some other problem.

6.6 Image Registration

Image registration refers to aligning source image to the target image in order to get the best match. Registration is necessary in order to be able to compare or integrate the data obtained from different measurements. Registration can differ based on the transformations used that can be either affine: combination of translation, rotation, scaling and shear. Affine transformations are global in nature, thus not being able to model local deformations which often need to be corrected in medical imaging. The second category includes elastic transformations that allow local warping of the image features, thus providing support for local deformations.

6.6.1 Classification of Registration Algorithms

Registration algorithms can be divided into two main groups: area based methods and feature based methods. For area based image registration methods, the algorithm looks at the structure of the image as a whole by using for example correlation metrics, Fourier properties and other means of structural analysis. Feature based methods, instead of looking at the overall structure of images, fine tune their mappings to the correlation of image features: lines, curves, points, line intersections, boundaries and others.

Another useful classification is between intermodal and multimodal registration algorithms. Intermodal registration algorithms are intended to register images acquired using the same kind of imaging device, while multimodal registration algorithms are capable of registering images acquired using different imaging devices. There are several examples of multimodal registration in the medical imaging field. Registration of brain CT/MRI images is the most notable one. In the OCT field registering OCT volume to a scanning laser ophthalmoscope (SLO) fundus image is an interesting approach.

Registration can be mathematically expressed as using one of the optimization algorithms to minimize the distance function between the source and the target. Several distance metrics are used in practice. The most common ones are cross-correlation, mean-square difference and mutual information. Mutual information is a measure of mutual dependency of two variables and thus suited for multimodal registration while the other two metrics are usually used for intermodal registration.

Most of the image registration methods operate in the spatial domain, using features, structures, and textures as matching criteria. This is also the most intuitive approach since in the spatial domain images look as the human eye might perceive them. However, there is a class of algorithms that use the properties of the frequency-domain to directly determine shifts between two images. By applying the Phase correlation method to a pair of overlapping images, it is possible to use the produced image which contains a single peak as the relative translation between the two images. Unlike many spatial-domain algorithms, the phase correlation method is resilient to noise, occlusions, and other defects typical of medical images. Additionally, the phase correlation uses the Fast Fourier transform to compute the cross-correlation between the two images, resulting in large performance gains. This method can be also extended to determine affine rotation and scaling between two images by first converting the images to log-polar coordinates. Due to properties of the Fourier transform, the rotation and scaling parameters can be determined invariant to the translation. This single feature makes phase-correlation methods highly attractive when compared to typical spatial methods, which must determine rotation, scaling, and translation simultaneously, often at the cost of reduced precision in all three.

6.6.2 Registration Metric

In the registration process, the metric typically compares intensity values in the fixed image against the corresponding values in the transformed moving image. When a point is mapped from one space to another by a transform, it will in general be mapped to a non-grid position. Therefore, interpolation is required to evaluate the image intensity at the mapped position (Ibanez et al., 2005). The simplest interpolation technique is nearest neighbour interpolation which uses the intensity of the nearest grid position. Linear interpolation assumes that intensity varies linearly between grid positions and unlike nearest neighbour interpolation, the interpolated intensity is spatially continuous. However, the intensity gradient is still discontinuous at grid positions. B-spline interpolation uses spline coefficients that are computed using recursive filtering. Sinc window interpolator corrects the well known signal processing phenomena which states that the process of sampling a spatial function using a periodic discrete grid results in a replication of the spectrum of that signal in the frequency domain. It can be shown that recovering the continuous signal from the discrete sampling is equivalent to the removal of the replicated spectra in the frequency domain. This can be done by multiplying the spectra with a box function that will set to zero all the frequencies above the highest frequency in the original signal which is equivalent to convolving the spatial discrete signal with a sinc function (sin(x)/x). The sinc function has infinite support, which of course in practice cannot really be implemented so the sinc is usually truncated by multiplying it with a window function.

6.6.3 Pyramidal Approach

Performing image registration using a multi-resolution pyramidal approach is widely used to improve speed, accuracy and robustness. The basic idea is to perform the registration at a coarse scale where the images have fewer pixels. The spatial mapping determined at the coarse level is then used to initialize registration at the next finer scale and the process is repeated until it reaches the finest possible scale. This coarse-to-fine strategy in general greatly improves the registration success rate and also increases robustness by eliminating local optima at coarser scales (Ibanez et al., 2005). Some methods use arbitrary coefficients for multi-resolution pyramid levels while using scale-space instead can be advantageous because of its image feature preserving properties.

6.6.4 Model-based Approach

Another approach that can be taken is registering a geometrical model to an image. Using this approach, a geometrical model is built first and a number of parameters are identified in the model. This same approach can be considered as a segmentation technique, since once the model has been optimally superimposed on the image, pixels can be labelled according to their associations with specific parts of the model. One of the main advantages of this approach is that, as opposed to image-to image registration, it actually provides insight into the anatomical structure contained in the image (Ibanez et al., 2005). This shows well how interconnected registration and segmentation problems are.

6.6.5 Registration of OCT Data

Registration in respect to the OCT in general refers to the registration of the stack that makes up the imaged volume. This is a more difficult task than a simple registration between the two images since it is important to perform this process in regard to the stack as a whole. Otherwise, if registered sequentially, serious errors can result as a consequence of the small errors that accumulate during the slice to slice registration. Another important issue is that the slices might actually overlap at some parts while some parts of the retina might end up being missed. The reason for this is the eye motion during the scan procedure that can occur in three degrees of freedom, including the rotation along the z axis. Thus, the ideal registration method would take all these facts into account and would correct for those transformations by applying reverse transformations and interpolating between the non-scanned areas. One approach would be to use a fundus image without the motion artefacts as a reference for the OCT stack registration. In the following, several methods will be reviewed that are specific to the OCT field or are relevant to this work in some other respect.

An automatic sub pixel registration algorithm that minimizes the mean square intensity difference between a reference and a test data set, which can be either images or volumes was presented in (Thevenaz and Unser, 1996). It uses an explicit spline representation of the images in conjunction with spline processing, and is based on a pyramid approach. The minimization is performed according to a fast, custom variation of the Marquardt–Levenberg algorithm for nonlinear least-square optimization. The geometric deformation model is a global three-dimensional affine transformation that can be optionally restricted to rigid-body motion, combined with isometric scaling. Good results were obtained for the registration of intramodality positron emission tomography and functional magnetic resonance imaging data. Based on this paper, a plug-in for ImageJ software was developed which was used as a basis for a modified version that is described later on.

Sorzano et al. presented an elastic registration algorithm for the alignment of biological images that represents deformation field as a B-spline model, which allows for a rich variety of deformations (Sorzano, Thevenaz and Unser, 2005). The registration problem is solved by minimizing a pixelwise mean-square distance measure between the target image and the warped source. The problem is further constrained by way of a vector-spline regularization that provides control over the deformation divergence and its curl. The algorithm is also capable of handling soft landmark constraints, which is useful when parts of the images contain very little information. An optimal analytical solution is proven in the case when only landmarks and smoothness considerations are taken into account. This paper was also a base for the ImageJ plug-in which was modified and used in the registration work.

Interesting registration and segmentation results were presented by applying primal-dual schema and MRF representation to a discrete grid laid over the images (Glocker et al., 2008). It guarantees finding a solution close to the global optimum in a reasonable time. However, the number of variables has to be relatively low. Although applicable to the MRI segmentation, this approach is not directly suitable for the OCT segmentation since the grid is too constrained to allow for a large variety of possible deviations from the normal retinal model.

An additional benefit for the stack registration would be use of an independent fundus image free from motion artefacts that would be used as a reference to register a fundus image created from the OCT stack. It was shown that an effective approach for generating a fundus image from the OCT signal in real time is to use the non-interference terms in the frequency spectra of the OCT signal to generate an intensity image (Jiao et al., 2005). The same author demonstrated use of these fundus images produced from the low density OCT scan as a reference for the registration of the high density OCT slices of the same subject (Jiao et al., 2006).

A method for mosaicing with correction of motion distortions and tissue deformations for in vivo fibered microscopy that is relevant to OCT stack registration has been presented (Vercauteren et al., 2006). It deals with a global registration problem that solves the error accumulation effect which occurs when registering a stack sequentially. However, some of the transformations used are specific to the fibered microscopy imaging so it cannot be used directly on the OCT stacks. The approach is based on a hierarchical framework that is able to recover a globally consistent alignment of the input frames, to compensate for the motion distortions and to capture the non-rigid deformations by modelling the relationship between the motion and the motion distortions. The proposed global alignment scheme is seen as an estimation problem on a Lie group, since the space of rigid body transformations is not a vector space but rather a Lie group that can be considered as a Riemannian manifold. A Lie Group is a smooth manifold together with a smooth composition map and a smooth inverse map, which satisfy the group axioms: closure, associativity, existence of a neutral element and existence of an inverse. Riemann manifold is a real differentiable manifold M in which each tangent space is equipped with an inner product, a Riemannian metric, which varies smoothly from point to point.

6.7 Data Segmentation

Since the focus of this work is on segmentation it is essential to give a detailed segmentation algorithms overview of both general and specific methods applied to the medical and OCT field in particular.

Segmentation refers to the process of partitioning a digital image into multiple regions. The general goal of segmentation is to simplify the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects or boundaries in images. The result of image segmentation is a set of regions that cover the entire image, or a set of contours extracted from the image. Pixels that belong to the same region are bounded by strong edges or have similar features or both. Some of the practical applications of image segmentation are in medical imaging, locating objects in aerial or satellite images, face recognition, fingerprint recognition and machine vision. In medical imaging segmentation can be used to locate tumours and other pathologies, measure tissue volumes, assist in computer-guided surgery, diagnosis, treatment planning, study of anatomical structure, etc.

In OCT in particular, existing published approaches and results vary in respect to the number of the layers segmented and on the tolerance to pathological cases. In general they tend to be very sensitive to pathologies and able to segment only the most prominent cases. The goal of this work is to develop an algorithm capable of segmenting all the visible layers and to be able to work on pathologies.

6.7.1 Classification of Segmentation Methods

Many general-purpose algorithms have been developed for image segmentation. Since there is no general solution to the image segmentation problem, these techniques often have to be combined with domain knowledge in order to effectively solve an image segmentation problem.

Histogram-based methods are simple and computationally very efficient when compared to other image segmentation methods because they typically require only one pass through all the pixels. An image histogram is computed and the peaks and valleys in the histogram are used to locate the clusters in the image. Usually colour or intensity is used as the measure. It is possible to extend this technique by recursively applying the histogram-seeking method to clusters in the image in order to divide them into smaller clusters. Main disadvantages of the histogram-seeking method are that it may be difficult to identify significant peaks and valleys in the image and that it does not take into account structural information.

Gaussian Mixture Model is a density model that is comprised of a number of Gaussian component functions that are combined to provide a multimodal density. It can be employed to segment an image based on the image statistics. Each mixture component is a Gaussian function with defined mean and covariance matrix.

The edge detection methods are based on gradient calculation. The assumption is that pixels with a low gradient constitute the background regions while pixels with high gradient values make edges or boundaries. Since the definition of low and high gradient value is obviously relative, a threshold value needs to be defined. The gradient is calculated using discrete differencing, in other words convolving the image with a differencing block that approximates continuous gradient at each point in the image (Hazaveh, 2003). However, the edges identified by edge detection are usually disconnected. To segment an object from an image it is necessary to extract closed region boundaries. This problem is usually solved by bridging the discontinuities if the distance between the two edges is within some predetermined threshold.

The region growing methods rely on expanding the initial regions. They can be divided into two basic subclasses: seeded and unseeded methods. Seeded method takes a set of seeds as input along with the image with each seed marking one of the objects to be segmented. The regions are iteratively grown by comparing neighbouring pixels to the attached region. The difference between a pixel's intensity value and the region's mean is usually used as a measure of similarity. This process continues until all pixels are allocated to one of the regions. The segmentation results are heavily dependent on the choice of seeds and in cases that seeds are placed in noisy regions, it can prevent the algorithm from advancing properly. Unseeded region growing is a modified algorithm that does not require explicit seeds and starts with a single region. It considers the neighbouring pixels in the same way as seeded region growing. If the neighbouring pixel differs from the region’s mean by less than a predefined threshold, it is added to the region. If not, the pixel is considered significantly different from all current regions and a new region is created containing that pixel.

The watershed transformation considers the gradient magnitude of an image as a topographic surface that is used for a water potential simulation. Pixels having the highest gradient magnitude intensities correspond to hills, which represent the region boundaries. Water placed on any pixel enclosed by common high ground flows downhill to a common local intensity minimum. Pixels draining to a common minimum form a catchment basin which represents one of the regions. Even though intuitively attractive, this method is very sensitive to noise present in the image.

Active contour (snake, curve propagation) is a class of popular techniques for object segmentation and tracking. The central idea behind this approach is to evolve a curve towards the lowest potential of a cost function, where its definition reflects the task to be addressed and imposes certain smoothness constraints. The contour is parameterized according to some sampling strategy and each element is evolved according to the image potential and internal smoothness terms. While such traditional techniques can be very efficient, they have a serious drawback that limits the evolution in special cases. One example is the case of multiple curves merging or splitting that is notoriously difficult to properly describe. The level set method can be used to efficiently address this problem. The central idea is to represent the evolving contour in a higher dimension, where its zero-level corresponds to the actual contour. According to the motion equation of the contour, it is possible to derive a similar flow for the implicit surface that when applied to the zero-level will reflect the propagation of the contour. The basic principle of operation is shown in Figure 12. However, if initialized far from the optimum, level set method can be extremely slow. It is also sensitive to noise and can get stuck at fake edges caused by noise or can evolve through the holes caused by noise present at the real edges.

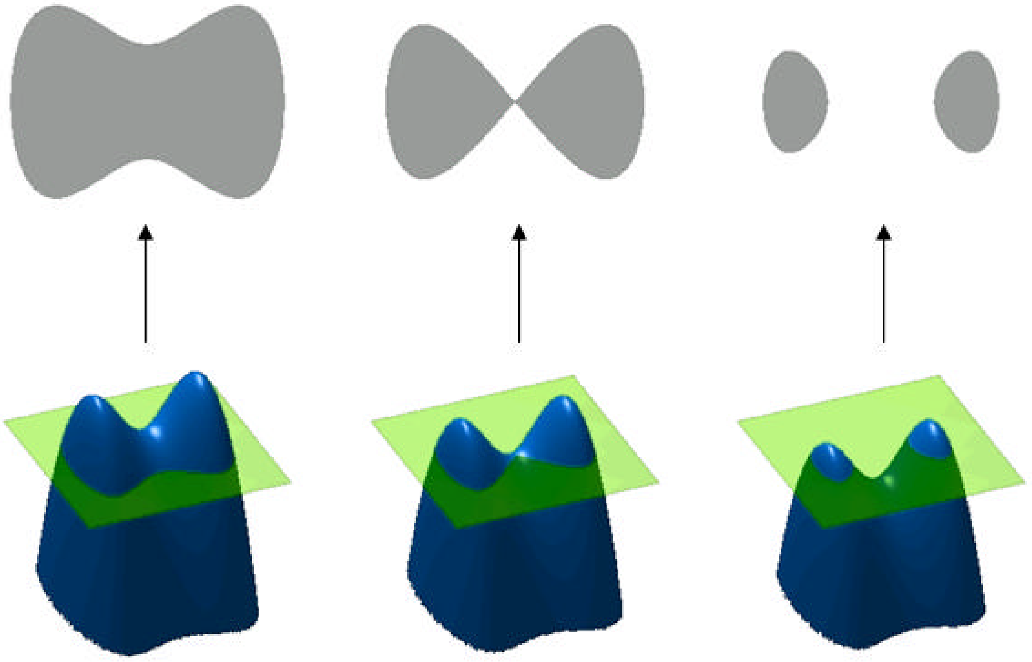


Figure 12 The level set method keeps the internal state of an evolving surface embedded in 3D (below). The plane cross-section produces the required segmentation (above). This type of representation allows for natural splitting of segmented areas, as the actual surface does not actually get split

Graph partitioning methods segment an image by modelling it as a weighted undirected graph. Each pixel is represented as a node in the graph and an edge is formed between every pair of pixels. The weight of an edge is chosen to represent the measure of the similarity between the pixels. The image is partitioned into disjoint sets by removing the edges connecting the segments. The optimal partitioning of the graph (the cut) is the one that minimizes the weights of the edges that were removed.

6.7.1.1 Model Based

Model based approach uses the variation obtained from the training set and imposes those constraints when segmenting an unseen image. This guarantees that the segmentation will be plausible and less sensitive to noise. However, it is extremely important to have a large, representative training set that includes all possible variation. It was described in the separate section that explains how statistical models can be built.

An extension to the standard AAM, called Robust AAM (RAAM) has been proposed in (Beichel et al., 2005) in order to enable more robust approach in presence of outliers. A two stage extension has been introduced; in the first step residuals are analyzed, while in the second stage optimization is used to detect gross outliers. This approach works well for cases when parts of the image are occluded or corrupted by strong noise, while the texture in the remaining part of the image is close to a prediction of the standard AAM model. Unfortunately, this approach is not directly applicable to the problem of OCT data segmentation, as the strong speckle variation is present in the whole image.

Interesting model based approaches have been applied in the face recognition and scene segmentation and classification fields. Cao and Fei-Fei presented a spatially coherent model that utilizes a visual bag of words representation of images but introduces spatial coherency and hierarchy of objects in an image (Cao and Fei-Fei, 2007). This approach could be useful if patches need to be segmented in OCT data that are not necessarily connected, but still obey some spatial consistency, which is often the case with layer detachments and fluid patches in pathologies. Another paper from the same research group performs scene classification, segmentation and annotation of objects utilizing a generative model (Li, Socher and Fei, 2009). This could provide very useful diagnostic benefit of not only segmenting various retinal layers but also providing a general prognosis for the patient. Kohli and Torr have presented a graph cuts approach in Markov random fields (Kohli and Torr, 2007). They applied it to a sequence of image frames in a video, utilizing the graph constructed from a previous frame to segment faster the current frame, without the need to construct a graph from scratch for each new image. This could potentially be speed wise beneficial to segmenting OCT stacks, as it can be noted that each slice in a stack is similar to its neighbours, just as videos frames are. These methods have been applied to problems where objects are usually well distinguishable, unlike the OCT data. Still, some of the ideas could be applied, modified, combined and used for OCT segmentation.

6.7.1.2 Extensions

Multi-scale approach is equally useful when applied to a segmentation problem as it is for registration. Segmentations are iteratively computed at multiple scales in scale-space and usually propagated from coarse to fine scales. This ensures that the process is both faster and less susceptible to noise.

Semi-automatic segmentation is often applied to difficult problems to enable the user to correct mistakes that might have been produced by the automatic algorithm or to initialize the algorithm and thus make it both faster and more accurate.

6.7.2 Segmentation of OCT Data

In the following, a series of publications will be reviewed that apply segmentation algorithms to the OCT and medical field in general or are relevant in some other respect.

A simple, yet rather effective, pre-processing algorithm for shadow compensation has been presented in (Fabritius et al., 2009a). Shadows are sometimes produced by the highly reflective upper layers of the retina, causing poor visibility of the lower layers. While a shadow compensation step is purely numerical and thus cannot remove the problem completely, it can alleviate some of the difficulties posed for the subsequent segmentation task.

Zawadzki et al. used a semi-automatic algorithm for the OCT segmentation where user could paint the areas of interest in any slice of the volume. For segmentation SVM was used with a feature vector that contained intensity, location, mean of the neighbourhood, standard deviation and gradient (Zawadzki et al., 2007).

A 3D graph search approach to OCT retinal layer segmentation was presented in (Garvin et al., 2009). The algorithm first aligned all the slices and straightened the RPE layer. After that, an optimal graph cut was performed with weights describing both edge and regional information. Good results were obtained, however they were obtained on slight pathologies and good signal quality data only. It would be too computationally demanding to apply this approach to more varied cases.

Fernandez et al. presented segmentation results of the peak finding algorithm (Cabrera Fernández et al., 2005). Since it is an iterative thresholding algorithm, it is sensitive to noise and deviation from the normal retinal data. It would be also difficult to extend it to any non-typical case since the training phase does not exist and the algorithm procedure is hard coded instead. Even though some good results were obtained, it is prone to failure and it is possible that detected boundaries will overlap.

Mujat et al. used the deformable spline algorithm (active contour) to determine NFL thickness (Mujat et al., 2005). They also detected blood vessels by detecting intensity holes in the RPE layer.

Vermeer et al. (Vermeer et al., 2011) used texture type classification, labelling each pixel independently, followed by a level set contour refinement stage. This approach suffers from an inefficient execution time and sensitivity to noise, the result of the level set method being unable to predict the probable shape based on training data.

Mayer et al. (Mayer et al., 2010) used a standard peak based detection approach for normal and glaucoma retina, however, for more demanding boundaries they used an energy minimisation term, compensating for the discontinuities and low contrast.

For rodent retinas Molnár et al. (Molnár et al., 2011) have presented an efficient ILM and RPE sliding window detection algorithm that assumes relatively flat boundaries, which in rodents is generally true. They then refine the results using an active contour approach.

Another rodent segmentation approach, in this case retinal degeneration, by Sarunic et al. (Sarunic et al., 2010) used a semi-automatic approach; user supplies a few initial boundary estimate points after which an active contour algorithm finds the boundary.

Yazdanpanah et al. (Yazdanpanah et al., 2009) used an active contour approach, minimizing a region term and assuming a consistent image intensity within one layer, with added terms for shape prior and regularization term to enforce smoothness. The initial position of the boundaries is user initialized, followed by the least squares optimization. Apart from this approach being semi automatic rather than fully user independent, it also assumes that layers are of uniform brightness, which is often not the case.

Rossant et al. (Rossant et al., 2009) combined several image processing and computer vision methods into a rather complicated, but slow algorithm. They first find the brightest boundaries by using a recursive low-pass filter and finding the maxima. For finding the weak boundaries, at roughly the same offset from the already determined boundaries, they use Kalman filtering, utilizing the principle of prediction/verification. Finally, remaining layers are found by performing texture classification using k-means clustering, after which the final boundaries are determined by optimizing a consistent area function using simulated annealing (SA).

A Markov boundary model was applied to connect the extracted boundary edge primitives in (Koozekanani, Boyer and Roberts, 2001). Even though more robust than standard column wise thresholding methods, it still relies on connecting 1D points. This makes it sensitive to noise and thus the detected layer boundary can easily drift off from the real boundary. Special rules have to be applied to correct for such cases which makes the whole approach less general.

In (Chiu et al., 2010) a similar approach has been used; first the layer boundary endpoints are detected, followed by a 1D tracking step. In this case optimal path is found using standard Dijkstra's algorithm. However, this approach suffers from exactly the same drawbacks as the previous one.

Yang et al. (Yang et al., 2010) used dynamic programming shortest path approach, based on working in dual scale edge; one image was created using axial gradient, while fine scale image was produced using the Canny edge detector. Dual scale approach alleviates to some extent the problems with shadows and other discontinuities present in OCT images.

A useful multiscale approach to segmentation is using wavelets since the multiscale image edges will correspond to the local maxima of the wavelet transform (Deng et al., 1993).

Highly regular wavelets that may be of general interest have been introduced for the detection of clustered micro calcifications in mammograms (Lemaur, Drouiche and DeConinck, 2003). Since they have higher regularity than classical wavelets, their performance is better for real signals that cannot be represented by lower order polynomials.

A novel approach to epidermal segmentation using shapelets has been demonstrated (Weissman, Hancewicz and Kaplan, 2004). Shapelet decomposition can be thought of as a two-dimensional wavelet decomposition in which an image containing a particular shape is used as a basis kernel rather than as an orthogonal or bi-orthogonal 1D or 2D waveform. After the shapelet transform has been performed, standard column wise search looks for the first maximum to locate the boundary.

An elegant approach to retinal segmentation based on spectral rounding was introduced in (Tolliver et al., 2008). It is a graph-partitioning algorithm that is based on the eigenvector calculation to determine the oscillation steps that represent the retinal edges. It performed well considering that no a priory information was available to the algorithm, simply dividing an image iteratively along the oscillation boundary. However, the number of extracted layers was low and it is unlikely that it would be possible to extract layers with weaker signal without using some structural information.

Based on this literature review it can be concluded that even though many segmentation methods and approaches have been introduced, a general and reliable solution for the OCT layer segmentation has not yet been developed.

7 Segmentation

Retinal and choroidal segmentation are the key to early pathology detection and therapy monitoring. Several novel methods have been developed to deal with real world, clinical, large datasets, consisting of normals and pathologies, compromised by noise and imaged using different commercial and non commercial OCT systems.

The chapter is divided into: work on registration algorithm (2.1) as it related to both the retinal (2.2) and choroidal segmentation (2.3). At the current stage, the segmentation works on each image in a stack independently, thus registration is not really necessary. However, extending the segmentation to 3D would make it essential. Additionally, to interpret the result, create thickness maps etc. it is important to have a stack registered at least approximately. Thus, in chapter 2.1, the early work on registration is presented, dealing with combining existing algorithms (rigid body and elastic) into a package applicable to OCT data registration, applicable to a whole stack in automatic fashion.

Retinal segmentation section contains description of an adaptive thresholding approach used for top and bottom foveal boundary detection (as a starting point for segmentation of the other inner layers), early work on data segmentation based purely on texture analysis, adaptive template matching approach to segmentation and finally a statistical model that unifies both the texture and shape information based on machine learning principles.

For segmentation of choroid affected often by advanced pathologies, it was necessary to develop a more robust top and bottom boundary detection algorithm. How it works and why it is applicable in large problem area outside OCT is explained in subchapter 2.3.1. In subchapter 2.3.2, choroid segmentation is explained using a statistical model similar to the one developed for the foveal segmentation (2.2.4)

7.1 Registration

Image artefacts resulting from motion may degrade image quality and cause inaccurate clinical interpretation of images. Motion artefacts arise when the object being imaged is moved during data acquisition but is assumed stationary in the image reconstruction process. Axial motion refers to a situation where the longitudinal distance from the probe to sample is changed in time, resulting in SNR reduction (Yun et al., 2004). Transverse motion results in both SNR reduction and degradation in transverse resolution, visible as blurring of an image along the direction of motion (Yun et al., 2004). Motion artefacts can be reduced by increasing the imaging speed, by applying eye tracking based OCT systems or by applying registration.

7.1.1 Stack Registration Algorithm

Stack registration algorithm has been developed based on the papers already mentioned in the registration section (Thevenaz and Unser, 1996), (Sorzano et al., 2005). The basic rigid body algorithm is used to register a source to target image based on a pyramidal approach. The stack registration algorithm uses the rigid body registration method and calls it iteratively to perform the stack registration. The elastic registration algorithm warps a source to target image and is not used by the stack registration algorithm.

These methods were modified in order to extend the existing functionality. The goal was to obtain a unified solution that will allow for approximate rigid body registration and fine warping producing a smooth stack. The first thing that had to be addressed, was the limited registration range of the rigid body algorithm, which meant that if there was a larger jump in the original stack it could not be registered properly. This issue was solved by increasing the number of pyramid levels. The next step was to apply elastic registration to each iteration of the stack registration procedure. To accomplish this, the elastic registration was successfully added to the stack registration loop. However, a problem appeared when registering larger stacks and using the warping sequentially. Errors would accumulate in each iteration due to numerical inaccuracies which meant that subsequent slices would be substantially deformed. To solve this problem, masking was applied to exclude the background from biasing the objective function and warping parameters were constrained heuristically to prevent unrealistic deformation of slices. However, this temporary solution did not solve the problem completely. The solution to this problem is planned for future work and has to be accomplished using a comprehensive solution that works simultaneously on the stack as a whole. However, since registration is not the focus of this PhD, this temporary registration solution is for now deemed satisfactory.

7.1.2 Results

The registration results presented compare the smoothness of the original input stack and stacks after the registration using different parameters. Results were subjectively judged based on the smoothness and geometrical distortions. No objective criteria were used for the evaluation, since it is a work in progress. Also, it is difficult to perform an objective comparison for this task since the motion artefacts occur during the scan of living subjects and thus performing the eye histology is not feasible. Has the histology been available one could evaluate the results either automatically or using a manual approach by marking the corresponding structures in two sets (OCT and histology).

Figure 13 shows a resliced side view of a stack obtained by a 800 nm system that is given as input. It is an appropriate representation because the motion artefacts along the stack are easily perceived. Figure 13 a) shows the original, unregistered stack and motion artefacts are very prominent. In Figure 13 b), the stack is displayed after the stack rigid body registration was applied using the modified rigid body plug-in. Most of the motion artefacts are corrected but the registration is still not completely satisfactory. The remaining motion artefacts are well visible at the top boundary where smaller jumps and roughness still remain. Subtle motion artefacts remain along the hidden axis as well. Figure 13 c) shows the stack after a modified stack registration plug-in that uses both the rigid body and elastic registration was applied. The stack is smoother but accumulated errors resulted in the progressive deformation and thinning of the whole stack. Even though useful results were obtained, a real 3D registration algorithm that works on the whole volume concurrently would be preferable.

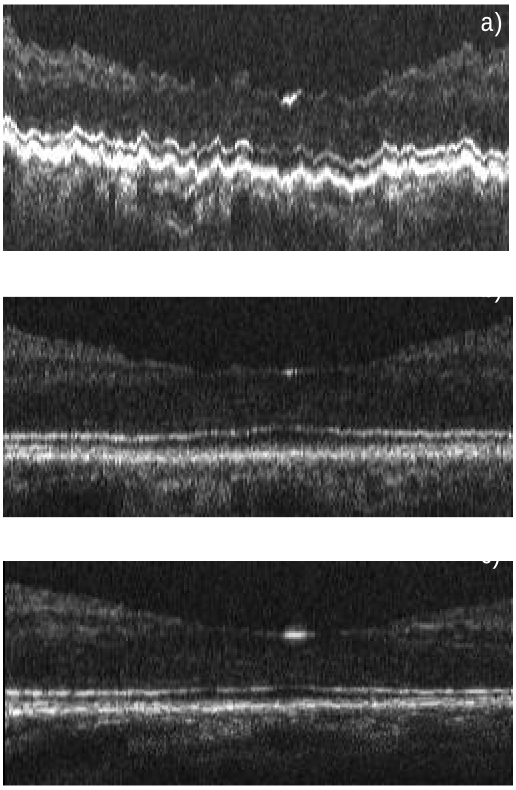


Figure 13 Side view of the resliced stack: a) without the registration, b) rigid body registration, c) rigid body and elastic registration

7.2 Retinal Segmentation

Retinal segmentation was done mostly with the goal of segmenting normal and close to normal retinas for purpose of early disease detection. Adaptive thresholding approach used for top and bottom foveal boundary detection (a starting point for segmentation of the other inner layers, replaced later by a more advanced approach described in 2.3.1), early work on data segmentation based purely on texture analysis, adaptive template matching approach to segmentation and finally the statistical model that unifies both the texture and shape information based on machine learning principles are presented in the subsequent sections.

7.2.1 Top and Bottom Retinal Boundary Detection

The next task to be solved was developing an algorithm that would detect top and bottom retinal boundaries since this information is necessary as a starting point for a subsequent multiple layer segmentation algorithm. This cannot be solved by simply thresholding column wise since the noise present in OCT data can cause either false edges or gaps in the real boundary. The first method tried was using a Level set approach, with force pushing the surface outwards in the normal direction and the inner force constraining the smoothness. This approach gave decent results but was not completely reliable (the curve would occasionally fall through in cases of noisy data) and could be extremely slow (in cases when there would be substantial amount of noise around the real boundary edges which would slow the curve down).

The adaptive thresholding algorithm starts by intensity thresholding, ILM boundary is found first by starting the thresholding process from the top of the image, while RPE boundary is found next by starting from the bottom. However, an image is divided into segments, allowing for different thresholds to be used in different parts of the image, as the image intensity is often not consistent throughout the image. Thresholding range and number of segments are chosen heuristically. Next, the algorithm converges to a close strong edge after the first estimate, additionally using constraints on distances between the boundaries. Robust polynomial fitting is afterwards used to eliminate outliers, followed by interpolation along the remaining points.

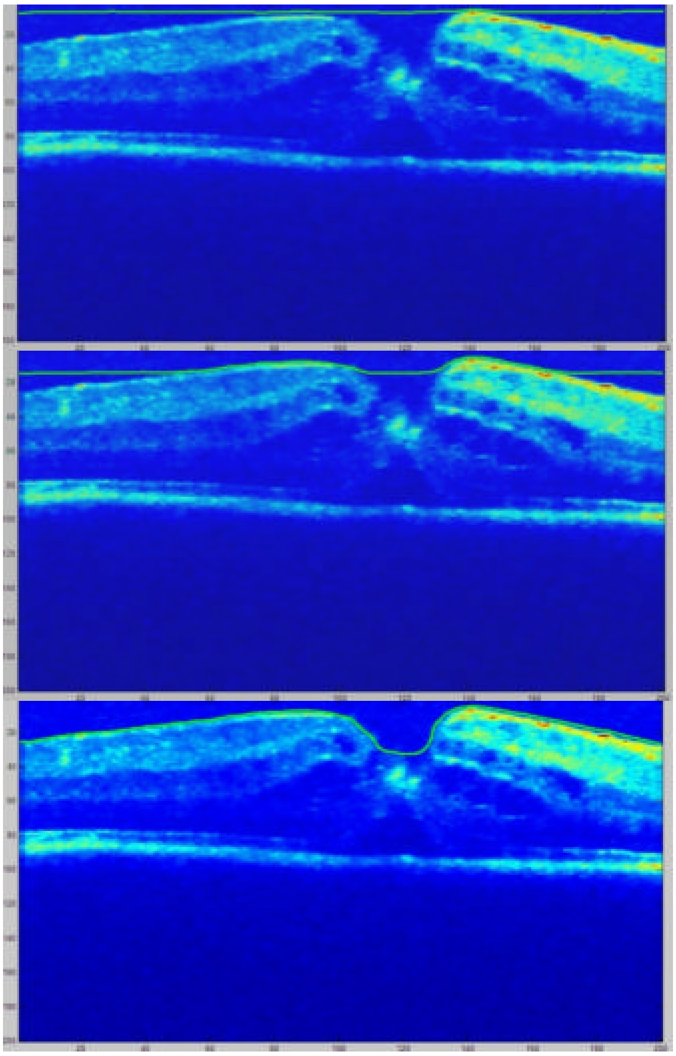
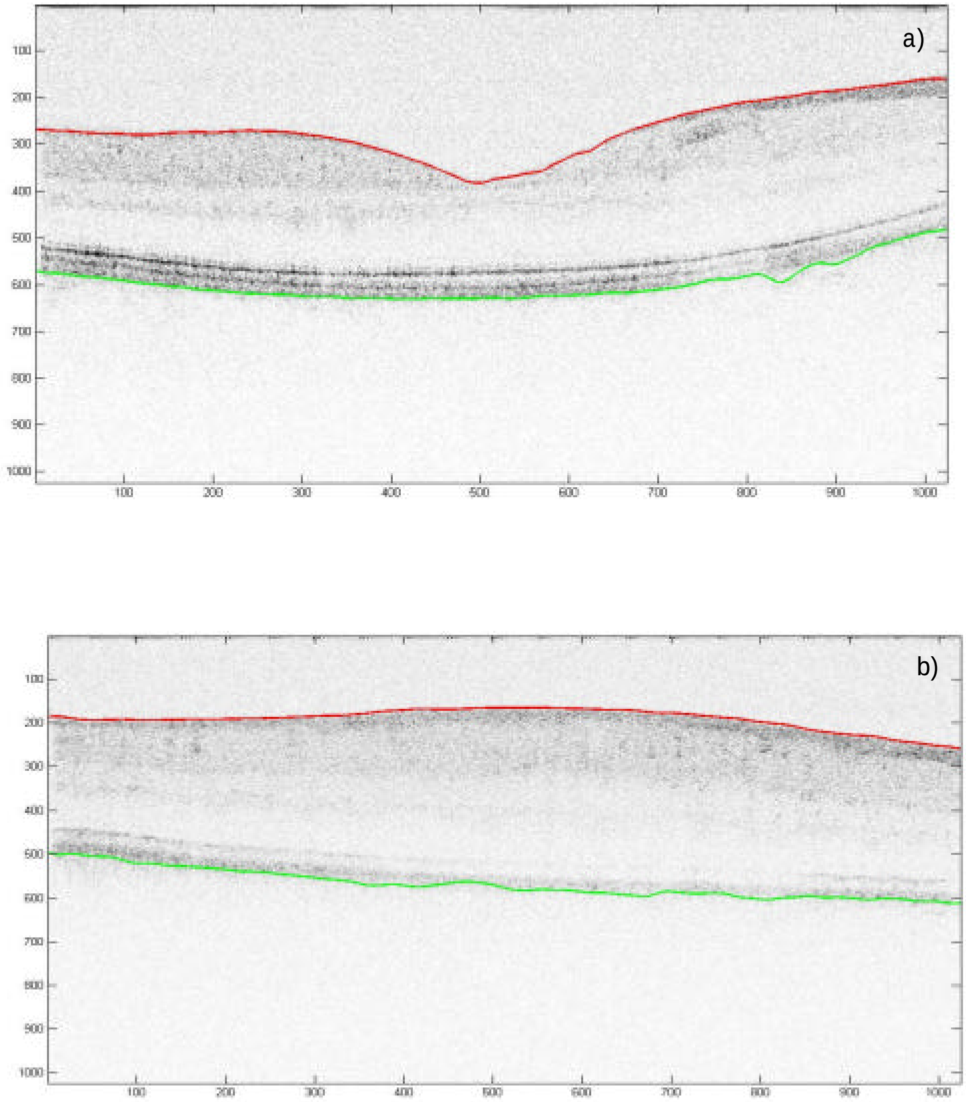


Figure 14 The level set method applied to the top boundary detection

7.2.1.1Results

The boundary detection methods seek a top and a bottom retinal boundary that is a prerequisite for the segmentation algorithm since it constraints the segmentation space. Results obtained by both the level set and adaptive thresholding methods are presented. In general the distance between the drawn curve and the boundary perceived by the observer is used as a subjective measure. Since faster and more robust methods have been developed for the same purpose in the later stages of the PhD, no further work was put into further refining and evaluating this method. Images come from 800 nm and 1060 nm systems, with mostly healthy retinas analyzed, as well as one age related macular degeneration (AMD) and two patients with changes in RPE. One stack comes from a time domain system.

Figure 14 contains three snapshots taken during the top boundary extraction level set run on a macular hole image. It can be noted that results are generally good. However, it took several minutes to obtain this simple extraction, while on the other hand if parameters are changed to make the evolution more aggressive there is a serious risk of contour penetrating the real boundary. Figure 15 shows a sequence of healthy retinas with the top and bottom boundaries marked as extracted by the adaptive thresholding. Processing times per B-scan are about 2 seconds, while the extracted boundaries match real boundaries very well. In Figure 16, three pathological B-scans are shown with difficult, irregular macular holes at the bottom. For all three cases, boundary detection performed equally well as in healthy subjects. Figure 17 shows a difficult retina that contains significant shadowing on the left which makes the NFL boundary hardly visible. Some noise is present around the boundaries as well. The algorithm did not perform as well as in previous cases but the extracted boundaries still roughly match real boundaries. With accurate stack registration applied, the algorithm could be extended to 3D, making the top and bottom surface extraction even more reliable.



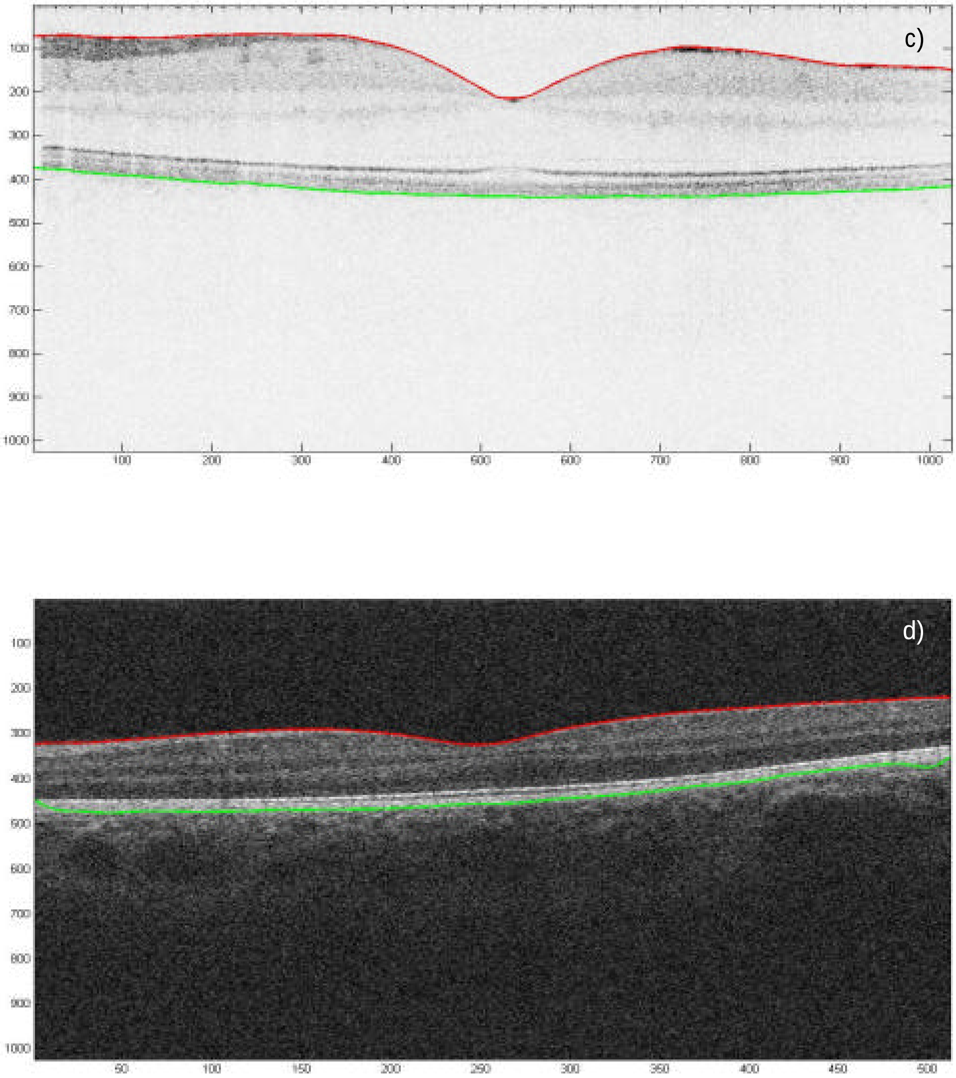


Figure 15 Top (red) to bottom (green) boundary detection using adaptive thresholding on healthy retinas imaged by two different OCT setups (high resolution prototype a) to c) and Zeiss commercial system d))

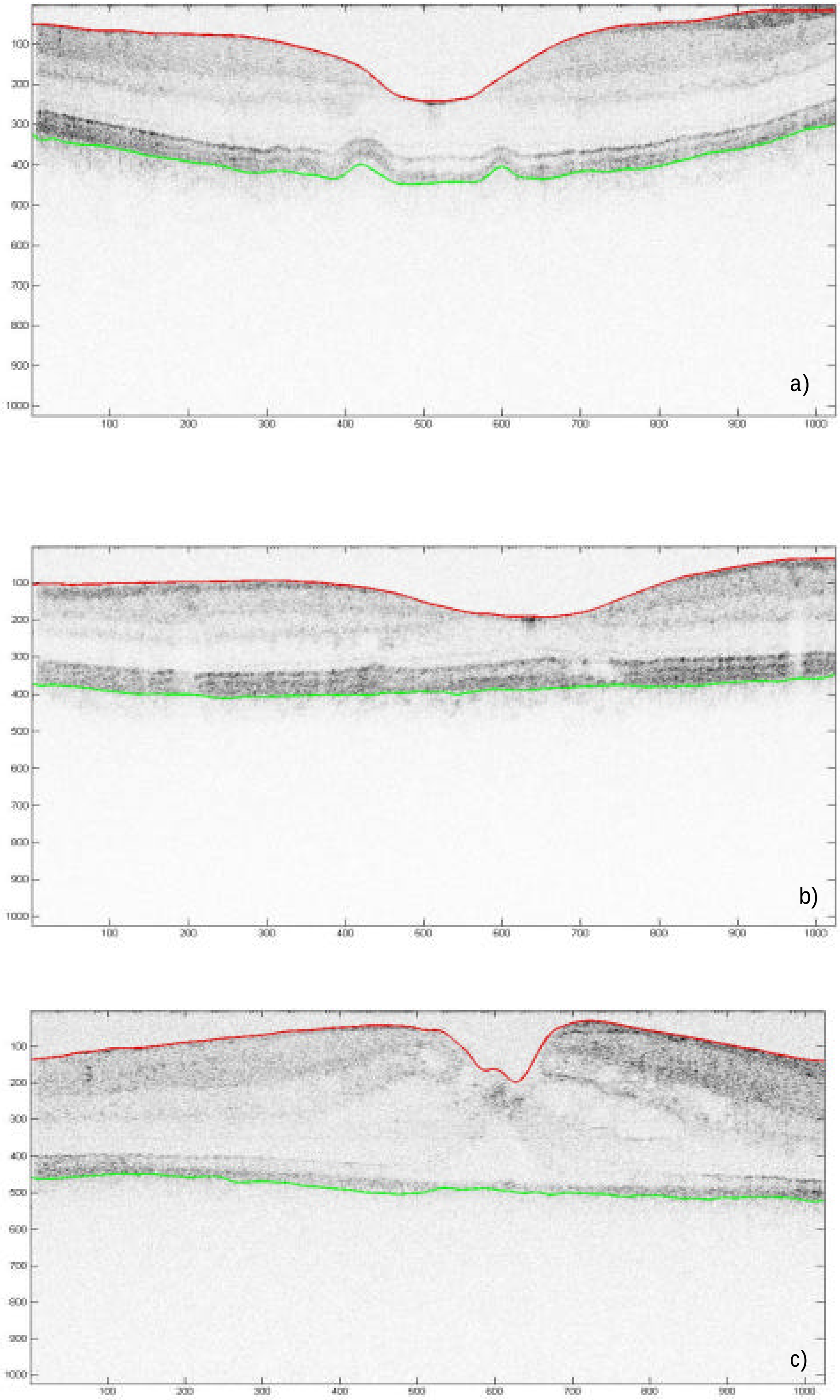


Figure 16 Top to bottom boundary detection on pathological retinas using adaptive thresholding (high resolution prototype OCT system). a) and b) show changes in the RPE, while c) is an AMD

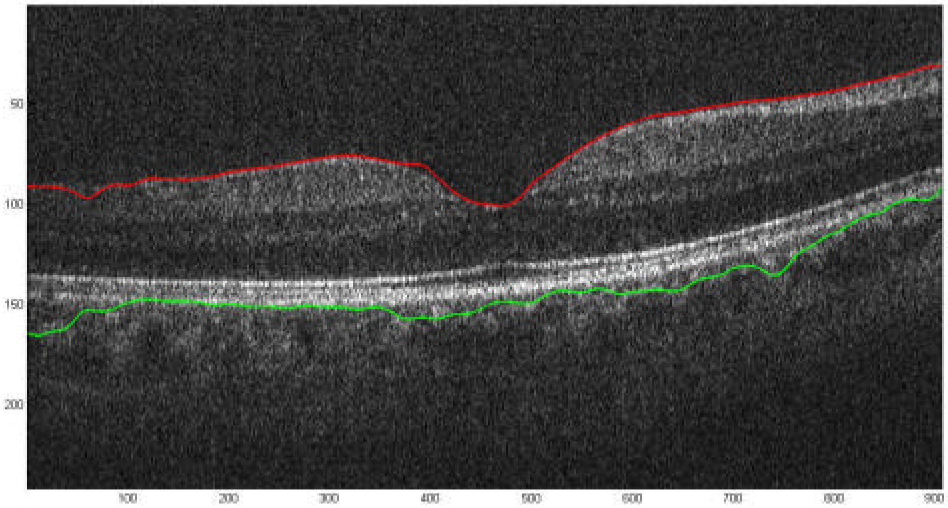


Figure 17 Top to bottom boundary detection using adaptive thresholding on a difficult, shadowed image of a healthy retina imaged by a Time domain OCT system

7.2.2 Texture-based Segmentation

Texture based segmentation, based on earlier work by Gavin Powell, was tried first. It is performed only on the region enclosed by the top (ILM) and bottom (RPE) retinal boundary. The basic algorithm scheme is given in Figure 18.

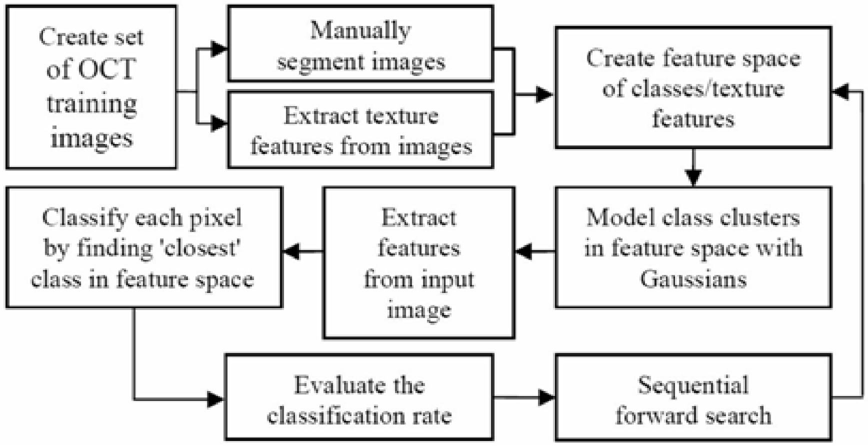


Figure 18 Texture based segmentation overview

The algorithm consists of a training and an application phase. In the training phase, a representative set of manually segmented images is used. The purpose of the training phase is to reduce the number of used features from currently 206 (including mean, entropy, Haralick, convex shell) to 70 by using sequential forward search algorithm (already mentioned in the segmentation section) and to obtain class distributions in the feature space, with every class representing a layer. Once the training is complete, the algorithm can be applied to the unseen data. Every pixel is simply mapped to the feature space and Mahalanobis distance is computed to the centre of each class. The smallest distance is chosen as the index to the belonging class. It is important to note that some geometrical features were used, that measure pixel's distance to the boundaries. Without these features, segmentation results tend to be patchy and inconsistent with layers overlapping. However, using geometrical features means that the training data layer geometry will be imposed on the unseen data. If the training set is too specific, the segmentation is too rigidly constrained. On the other hand, a varied training set does not constrain segmentation sufficiently to prevent the patchy, inconsistent results. This is illustrated best in the next section. Therefore, texture segmentation showed some potential but was unable to perform reliably on its own.

Texture based segmentation classifies each pixel as belonging to one of the layers and colours it accordingly. Results of the automatic segmentation can be compared to the manually segmented data and numerical evaluation can be used to rate the accuracy. While segmenting data that was very similar to the training data gave reasonably accurate results, the algorithm performed rather poorly on more diverse data (Figure 21). Since the errors were quite large, numerical measurements were not performed and focus was placed on extending and developing new segmentation approaches.

7.2.2.1 Results

Texture segmentation results of a healthy subject from the data set closely matching the training set are presented in Figure 19 with retinal and individual layer thickness maps shown in Figure 20.

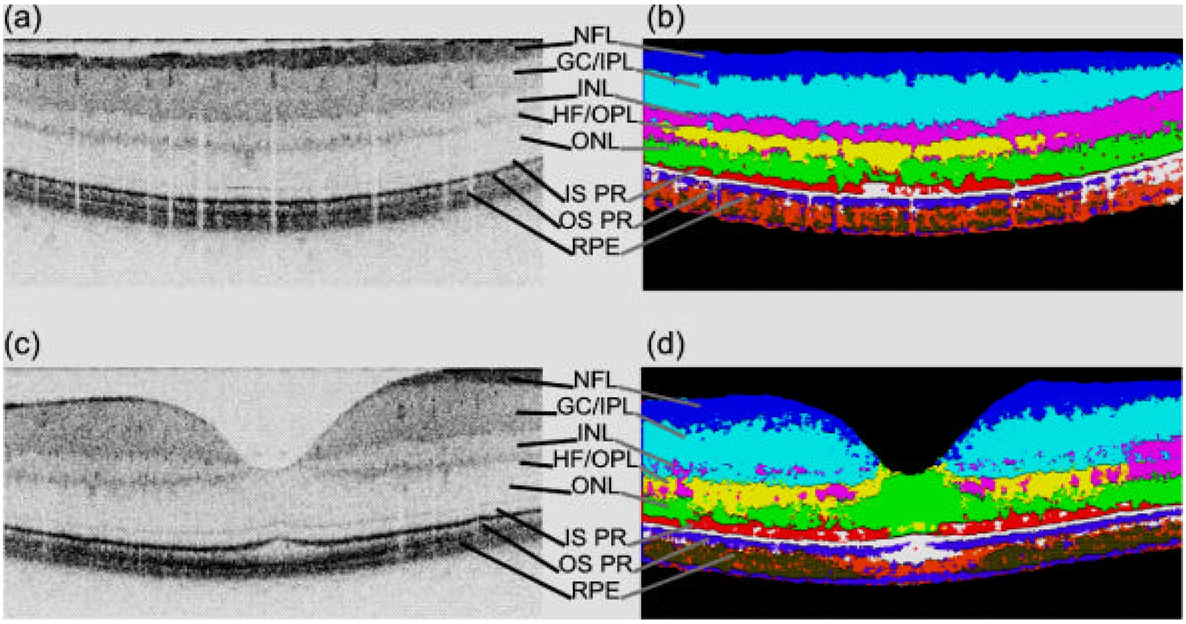
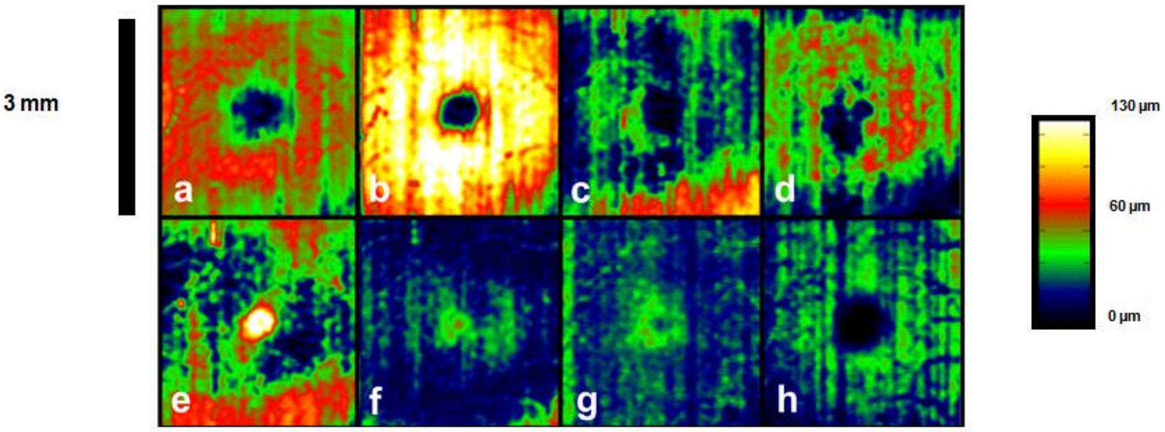


Figure 19 Texture segmentation results on a healthy retina



(a) NFL (b) GCL+IPL (c) INL (d) HF+OPL (e) ONL (f) IS PR (g) OS PR (h) RPE

Figure 20 The retinal thickness maps for a healthy subject

The obtained results show a decent performance of the texture analysis with segmented layers roughly corresponding to the real positions. However, as mentioned in the previous section, it is clear that the boundaries are not smooth but patchy since the classification of every pixel is independent. This effects mostly the segmentation of thin layers such as inner segment (IS), IS/OS, outer segment (OS) and RPE. It should be noted that some of the used features were geometrical, measuring the top and bottom retinal distance.

In Figure 21 results of the algorithm applied to two pathologies are shown. In Figure 21 a) the discontinuities in the IS, IS/OS and OS are visible in the segmented image. However, the segmentation in general is not accurate enough. It can be additionally seen how the geometry was imposed on the outer plexiform layer (OPL), ganglion cell layer and inner plexiform layer (GCL/IPL), with IPL/INL (inner nuclear layer) boundary not following the real boundary through the whole image.

Figure 21 b) shows the algorithm applied to the macular hole retina using the geometrical features, while c) displays segmentation results of the same image without the geometrical features being used. While in the b), results show what the ordering of the layers is correct, the segmented areas do not match the real image very well since the geometry of the training set is intrinsically imposed on the unseen image. In the c), row segmentation matches the real structure better, but the ordering of the layers is inexistent, with areas detected as NFL appearing in what is actually RPE because of similarities in texture.

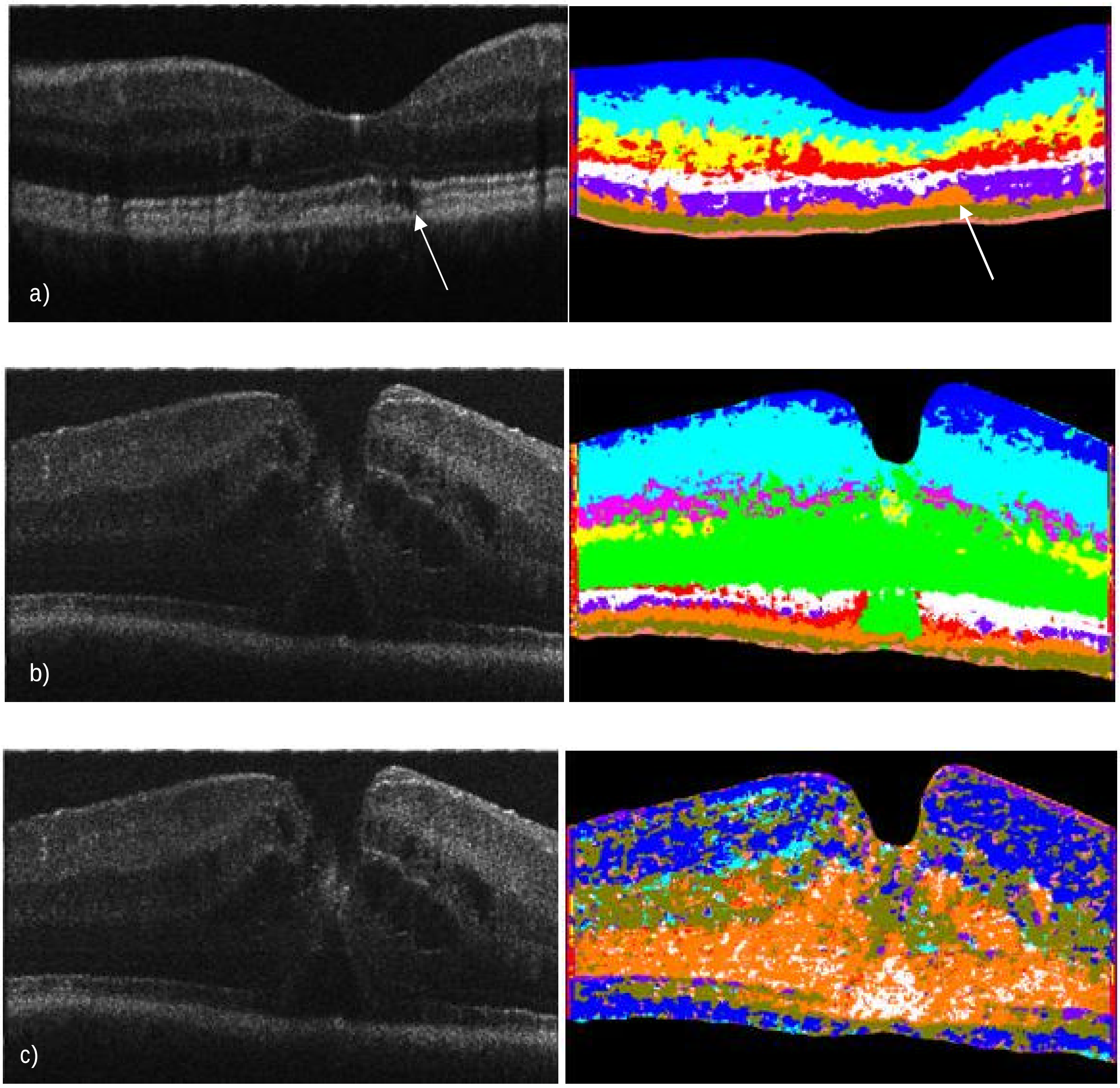


Figure 21 Texture analysis of pathologies with the algorithm trained on a healthy subject data set. a) shows result on a slight pathology, with localized damage to the RPE (arrow). The damaged area can be spotted in the segmentation result. However, the spatial constraints cause poor performance close to the fovea. In b), spatial constraints also cause poor performance, as the AMD geometry is significantly different from a healthy retina. In c), the same image is segmented without using the spatial features. Even though the layers are patchy and mixed, the segmented areas follow more closely the actual underlying geometry.

7.2.3 Adaptive Template Matching-based Segmentation

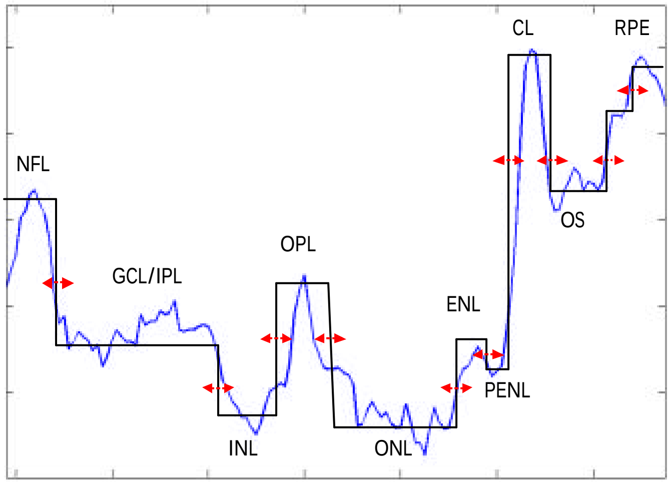


Figure 22 1D adaptive template matching: blue is an OCT A-scan, black the fitted template, with red arrows representing movable layer boundaries. Layer names are given above/below each layer

To address this problem, a geometry based, adaptive template matching solution was developed. The basic idea is to try to fit adaptive intensity based profile to each A-scan and assume that it will deform to best match the existing layer boundary. Figure 22 shows how a set of step adaptive templates are imposed on the A-scan intensity profile. The modelled layers include: nerve fibre layer (NFL), ganglion cell layer and inner plexiform layer (GCL+IPL), inner nuclear layer (INL), outer plexiform layer (OPL), outer nuclear layer (ONL), connecting cilia (CL), external limiting membrane (ELM), post-ELM (PELM), outer segment (OS), two segments of retinal pigment epithelium (RPE1 and RPE2). An active set MATLAB optimization method is used to deform them and best value is chosen as correct segmentation. Layer boundaries are simply the step transitions of the template. The optimization objective function is obviously the key for the algorithm performance. It is defined as a sum of normalized square difference of the template and the signal, the penalty for changing the template intensity and the penalty for changing the template step width with weights chosen heuristically. The optimization problem is constrained by forbidding height inversion of the neighbouring template steps. Also, the template step horizontal boundaries are not allowed to overlap. However, an individual A-scan is usually extremely noisy and it is very difficult to determine correct boundaries. To visualize this it is best to zoom into a retinal image and see how an A-scan looks like, Figure 23.

It is very difficult to determine correct layer boundaries looking at one A-scan even though by looking at the whole image this task is quite easy. Therefore, the algorithm has one more step. After every A-scan was separately segmented the connection phase will join the neighbouring A-scans into growing segments.

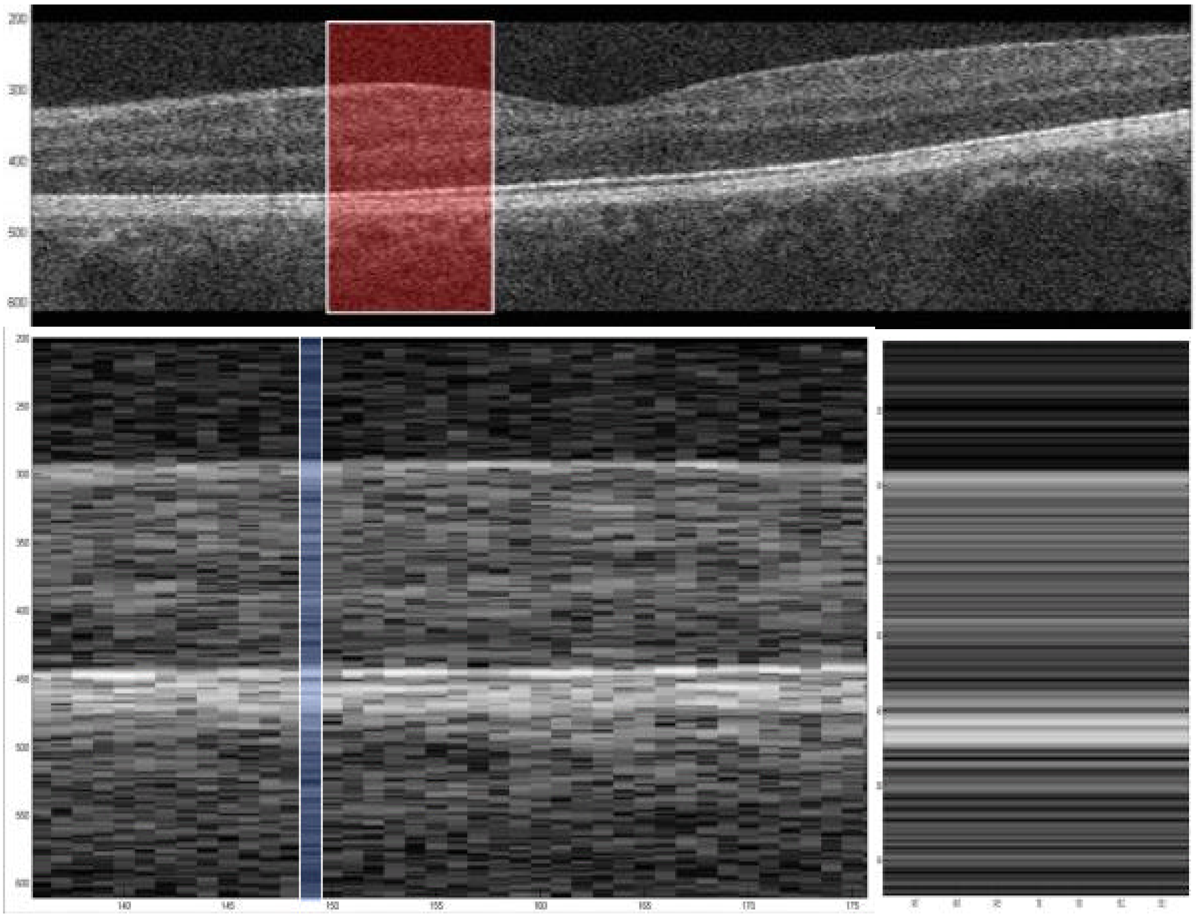


Figure 23 Zoom in to the A-scan level

This progression is shown in Figure 24. The left and right boundaries between the segments are shown in red and blue respectively, with rest of the template positions shown in green. First, the segments will start expanding to the “free” territory. As the segments come into contact they will merge if they are boundary-wise close enough, in other words below certain threshold. If no more segments can be merged the threshold is adaptively relaxed by a fixed step until all the segments are connected. When two segments actually connect, neighbouring values of each segment propagate through the other segment and replace the worse existing values.

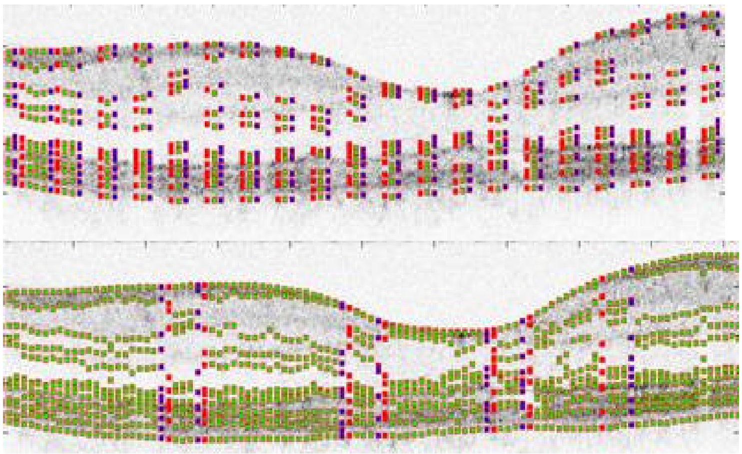


Figure 24 Connection phase. Starting from the initial positions (green), the segments are connected (red and purple)

However, this correction scheme does not correct the problem thoroughly. As registration should be done on the whole stack simultaneously, segmentation should be performed simultaneously on the whole image or volume using smoothness term in the objective function. To illustrate why this is important consider Figure 25. This figure illustrates two imaginary neighbouring A-scans. The blue line follows the strong edge while the red line follows the weak edge with reasonable smoothness. Using independent A-scan segmentation, the blue line would be chosen as the proper boundary, even though it is actually the red line that approximates it best. Therefore, the segment connection phase cannot correct for such a problem since it would impose one of the A-scan segmentations on the neighbour, while the correct solution is neither of those. Thus, to have a reliable method it is necessary to perform segmentation simultaneously. However, this introduces a new difficult problem to the optimization method used. The number of variables becomes large, even when reducing the original resolution for the rough fit. Even worse, the large number of inequality constraints makes the objective function domain extremely difficult to optimize. Trust region method implemented in MATLAB for the large scale problems and SPSA algorithm have been tried without any success. This result is not unexpected and leads to the model based approach as probably the optimal solution. It is important to note that the work done so far is a natural evolution of trying first the separate methods and then combining them together. It would be ungrounded to conclude at the start that a complex statistical model is needed if the simpler solutions have not been thoroughly tested.

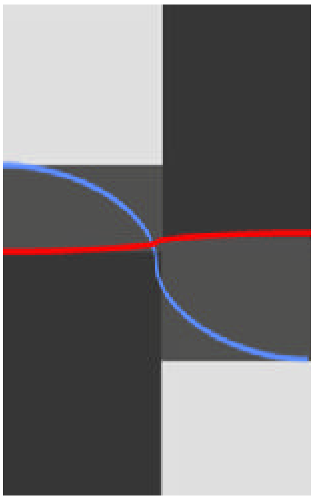
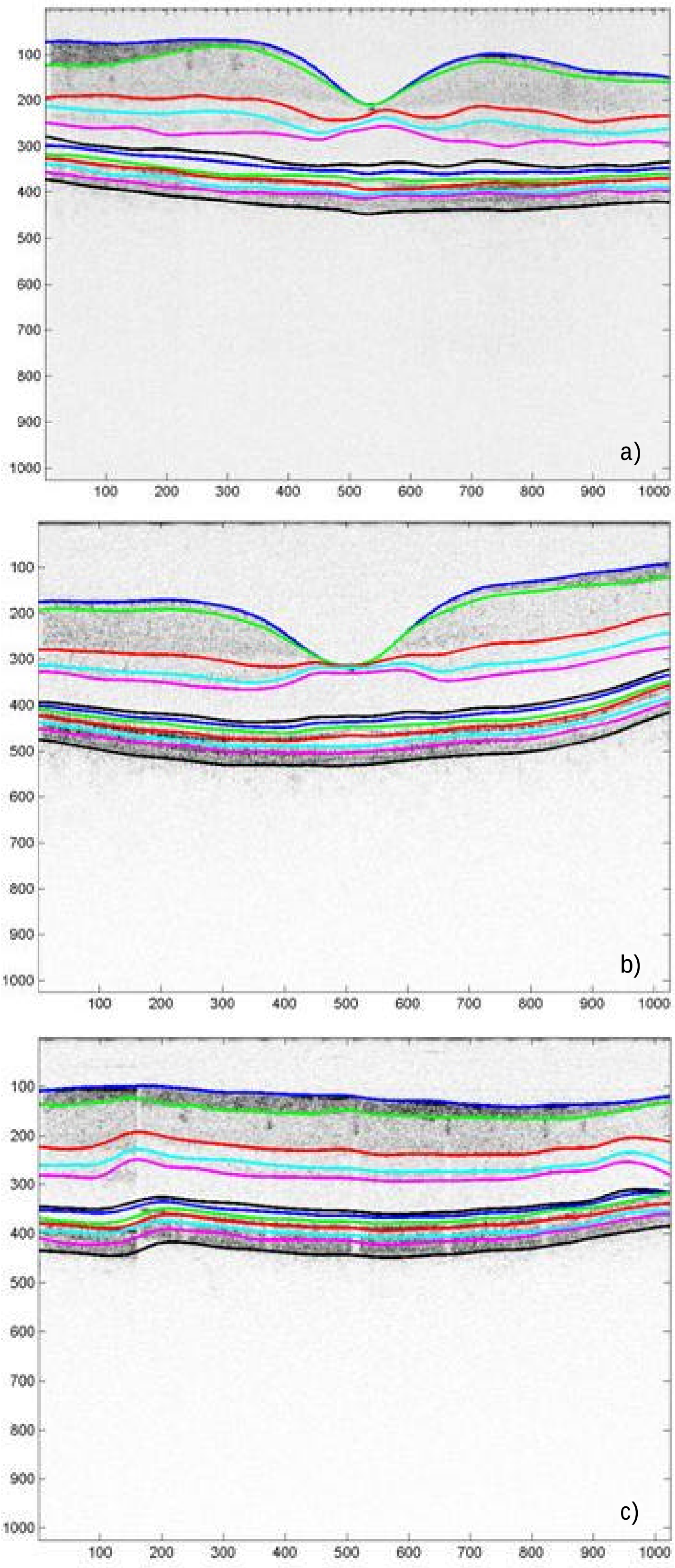


Figure 25 Weak (red) and strong (blue) edge



7.2.3.1Results

The principle of adaptive template matching based segmentation is different than that of the texture segmentation. Instead of classifying each pixel individually, boundaries are adaptively fitted to give the best match according to the objective function. Thus, on the images shown, boundaries are marked with coloured contours and ten inner layer/interface boundaries were classified. The numerical comparison to the manually segmented data was not performed, as in some cases considerable inaccuracies were already visible from the visual inspection of the results. Thus, the subsequent model based approach was developed and evaluated.

A series of results obtained using adaptive template matching is shown in Figure 26

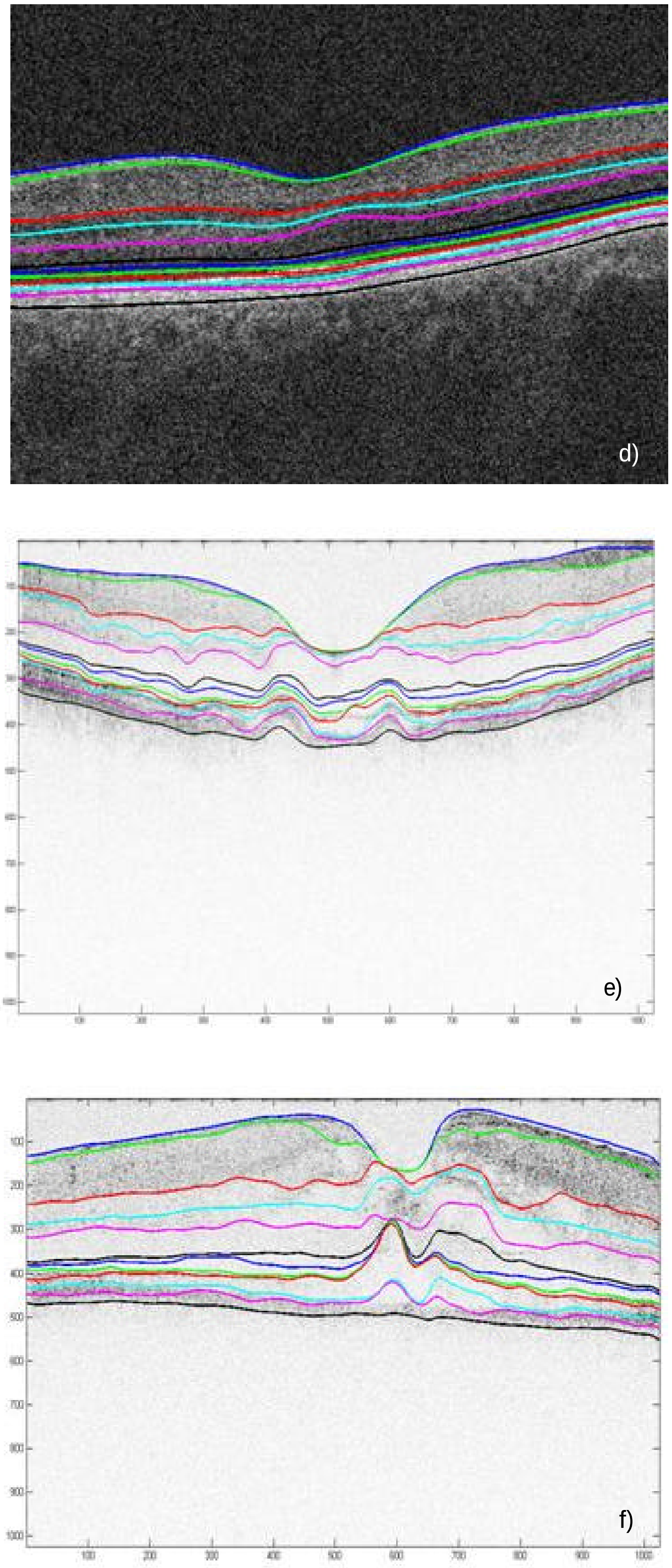


Figure 26 Adaptive template matching applied to normals (a) to d)) and pathologies (e) and f))

with thickness maps of two normals displayed in Figure 26. These results demonstrate that by using global geometrical information smooth boundaries can be obtained that match roughly the real boundaries. For major layers the obtained boundaries are usually close to their real values while the less prominent layers are less accurately positioned. In some areas matched templates do not follow the real boundary very well that can be explained by recalling that A-scans individually are very noisy. As was mentioned in the previous section, penalty term was used to contain too large shifts from the original template shape. That explains why the template will often not match well to the real boundary since this boundary has to be strong in order to attract the template despite the penalizing term. However if that term was not used, noise in the individual A-scans would be overwhelming, and likely cause even greater inaccuracies. Another problem is that, usually, in the whole processed stack, segmentation of some slices fails significantly in certain areas. This can be seen in thickness maps where certain spots are visible that actually do not match any real pathological thinning or displacement since normal subjects were segmented. Those spots correspond to the areas were the algorithm failed. The two pathologies shown in the last two rows of Figure 26 demonstrated the performance of the algorithm in difficult cases. A macular hole is especially difficult since this example shows an advanced pathology. It is encouraging that the RPE detachment was detected despite the major dissimilarity between this case and a normal retina.

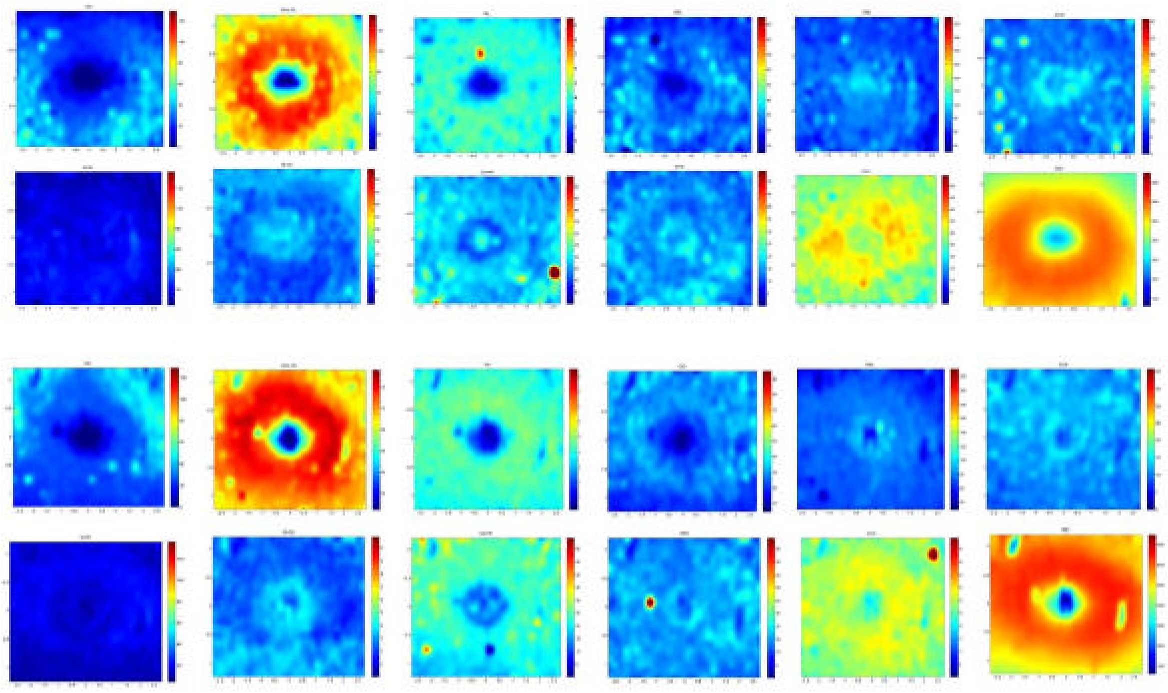


Figure 27 Thickness maps of eleven layers (nerve fibre layer (NFL), ganglion cell layer and inner plexiform layer (GCL+IPL), inner nuclear layer (INL), outer plexiform layer (OPL), outer nuclear layer (ONL), connecting cilia (CL), external limiting membrane (ELM), post-ELM (PELM), outer segment (OS), two segments of retinal pigment epithelium (RPE1 and RPE2)) and retina obtained by applying adaptive template matching to two healthy subjects.

7.2.4 Model based retinal segmentation applied to normal human fovea

The idea behind the model based approach is to utilize information from the manually segmented example images (training data) to constrain the space of possible solutions to a plausible set. Generated speckle noise was progressively added to the data to test the robustness of the algorithm. Also, accurate segmentation is achieved in cases even when large shadows and low signal are present.

A three-dimensional OCT system was used for imaging. It uses a superluminescent light source, with 840nm central wavelength and 50nm optical bandwidth. Axial resolution is 5-6 microns, while transverse resolution is 15-20 microns. Data acquisition speed was 27 klines/sec. Optical power was 500 µW and SNR was 96dB with a sensitivity roll off - 6dB/mm. Depth range was 3.5mm and axial sampling 2.3 µm/vx.

7.2.4.1Algorithm overview

As seen in the overview of the algorithm (Figure 28), one can observe that the pre-processing stage is performed for both the training step and the segmentation of the unseen data. Once the variation parameters have been learned from the manually segmented training data, they can be used to drive the model to perform segmentation of unseen data. The actual segmentation process is essentially an optimization run that changes the model parameters in order to minimize the objective function that defines the difference between the model and a given unseen image that is to be segmented.

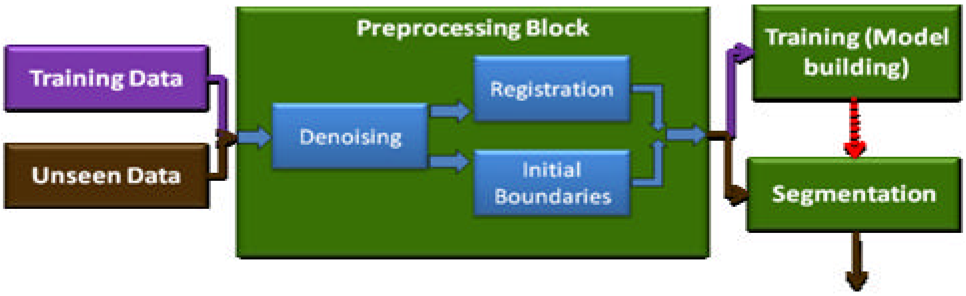


Figure 28 Algorithm overview: manually segmented data is used as the input to the training phase of the algorithm. After passing the pre-processing block, a statistical model is constructed that captures the variance in the training data, which can be then used to segment unseen data.

7.2.4.2 Pre-processing

Before the segmentation process, dual-tree complex wavelet (DTCW) denoising is applied to the data. The denoising algorithm exhibits very good performance, while being computationally efficient (Selesnick, Baraniuk and Kingsbury, 2005). This reduces the speckle noise present and thus makes the subsequent segmentation tasks easier. However, even without this step the model based approach outperforms the previously developed methods.

Denoising based on quasi-random nonlinear scale space described in (Mishra et al., 2010a) and applied to OCT speckle reduction in (Wong et al., 2010) would likely be more effective. It is an effective and fast method based on formulating the denoising problem as a general Bayesian least-squares estimation problem. A quasi-random density estimation approach is introduced for estimating the posterior distribution between consecutive scale space realizations. However, the relatively small performance difference (larger speed difference) in not significant for the performance of the statistical model, thus a well-tested and freely available DTCW code was used.

After that, registration of the stack and segmentation of the three initial well-defined boundaries (ILM, connecting cilia (CL) and end of RPE) is performed. Registration and initial boundary location finding are currently independent since detection of the initial boundary location operates on each B-scan independently.

A stack registration algorithm has been developed based on B-spline multi-resolution pyramid registration approach (Thevenaz and Unser, 1996) and (Sorzano et al., 2005). The basic algorithm for translation and rotation is used to register source to target image (7.1).

ILM, CL and end of RPE boundaries are found using an adaptive thresholding algorithm (auto adjusts to appropriate power, similar to 7.2.1) that converges to a close strong edge after the first estimate, additionally using constraints on distances between the boundaries. Robust polynomial fitting is afterwards used to eliminate outliers, followed by interpolation along the remaining points. ILM boundary is found first by starting the thresholding process from the top of the image, while RPE boundary is found next by starting from the bottom. The CL boundary is determined the last and depends on the positions of the already found ILM and RPE boundaries. It is found starting from the top after eliminating the pixels in the neighbourhood of the already found ILM boundary and imposing constraints on the distance from the RPE boundary. An example with a large shadowed area is shown in Figure 29.

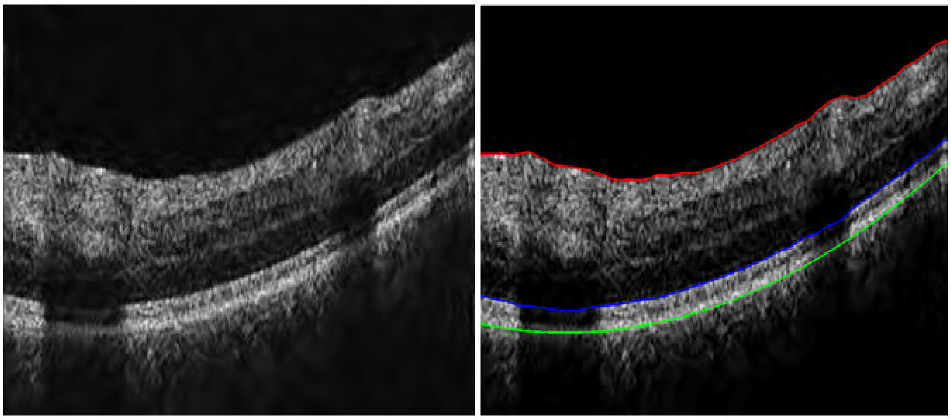


Figure 29 Initial segmentation step of a despeckled OCT frame (on the left) after adaptive thresholding boundary detection demarking (on the right): internal limiting membrane (ILM, red), connecting cilia (CL, blue), retinal pigment epithelium (RPE, green).

Initial boundary detection would also be possible based on a decoupled active contour (DAC) approach as presented in (Mishra, Fieguth and Clausi, 2010b). The level set active contour approach was tested and discarded for its slow convergence. However, DAC is both robust and fast, as it decouples the measurement (solved by using Hidden Markov Model (HMM) and Viterbi search) and prior active contour energy terms. As it was found that the initial boundary estimation approach was sufficient for the current application, not all of the other available methods were tested. For future work, however, algorithms such as DAC could prove valuable.

7.2.4.3Model building

After the pre-processing stage the statistical model is first trained on a set of manually segmented images and can be then applied to the unseen data. Using a statistical model based on the training data is a potentially effective tool for both segmentation and registration (T.F.Cootes, Edwards and C.J.Taylor, 2001). Its main advantage is that knowledge of the problem can be used to resolve the confusion caused by structural complexity, provide tolerance to noisy or missing data, and provide a means of labelling the recovered structures. The idea is to perform supervised learning by applying knowledge of the expected shapes of structures, their spatial relationships, and their textural appearance to restrict the automated system to plausible interpretations. Supervised learning is a type of machine learning for learning a function based on training data, which consists of pairs of input objects, and desired outputs. The task of the supervised learner is to predict the value of the function for any valid input object after having seen a number of training examples. To be useful, a model needs to be specific, capable of representing only legal examples of the modelled object.

From the manually segmented images, the shape and texture features were extracted and for each image all the extracted shape features are arranged into one vector and all the texture features into another vector. Separate models for shape and texture are constructed similarly, so only the shape model construction will be explained. If there are *m* training images, for each layer (*n* layers) one vector of offsets $\mathbf{v}$ is obtained per layer, per image of width w, which stacked together for all the layers define $\mathbf{x}$. All of the manual segmentations then comprise the matrix $\mathbf{X}$ Eq. (1).

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_m \end{pmatrix} = \begin{pmatrix} \mathbf{v}_{11} & \cdots & \mathbf{v}_{1n} \\ \vdots & \ddots & \vdots \\ \mathbf{v}_{m1} & \cdots & \mathbf{v}_{mn} \end{pmatrix}$$

$$\mathbf{v}_{ij} = [off_1 \dots off_W] \tag{1}$$

Shape features that are used are sparsely sampled distances of the boundaries from the top boundary (ILM). Texture features that are currently used are simple, although it is trivial to include additional features if needed to further increase performance in cases of vessels, large shadows and pathological tissue; currently used features are the mean of all the pixels for each of the layers in the original image, standard deviation and mean of all the pixels for each of the layers in the median filtered image, as well as the multiple-scale (a pyramid of Gaussian filtered versions of the image) edges sampled along the boundaries. In practice, for an image of width 512, each boundary was sampled at 26 positions. Thus there are 26 spatial features and 4 texture features per each layer, and for eight layers, 208 spatial and 32 texture features were obtained.

Statistical models can reproduce specific patterns of variability in shape and texture by analyzing the variations in shape across the training set. It is difficult to achieve this selectivity, whilst allowing for natural variability, without using very large descriptors and thus it is essential to select good features from the training set for the model-building phase. The key step of the statistical model-training phase is the dimensionality reduction of the large set of features from the training data set. The reason for dimensionality reduction is to reduce the computational cost of the optimization method that is used to fit the model to the real data later on. The idea behind this concept is to find statistical dependencies between the produced features and reduce the dimensionality of the space by identifying only a certain number of the most prominent properties in the data set, represented by the most important eigenvectors.

Principal component analysis (PCA) is the standard vector space transform technique used to reduce multidimensional data sets to lower dimensions for analysis. It works by calculating the eigenvalue decomposition of a data covariance matrix or singular value decomposition of a data matrix. Usually a relatively small number of eigenvectors with greatest eigenvalues can describe the original data well. If $\mathbf{X}$ is the original data matrix, as defined in Eq. (2), after the decomposition only L principal components can be selected and in that way project the data into a reduced dimensionality space to get $\mathbf{Y}$ Eq. (2).

$$\begin{aligned} \mathbf{X} &= \mathbf{W}\Sigma\mathbf{V}^T \\ \mathbf{Y} &= \mathbf{W}_{\mathrm{L}}^T\mathbf{X} \end{aligned} \tag{2}$$

$\mathbf{W}$ is a matrix of eigenvalues of $\mathbf{X}\mathbf{X}^T$, $\mathbf{V}$ is a matrix of eigenvalues of $\mathbf{X}^T\mathbf{X}$, while Σ is a diagonal matrix with nonnegative numbers on the diagonal.

However, rather than PCA, neural network based dimensionality reduction was used since it offers nonlinear eigenvectors and therefore can reduce the space more compactly if the data is nonlinearly distributed than the linear representation obtained by PCA (Scholz, Fraunholz and Selbig, 2007). The shape features proved to be nonlinear and thus a more compact representation was obtained using nonlinear dimensionality reduction that required a smaller set of eigenvectors than the PCA to capture the same variability within the training set. A Neural network (NN) is a mathematical or computational model based on principles found in biological neural networks. It consists of an interconnected group of artificial neurons and processes information where each connection between neurons has a weight, with the weights modulating the value across the connection. The training phase is performed to modify the weights until the network implements a desired function. Once training has completed, the network can be applied to data that was not part of the training set. It is useful to note that a special type of neural network (inverse) (Scholz, 2005) can be used to perform dimensionality reduction on the training feature set that is produced which contains missing values. Missing values occur when no data value is stored for the variable in the current observation. The generating function is used to produce larger dimensionality data $\mathbf{X}$ from the parameters $\mathbf{z}$ (equivalent to $\mathbf{W}_L$ in PCA) Eq. (3). The extraction function does the reverse.

$$\begin{aligned} &\Phi_{gen} : \mathbf{z} -> \hat{\mathbf{X}} \\ &\Phi_{extr} : \mathbf{X} -> \mathbf{z} \end{aligned} \tag{3}$$

The problem of missing data occurs because during the registration process slices are moved, and since input to the dimensionality reduction step has to be a rectangular matrix, it is necessary to fill the missing values. In practice these can be set as “not a number” (NaN) and perform the nonlinear PCA (Figure 30). After that, a reduced number of variables is obtained which can reasonably well describe any variation observed in the training data. The dimensionality of the original spatial feature space was reduced from 208 to 12, and the texture feature space from 32 to 2. This number of eigenvectors allowed for an efficient optimization in the subsequent steps, while still preserving the original data variation well.

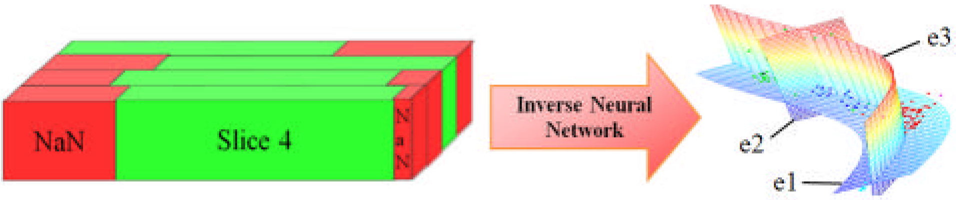


Figure 30 Filling the gaps after the registration with NaNs and applying inverse neural network nonlinear PCA dimensionality reduction. In the case of the example data shown on the right, it can be seen that already the first eigenvector (e1) captures most of the variance in the original data set. This illustrates the idea behind the dimensionality reduction.

The approach is based on a similar concept to the Active Appearance Model (AAM). For completeness, it will be first explained how the basic AAM model works, followed by an explanation of how the proposed statistical model differs from that concept. An AAM manipulates a model capable of synthesising new images of the object of interest by finding the model parameters which generate a synthetic image as close as possible to the target image (T.F.Cootes et al., 2001). An AAM will, based on learned shape deformation, generate a new image with a texture learned from the texture variation and then compute the distance between the synthesized and the given image that is to be segmented. $\mathbf{x}$ is the shape vector (which is normalized by subtracting the mean shape and rescaling, Eq. (4)) and $\mathbf{g}$ is the texture vector obtained from an image $\mathbf{I}$ and the shape vector (it is also normalized) Eq. (5).

$$\mathbf{x} \longrightarrow (\mathbf{x} - m(\mathbf{x})\mathbf{1})/s(\mathbf{x}) \tag{4}$$

$$\mathbf{g} = G(\mathbf{x}, \mathbf{I}) \tag{5}$$

Function $S(\mathbf{s})$ produces new shape vectors by adding the shape parameters $\mathbf{s}$ multiplied by the shape matrix $\mathbf{Q}_s$ (a matrix of sorted eigenvectors learned from the training set, usually produced by PCA decomposition) to the mean shape vector $\bar{\mathbf{x}}$ Eq. (6). The same procedure is used to generate new texture vectors.

$$\begin{aligned} \mathbf{x} &= S(\mathbf{s}) = \bar{\mathbf{x}} + \mathbf{Q}_s\mathbf{s} \\ \mathbf{g} &= T(\mathbf{t}) = \bar{\mathbf{g}} + \mathbf{Q}_g\mathbf{t} \end{aligned} \tag{6}$$

However, unlike the AAM which compares pixelwise synthesized images, the layer boundaries produced by the model were used during the optimization to compute texture features of the bounded area and compare them to the expected texture properties of each layer learned from the training set. This approach is used since unlike the areas in which AAMs are usually applied, the texture of retinal OCT scans varies so much within one layer that the direct comparison with a synthesized image is unusable. The objective function (Eq. (7)) evaluates how well the model matches real data and is minimized during the optimization.

$$f(\mathbf{s},\mathbf{t}) = \left|T^{-1}(G(S(\mathbf{s}),\mathbf{I})) - \mathbf{t}\right| + \frac{b^* \sqrt{\sum (S(\mathbf{s})_b - \mathbf{b}_{init})^2}}{w} \tag{7}$$

b is the number of boundaries, w is image width and $\mathbf{b}_{init}$ defines the initial three boundaries positions found by the adaptive thresholding algorithm. T^{-1} is the inverse of T; T is defined in Eq. (6). T^{-1} returns the model texture parameters **t** that are most likely to generate a given vector of texture features **g**. The first term of the objective function defines the main measure for evaluation of the model fitting, determined by the difference of the model texture parameters and the texture parameters extracted from the image regions defined by the model shape parameters. The second term penalizes deviations from the initial boundary as found by the initial three boundaries algorithm and the one produced by running the optimization function for the statistical model. This is an important novelty, when compared to the standard AAM, which helps to constrain the optimization process to valid solutions. Additionally, the optimization process does not start from the mean of the model, but rather the median distance between ILM and RPE boundaries is found by the adaptive thresholding algorithm, as well as the ratio of the foveal pit distance to the greatest thickness found in the image. Using these values, the closest example from the training set is chosen and used to obtain the parameters for the initial model position. This way a faster and more robust convergence is ensured.

Another novelty is introduced in the second stage of the algorithm based on fitting a model for each independently used A-scan (depth-scan) to further improve the accuracy. This stage starts from the position defined by the result of the first stage B-scan fitting. The image area was heuristically divided into four segments and built an A-scan model for each segment since different types of variation can be expected at different offsets from the foveal depression. The A-scan model is trained on offsets produced by back projecting the manual segmentation data using the main B-scan model and computing the boundary offsets between the back projections and the original segmentations Eq. (8) (n is the number of layers and u is the number of A-scans from all the images in the given segment).

$$\mathbf{A} = \begin{pmatrix} aOff_{11} & \cdots & aOff_{1n} \\ \vdots & \ddots & \vdots \\ aOff_{u1} & \cdots & aOff_{un} \end{pmatrix} \tag{8}$$

In Figure 31 it can be seen how the second refinement stage of the algorithm improves precise tracking of the layer boundaries.

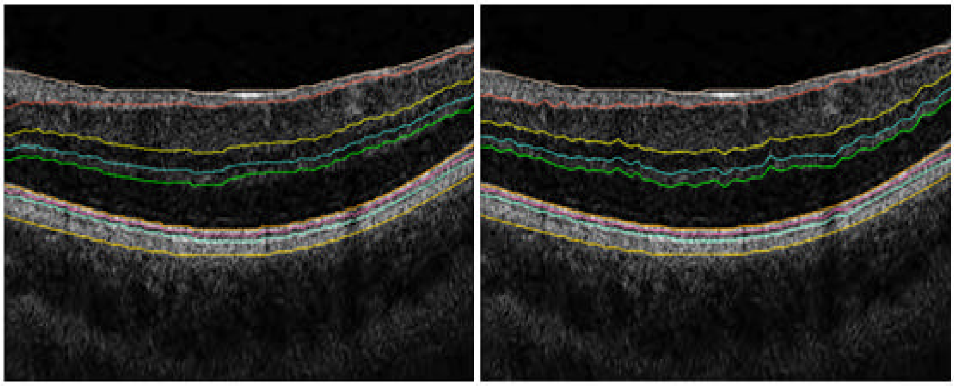


Figure 31 On the left is the result after the global low-res optimisation followed by, on the right, the refined result by the A-scan optimization.

7.2.4.4 Mechanical Turk

A large training data set has been efficiently obtained via an Amazon service called the Mechanical Turk (AMT), designed to offer a large international work force for completion of user defined tasks (Human Intelligence Task or HIT). Using this approach it was possible to evaluate the algorithm performance to a real golden standard, unlike the standard approach which can only evaluate repeatability of the algorithm in comparison to itself. In principle, the task of segmenting was divided to the detailed description of the task by a skilled person, manual delineation of the interfaces by a large number of less skilled workers, the comparison of multiple results for the same task and the supervision of the whole process by the skilled operator. Two account types are used: worker and requester. The worker account type is used for performing the tasks, while the requester type is used for submitting them. Submitted tasks are usually simple but it is possible to define criteria for the workers and in that way use skilled workers, for larger payments, of course. In this case no testing was performed for selection of the workforce apart from the general ranking of a worker based on previous performance recorded by the AMT-system. However, it was necessary to supervise the work relatively often and update the instructions based on the input from workers and give bonuses for good work to stimulate reliable workers to continue doing the provided tasks.

The architecture of the whole system is comprised of a web page with JavaScript to handle the user input that was designed through the AMT interface. Inside a Java applet was embedded through which the workers perform the segmentation. For storage of the B-scans, example images and results to be saved, Amazon S3 storage service was used. 505 B-scans were submitted and each image was set to be segmented twice by different workers respectively. That way, the inter worker variability can be computed and inaccurate results left out, while still having another one which is usually good. Inter worker variability was computed only on the images for which both results were deemed to be accurate. Bonuses were paid out for good work, approximately equivalent to the initial payment. In cases of inaccurate or inappropriate results it is not necessary to pay the worker. Since the behaviour of the AMT system can be better described by the rules of sociology than simple mathematical relations, the processing speed is nonlinear. It is important to note that while half of the results were obtained in just a few days; it usually took significantly longer to get all the tasks completed. That is not a problem since it is possible to use results as they are produced without having to wait for the completion of the whole batch. One most likely reason for the reduced work speed is that workers use the default sorting for viewing the available tasks, based on the number of available tasks. It took four weeks to complete the segmentation of about 2700 images. Workers seemed to be more interested in the task once the purpose of the work was given in the introduction and it was pointed out that it serves a valuable medical goal. A few questions were included in the form of a web form so that workers can give the feedback on the work that they were doing.

7.2.4.5 Results

For evaluation purposes 466 manually segmented B-scans were used, almost (in some cases the manual segmentation had to be discarded) uniformly sampled from 17 eyes (each stack contains 128 B-scans). The performance of the algorithm was tested on this data set using the leave-one-out test; all data from one person was iteratively left out, the model trained on the remaining data and then the performance tested on the data from the person left out. This procedure is performed for each person in the training set. This way it was ensured that the performance of the algorithm is tested on the “unseen” data.

For evaluation, automatic segmentation results were compared to manual segmentation done by the AMT workers. Two types of error measures were used, computed for each boundary *i* separately and from these error measures for an entire B-scan or for an individual layer were computed, Eq. (9).

$$^{i}E_B = \sum_{j=1}^{j=w} \left| yAut_{ij} - yRef_{ij} \right|,\ ^{i}E_{LDEV} = \sqrt{w * \sum_{j=1}^{j=w} (yAut_{ij} - yRef_{ij})^2} \tag{9}$$

E_B (Basic) is the basic error measure that defines the number of misclassified pixels. E_{LDEV} (Layer DEViation) uses the $\sqrt{w}$ term for normalization so that for the special case when the two boundaries are equally distant from each other along their whole length ($yAut_{ij} - yRef_{ij} = d$ for all *j*), it is equal to E_B (proved in Eq. (10)).

$$\begin{aligned} ^{i}E_B &= \sum_{j=1}^{j=w} \left| yAut_{ij} - yRef_{ij} \right| = \sum_{j=1}^{j=w} |d| = w * |d| \\ ^{i}E_{LDEV} &= \sqrt{w * \sum_{j=1}^{j=w} (yAut_{ij} - yRef_{ij})^2} = \sqrt{w * \sum_{j=1}^{j=w} d^2} = \sqrt{w * w * d^2} = w * |d| = {}^{i}E_B \end{aligned} \tag{10}$$

For all other cases E_{LDEV} is larger than E_B. Thus E_{LDEV} will penalize large deviations from the reference position of a boundary, unlike E_B which only measures the number of misclassified pixels. E_{LDEV} is therefore useful for penalizing specific types of poor algorithm performance which could show as, for example, a large jump in a boundary position that could be narrow and thus not affect E_B significantly since the misclassified area would be relatively small.

The error for a whole image (this refers to both E_B and E_{LDEV}) is defined in Eq. (11).

$$E = \frac{\sum_{i=1}^{i=b} {}^{i}E}{A} \tag{11}$$

${}^{i}E$ is the error for each boundary and A is the area between top (ILM) and bottom boundaries (RPE/CH).

In the case when the error for layer *k* is expressed separately, instead of summing up across all boundary errors, only the two boundaries that define a layer are added and divided by the sum of the layer area as given by the automatic segmentation (A_A) and the layer area as given by the reference segmentation (A_R), Eq. (12). This approach is used to normalize for double counting of misclassified pixels, as each layer is bounded by two boundaries.

$$E = \frac{\sum_{i=k}^{i=k+1} {}^{i}E}{A_A + A_R}, 0 < k < b \tag{12}$$

A confidence measure could be introduced based on the values returned by the objective function after the optimization step. Large values are proportional to the low confidence in the boundary positions determined by the model fitting. This would be useful for the operator to decide whether the obtained results are reliable.

In Table 1 the inter-worker variability of the manual segmentations used in training is presented for each boundary and in total, while in Table 2 and Table 3 results are presented for both the initial segmentation and after the second step refinement.

Table 1 Variability of manual segmentations on 75 B-scans in percent (the data has been previously examined and "bad" results left out)

Error Type	NFL	GCL +IPL	INL	OPL	ONL	CL	OS	RPE	Total
E_B	13.6	11.4	22.8	25.0	6.0	28.0	23.3	18.7	16.1
E_{LDEV}	17.9	14.4	28.4	31.3	7.4	35.4	28.6	22.5	19.9

Table 2 Error values on 466 B-scans at various positions from 17 eyes in percent before the A-scan optimization

Error Type	NFL	GCL +IPL	INL	OPL	ONL	CL	OS	RPE	Total
E_B	23.2	14.3	31.6	41.9	8.6	35.2	32.1	22.1	22.4
E_{LDEV}	31.9	17.4	39.7	55.4	10.7	47.3	41.0	27.0	27.8

Table 3 Error values on 466 B-scans at various positions from 17 eyes in percent after the A-scan optimization

Error Type	NFL	GCL +IPL	INL	OPL	ONL	CL	OS	RPE	Total
E_B	20.0	10.1	22.1	31.6	7.1	34.9	30.8	21.6	18.7
E_{LDEV}	29.2	13.2	30.4	46.4	9.3	47.1	39.5	26.5	24.2

It can be seen that the total error rates (especially E_B which is the main measure) are close to the inter-operator variability (18.2% compared to inter-operator's 16.1%). E_{LDEV} difference is somewhat larger. Thus, it can be concluded that the algorithm performance is almost the same as ground truth.

The algorithm performs well even when artefacts are present, such as strong shadows, which can cause problems for less robust algorithms (Figure 32).

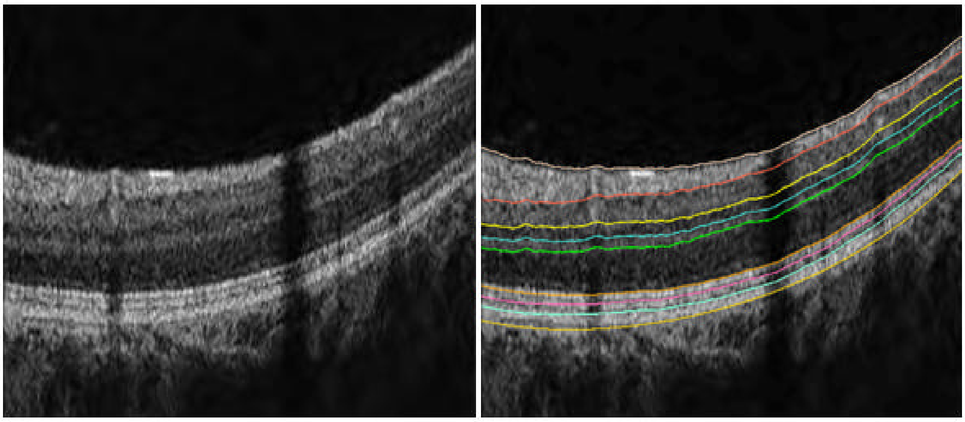


Figure 32 Robust performance for all the layers is achieved even in presence of shadowing. A despeckled image is shown on the left; the segmented image is on the right.

In Figure 33 thickness maps are shown for 17 different eyes after registering them and computing median and coefficient of variation (expressed as absolute variation in pixels), since it would take too much space to present the results for each eye individually. It can be seen that despite the data being affected by artefacts, the results are accurate and show larger variation only around the foveal pit region, as can be expected.

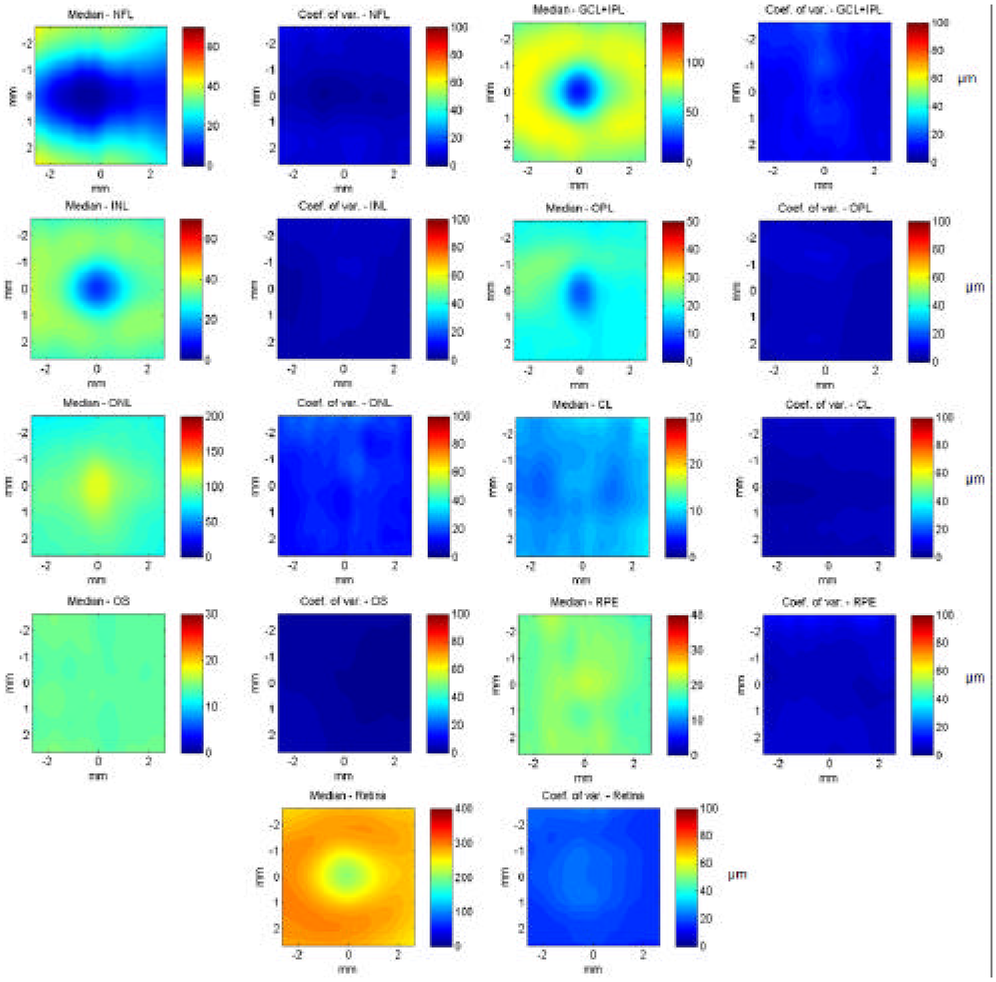


Figure 33 Median and coefficient of variation computed on thickness maps of all the individual layers (nerve fibre layer (NFL), ganglion cell layer and inner plexiform layer (GCL+IPL), inner nuclear layer (INL), outer plexiform layer (OPL), outer nuclear layer (ONL), connecting cilia (CL), outer segment (OS), retinal pigment epithelium (RPE)), as well as the retina, obtained from 17 eyes.

To evaluate the performance of the algorithm in conditions of increased noise (reduced dynamic range) that frequently occur in clinical measurements for a number of reasons (opaque cornea of cataract lens, residue in vitreous humour, non optimal imaging conditions, etc), background noise (speckle, multiplicative random noise) has been added to tomograms (Figure 34) and results plotted on a graph. The background was generated using a texture synthesis approach (Efros and Freeman, 2001). This enabled for efficient creation of a different speckle noise pattern for each image even though they are all based on the same physical speckle template, which is only one image of background noise with the typical spatial frequency distribution. Using this approach an arbitrary number of synthetic, but uncorrelated and realistic, images of background noise can be generated that are added subsequently to each given image to simulate low dynamic range. Using this novel approach it was possible to evaluate the performance in an arbitrary range of realistic speckle conditions. The algorithm shows robust performance under such conditions shown by a gentle rise of the error/dynamic range curve.

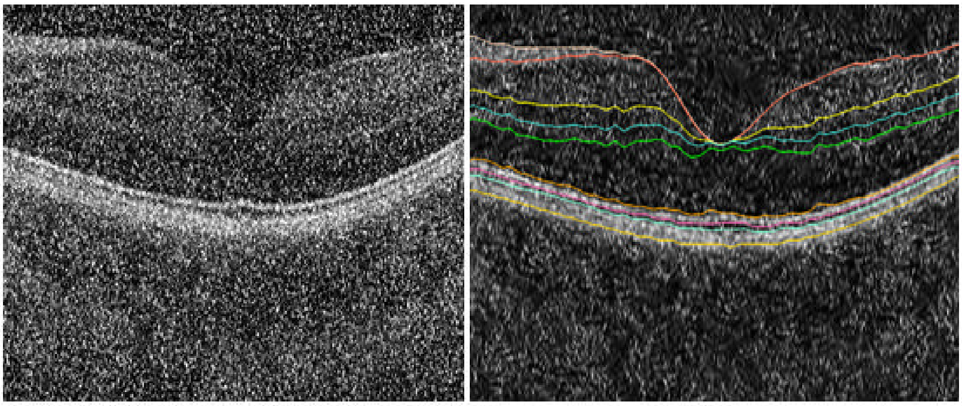


Figure 34 Segmentation in a case of added strong noise. Left original image. Right filtered, denoised image with segmentation results superimposed.

This can be seen in two graphs showing error rates E_B and E_{LDEV} plotted versus the dynamic range for a set of images for all the layers combined and with the confidence interval (1.96 std. dev.) plotted as dashed lines (Figure 35), as well as two graphs showing the error rates for each individual layer (Figure 36). The individual boundaries most affected by decreasing dynamic range are those defining INL and OPL, as could be expected since these layers exhibit normally significant variation and have weak boundaries which are affected early by the noise increase. Also, the boundaries between CL, OS and RPE are difficult to determine.

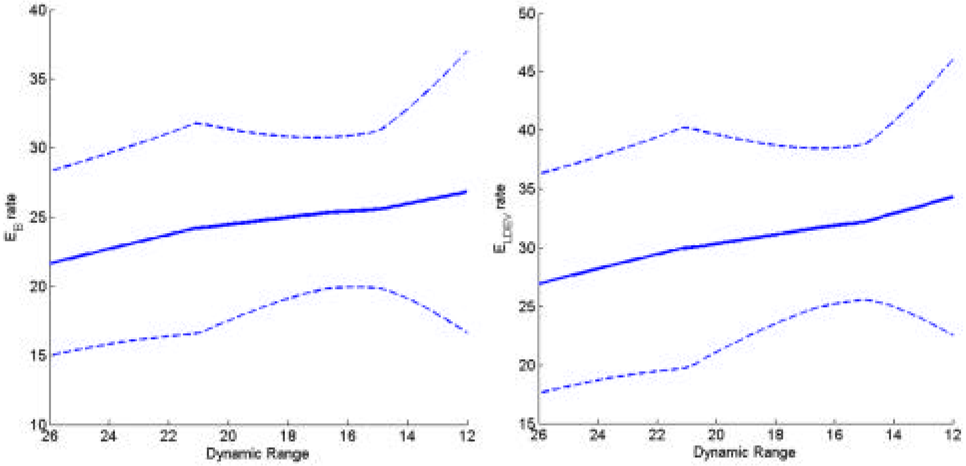


Figure 35 Error rates E_B (Basic) and E_{LDEV} (Layer DEViation) with decreasing dynamic range for all the data sets, with confidence interval (1.96 * standard deviation) marked by the dashed lines. For both error measures a slow rise in the error values can be observed, which guarantees robust performance with noisy data.

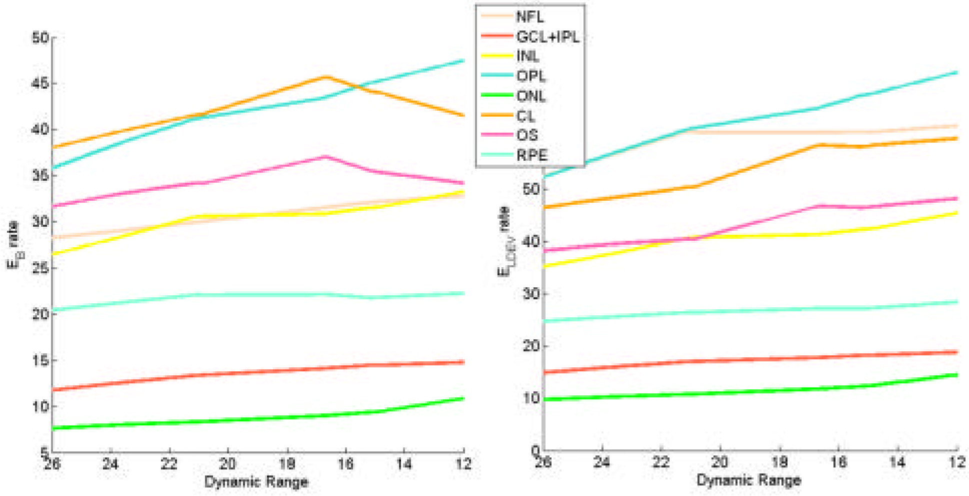


Figure 36 Error rates for the individual layers E_B (Basic) and E_{LDEV} (Layer DEViation) with decreasing dynamic range for all the data sets. For all the individual layers (nerve fibre layer (NFL), ganglion cell layer and inner plexiform layer (GCL+IPL), inner nuclear layer (INL), outer plexiform layer (OPL), outer nuclear layer (ONL), connecting cilia (CL), outer segment (OS), retinal pigment epithelium (RPE)) a slow rise in the error values can be observed. Thin layers inherently exhibit greater error values, as both errors are normalized by the layer area.

7.2.4.6 Conclusion

An algorithm for automatically segmenting all major retinal layers based on a novel statistical model has been developed. Two important novelties with respect to the standard AAM were introduced: a second term in the optimization function that penalizes large deviations from the three boundaries found by the adaptive thresholding algorithm and the second algorithm stage that refines the model fit for each A-scan independently, giving increased accuracy.

It has been thoroughly tested and evaluated against a manually segmented large dataset from a 800 nm OCT system. The algorithm proved highly robust in full foveal scans even in the presence of artefacts and added strong background noise that reduce dynamic range down to 12dB. It is the first time that a large, representative data set (466 B-scans from 17 eyes) has been used for evaluation of an OCT segmentation algorithm. Manual segmentations of the large data-set were used as ground truth, rather than the frequently used error computed between the results of the algorithm on inter-visit measurements, as it is susceptible to underdetermine the real error value, likely ignoring systematic error of the algorithm. Apart from the basic error measure that counts the number of the misclassified pixels, a second error measure was used to penalize large deviations from the ground truth.

It can be concluded that the algorithm successfully demonstrated reliable performance under conditions which prove extremely challenging for the pre-existing methods. It was used as a basis for the development of the model based choroid segmentation algorithm described in 2.3.2, proving that indeed it is well suited to the weak and missing data conditions. It would be also possible to extend the proposed algorithm to segmentation of pathological cases, as well as segmentation of ONH (optic nerve head) scans which contain discontinuous boundaries. In cases when the stack registration is very precise, the initial ILM and RPE boundary finding step could be replaced by the algorithm proposed by Fabritius et al. (Fabritius et al., 2009b) that relies on full 3D information present in the stack, since it is very efficient. Clinically, fully automated segmentation of all major layers is essential in making medically useful the possibilities given by the method of high resolution, high speed OCT of large portions of the human retina at microscopic detail.

7.3 Choroid segmentation

For segmentation of choroid affected often by extreme pathologies, it was necessary to develop a more robust top and bottom boundary detection algorithm. subchapter 2.3.1 presents how it works and why it is applicable to a large problem area outside OCT. In subchapter 2.3.2 choroid segmentation is explained using a statistical model similar to the one developed for the foveal segmentation (2.2.4)

High-speed 3D OCT-imaging at 1060 nm was performed with less than 2.5 mW at the cornea, well below the maximum power limit for 10-second exposure. Three-dimensional OCT volumes were acquired with 15 to 20 µm transverse resolution, approximately 7 µm axial resolution, and 512 voxels per depth scan (A-scan). Scans across a 36° by 36° ?eld with 512 by 512 A-scans at 47,000 A-scans/s were centred on the fovea and resulted in to 70 frames/s (B-scans/s). The spectrometer utilized a line camera with 1024 px (Goodrich SU-LDH2) clocked at 92 kHz linerate, which led to 512 voxel in each depth scan between the zero delay and the maximal depth. The signal-loss of -14dB along the depth scan was counteracted by placing the closer, less attenuated side of the tissue towards the zero delay. For registration and noise reduction ImageJ software was used. The OCT volume was averaged in both transverse directions within a ?eld of approximately 1°, to remove speckle and increase sensitivity. Axial choroidal thickness (ChT) was defined as the distance between the centre of the signal peaks originating from the RPE/Bruch’s membrane/choriocapillaris (RBC) complex and the choroidal–sclera interface (CSI).

The pre-processing stage is performed for both the training step and the segmentation of the unseen data. Once the variation parameters have been learned from the manually segmented training data, they can be used to drive the model to perform segmentation of unseen data. The actual segmentation process is essentially an optimization run that adjusts the model parameters in order to minimize the objective function that defines the difference between the model and a given unseen image that is to be segmented. The overview is shown in Fig. 37.

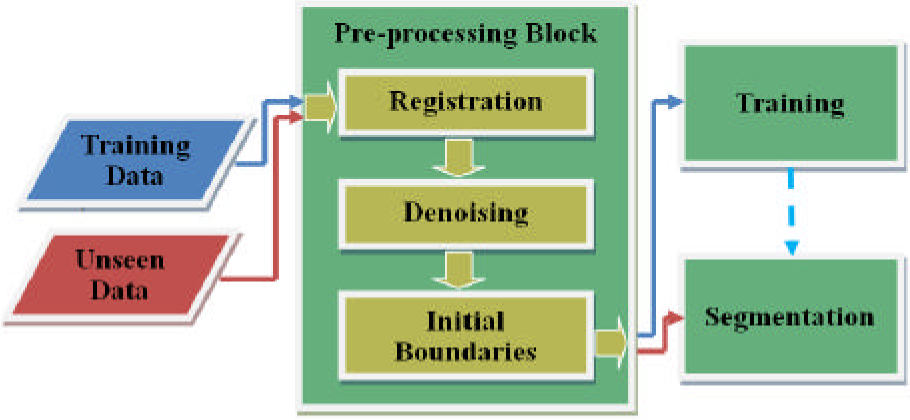


Fig. 37 Algorithm overview: both the training and unseen data pass through the pre-processing block. The registration is used only for averaging in z-direction as a despeckling procedure, while the segmentation is independent for each B-scan. A statistical model is subsequently constructed that captures the variance in the training data, which can be then used (dotted line) to segment unseen data.

7.3.1 Choroid initial boundary detection

To successfully segment internal limiting membrane (ILM) and RBC boundaries of choroidal data several challenges have to be overcome. RBC defines the top choroidal boundary and is used as initialization for the statistical model that subsequently finds the CSI (choroidal sclera interface) boundary. ILM is relatively easy to find and is of clinical use, thus it is segmented as well.

The RBC is usually almost at the same position as RPE, but in cases of RPE detachment it is often not. The OCT data obtained was from a 1060 OCT system by scanning over wide angle, which causes low signal strength on the image edges, especially towards the stack start and end, where the retina is narrow (Fig. 38 a), the right side of the image). Another problem were various pathologies that present different challenges; the boundaries can be discontinuous, also their shape is highly unpredictable and of highly changing curvature, so it is impossible to efficiently parameterize them, using a small set of parameters, with polynomials or other models that can be used for fitting. The third problem is that the data itself is often coming from different OCT systems and pre-processing pipelines; obviously it is highly beneficial if the correct segmentation can be achieved without retuning the algorithm parameters, thus a robust approach is required.

To meet all these goals an algorithm has been devised that consists of several basic steps, however special care was taken to keep the number and importance of free parameters as small as possible. This makes the algorithm intuitive and general while avoids over specializing.

First non-signal areas at the edges of an input image (Fig. 38 a)) are found by performing Canny edge filtering using large sigma (32) and high threshold (0.35). The parameters were heuristically chosen to extract only the major edges in an image. Columns on the sides where there is no edge detected are considered to be non-signal areas. The first column with an edge detected is considered to be the start of the signal area; this is determined both from the left and the right side (Fig. 38 b)).

A modified Dijkstra's algorithm is used to find a minimal path from the left to the right edge of an image, representing the RPE boundary, as it is almost always the brightest and longest one (in the sense that it extends furthest to the sides of the image). It is important to note that this algorithm has been used only to find the RPE which is a high intensity boundary, unlike some previous approaches which used path finding algorithms to find all the retinal boundaries, making the segmentation sensitive to any changes in the input data.

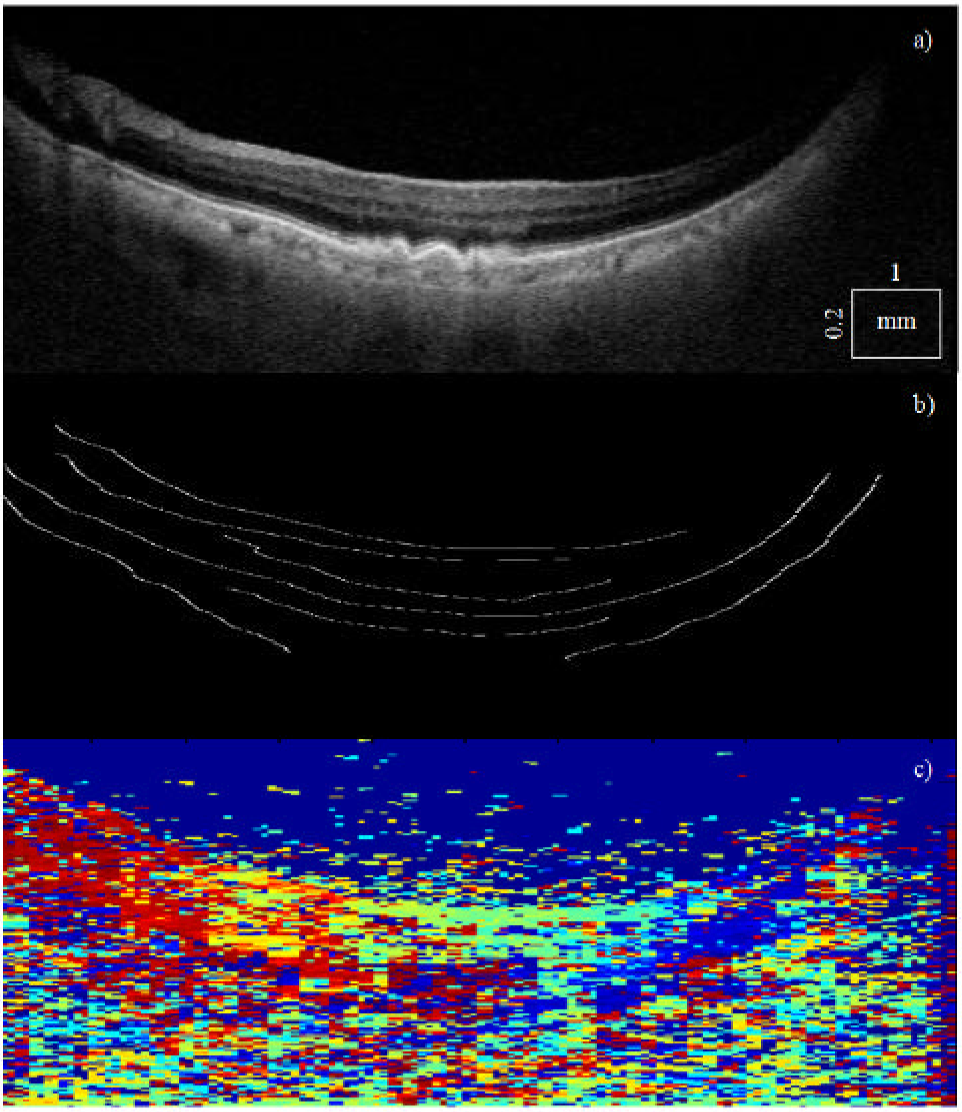


Fig. 38 Pre-processing steps for finding ILM and RBC: a) is the original image, b) is obtained after Canny edge filtering, c) are colour coded dominant wavelet domain orientations (from dark red "descending" (-90deg) angle, to dark blue "ascending" (90deg))

A graph is constructed, where each node represents the inverted pixel intensity in an image (i.e. the dark pixels become bright and bright pixels dark). Since the path to be found needs to be minimal, traversing bright pixels will mean going through low intensity values in the inverted image. A diagonal block connectivity matrix **A** consists of $m*n$ rows, where m is height and n is width of an image **I**. Thus, every row represents connectivity of certain pixel to pixels in the next column of the original image. Considering that the boundary cannot track backwards or skip columns, matrix **A** has only m columns, as each pixel is connected only to the next column of the original image. In this case it was assumed that plausible steps from column to column do not exceed 10 pixels, thus *jmp* was set to 10, Eq. (13). Vector **c** represents connectivity of a particular node (pixel) to other nodes and it is defined within the *jmp* range, with the rest of pixels set to 0 (no connection).

$$\mathbf{A} = \begin{pmatrix} \mathbf{c}_{11} & & & & 0 \\ \vdots & \ddots & & & \\ 0 & \mathbf{c}_{1m} & & & 0 \\ \vdots & & \ddots & & \\ 0 & & \mathbf{c}_{n1} & & 0 \\ \vdots & & & \ddots & \\ 0 & & & & \mathbf{c}_{nm} \end{pmatrix}$$

$$\mathbf{c}_{ij} = 1 - \mathbf{I}(x, j+1), \max(0, i - jmp) < x < \min(n, i + jmp) \qquad (13)$$

To increase robustness of this approach it was expanded in two ways. Firstly, derivatives were added to prevent unrealistic boundary jumps caused perhaps by just slightly more favourable intensity information.

Second, a complex dual-tree wavelet decomposition, being multi-scale, orientation sensitive and translation invariant, making it more efficient than standard edge enhancement methods, such as Sobel, Prewitt, Roberts or Canny, was used to obtain the edge orientation information (Fig. 38 c). Smooth dominant directions were calculated from -90 (dark red) to 90 (dark blue) degrees, which was applied in the modified algorithm to obtain the third term expressing how well the path is aligned with the real edges in an image. This prevents path drifting over high intensity areas that possibly have completely different orientation, thus producing a non-realistic path.

Unlike working with static weights and having a complicated workaround the problem that each node can be part of multiple paths and thus cannot be assigned an optimal weight using static values (Perkon et al., 2010), the problem has been solved by making an elegant extension to the optimal path algorithm itself, while preserving the same time complexity.

The Eq. (14) shows the modified algorithm distance function.

sc is the current node index, corresponding to any pixel in the original image, *xc* and *yc* are x and y coordinates obtained of the node *sc*, **y** are the neighbouring column's connected nodes (pixels), *yp* is *sc's* parent's y value (inherently always known from the path history), $\mathbf{y}''$ is the second derivative of the potential path, while $\overline{\mathbf{y}'}$ is the derivative mean used as an approximation to obtain the angles **j** of the potential path. Weights $\mathbf{ew_1}$ and $\mathbf{ew_2}$ indicate how close to each of the two neighbouring out of 6 total discrete orientations possible path is. **IW** is the tensor containing directional wavelet coefficients (amplitudes) for each of the six directions, that can be index using the potential path directions **j**. *n* is the number of wavelet levels used, for this application it was set to 4.

$$d(sc) = w_1\mathbf{A}(sc,\mathbf{y}) + w_2\left|\mathbf{y}''\right| + w_3\left(1 - \frac{\mathbf{ew_1} \cdot \sqrt[n]{\mathbf{IW}(yc, xc, \mathbf{j}_1)} + \mathbf{ew_2} \cdot \sqrt[n]{\mathbf{IW}(yc, xc, \mathbf{j_2})}}{\sum \sqrt[n]{\mathbf{IW}(yc, xc, f_i)}}\right)$$

$$xc = \left\lfloor \frac{sc-1}{M} \right\rfloor$$

$$yc \equiv sc\text{-}1 \pmod{M}$$

$$\mathbf{y} = \mathbf{A}(sc, i) > 0$$

$$\mathbf{y}'' = (\mathbf{y} - yc - (yc - yp)) = (\mathbf{y} - 2 * yc + yp)$$

$$\overline{\mathbf{y}'} = \frac{\mathbf{y} - yc + (yc - yp)}{2} = \frac{\mathbf{y} - yp}{2}$$

$$\mathbf{ew}_1 = \frac{\mathbf{j}_2 - \mathbf{j}}{\mathbf{j}_2 - \mathbf{j}_1}$$

$$\mathbf{ew}_2 = \mathbf{1} - \mathbf{ew}_1 \qquad (14)$$

The weights *w* of the distance measure's three terms have been set heuristically to 1, 0.08 and 0.2 respectively; the algorithm is robust enough that the exact weight values are not very important. For some other application another values might be more appropriate. Having the distance measure defined between each pair of nodes in the input image, the task of finding the optimal path is simply that of finding a minimum sequence of distances from end to end.

After the initial estimate of RPE boundary has been found (Fig. 39 a)), a convex hull is fitted to the points that comprise the RPE boundary (Fig. 39 b)). A formal definition of the convex hull for a set of points X is the minimal area convex set containing X.

Fitting the convex hull allows for an efficient estimation of which regions should be interpolated over. The convex edges are always left intact, while concave regions (intuitively dents) are considered for interpolation. Following the efficient representation of the problem allowed for a simple statistical measure to be developed. In general, any segment which has a high boundary curvature, or is significantly shifted from the estimate of the convex shape is likely to be caused by pathological detachment of the RPE (i.e. drusen). Thus, an assumption can be made that the position of the RPE does not mark the start of the choroid, while in most cases a cubic interpolation gives a good estimate of the choroid. While the RBC estimate produced by the interpolation over these non convex regions is quite close to the boundary determined by experts, it is not always completely correct. For the simplicity and rarity of these occurrences, the algorithm was found to be performing sufficiently well even with simple interpolation. However, a more complex approach can be devised. After finding the cubic spline estimate of the start of the choroid region, a neighbourhood of the spline can be considered for a second run of the path finding algorithm. This will remove the area where the detached RPE is located, allowing for a more precise boundary estimate. That boundary is often very weak, and in some cases it is even marked by a black edge rather than a white one. Thus, for that run of the path finding algorithm, a more complex edge detector would have to be considered, instead of simply using intensity values. A straight forward approach would be to use a trained texture descriptor to obtain edge probabilities. If encoded as an intensity image, the same shortest path procedure could be subsequently applied.

Once the RBC boundary has been found, the neighbourhood and the image areas below it are removed (current position-50 pixels, to the bottom of the image). In the next step this image is used to find the real ILM position. Columns in which no edge was found signify gaps in the ILM, as the removal of the image signal around and below the RPE will result in no edge detection in the whole column (ILM and RPE being the brightest boundaries). The final ILM boundary is then found using piecewise cubic spline interpolation.

At the non-signal areas of the image, RBC is set as straight, while ILM is set to the value of the RPE (Fig. 40). Additionally, the program returns the non-signal area positions, so that the statistical model can use it later to find the choroidal-scleral interface, as at the non-signal areas of the image that boundary is set to the RBC value as well and is excluded from the optimization run (thus noise at the sides does not confuse the optimization function).

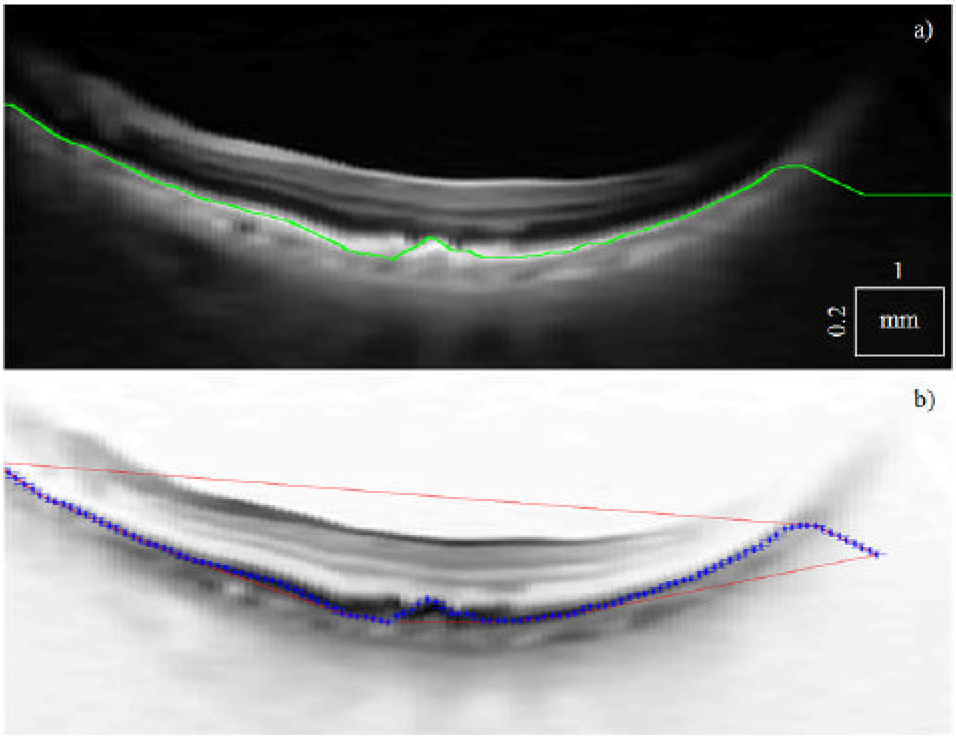


Fig. 39 RPE estimate (green line and blue dots) and convex hull (red line)

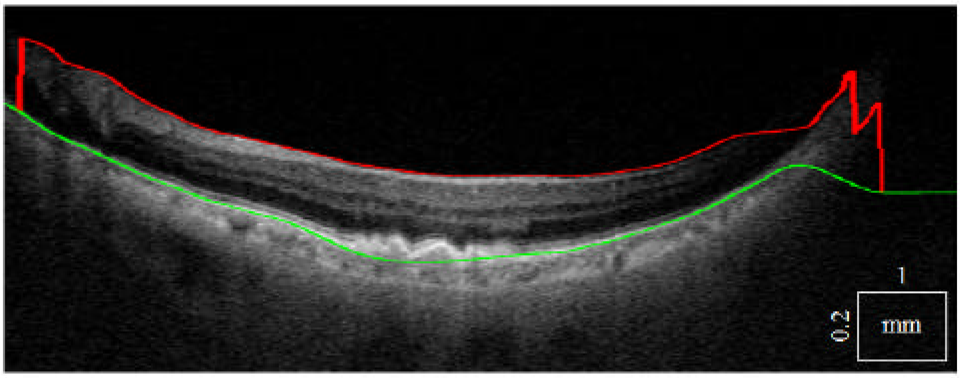


Fig. 40 The final result: ILM (red) and RBC (green)

7.3.2 Model Based Approach to Choroid Segmentation

Choroidal segmentation is an ideal application for the proposed statistical model, as the choroid is notoriously difficult to detect using standard approaches that have been applied so far to the foveal segmentation. There are three main reasons for that:

- the choroid is usually deep underneath the retina and thus OCT signals can degrade significantly, especially if the choroid is thick due to multiple scattering
- the CSI, its lower boundary is, unlike that of foveal layers, often very weak or in some locations invisible
- the texture is inconsistent comprising of large and small vessels

Essentially the choroid is more defined by its various vessel features as a textured region, rather than an area clearly separated by a strong boundary. All these factors make it an ideal application for a statistical model. The statistical model already assumes certain shape information and is thus far likely to perform poorly due to, at segments missing boundary and weak signal. Secondly, it considers the area bounded by the lower boundary being optimized, unlike many other segmentation methods that only try to track or in other ways segment the boundary itself without any knowledge about the area that that boundary defines.

The statistical model relies on initial boundary algorithm that defines the ILM and RBC and provides the offsets for the non-signal area of an image. The task for the model is then to find CSI boundary (the layer between the RBC and CSI). ILM is not used by the model, but is clinically useful to compare produced choroid thickness maps to the foveal shape.

The pre-processing and statistical model construction were done similarly to the procedure described in (Kajic et al., 2010), thus only algorithm elements that differ will be described in more detail. There are two distinctions. The first one is that the stacks are not registered by finding the position of the foveal pit, as it would be too unreliable to determine the exact location utilizing the same method as for the normal retinal segmentation. Dealing with pathologies means that there is no foveal pit, that part can be a bulge (i.e. AMD). Thus the stacks are simply used as they are, which does not cause any problems since the input to the shape training model is simpler, consisting of only one boundary (RBC), unlike with the foveal segmentation which had many more. The second difference is that the texture information is not included in the statistical model; texture is however used in form of a blob detector, but how that information is used by the objective function is fixed.

Given the RBC position the model is fitted to the image area below the RBC. A simple objective function is given: it tries to optimize the ratio of area covered by blobs (choroid being an area containing many vessels) versus the total area. Blobs are computed using a maximally stable extremal regions method (MSER (Matas et al., 2004)). An extremal region Q is a region for which all the elements are either larger (maximum intensity region) or smaller (minimum intensity region) than the neighbouring elements (pixels).

The concept more simply can be explained by thresholding. All the pixels below a given threshold can be considered black and all those below or equal white. In a progressive sequence of thresholded images the first one would be completely white, then black spots corresponding to local intensity minima would appear, then grow larger and eventually merge, until the whole image would be black. The set of all connected components in the sequence is the set of all extremal regions. A formal definition of maximally stable extremal region is given below:

Let $Q_1,..,Q_{i-1},Q_i,...$ be a sequence of nested extremal regions $Q_i \subset Q_{i+1}$. Extremal region Q_i is maximally stable if and only if $q(i) = |Q_{i+\Delta} \setminus Q_{i-\Delta}| / |Q_i|$ has a local minimum at i. Δ is a parameter of the method. Intuitively it can be understood as a region that stays stable through a relatively large number of thresholds, i.e. it is a blob with a strong boundary, making it suitable for vessel cross-section detection.

Over a large range of thresholds, the local binarization is stable in certain regions, and has the properties listed below:

- Invariance to affine transformation of image intensities
- Covariance to adjacency preserving (continuous) transformation on the image domain
- Stability, as only regions whose support is nearly the same over a range of thresholds are selected.

- Multi-scale detection without any smoothing involved, both fine and large structure is detected

For this application the last two properties are important. The MSER implementation used is part of VLFeat, cross-platform open source collection of vision algorithms (Vedaldi and Fulkerson, 2008). It bundles a MATLAB toolbox, a portable C library and a number of command line utilities. Several parameter that the MSER function takes will be explained:

- Delta: already mentioned in the MSER definition
- MaxVariation: defines the maximum variation (absolute stability score) of the regions
- MinDiversity: the minimum diversity of the region. When the relative area variation of two nested regions is below this threshold, then only the most stable one is selected
- MinArea and MaxArea: these parameters define the minimum and maximum allowed areas of detected regions as percentage of the total area of an image

For this application the parameters were heuristically chosen as follows: 1, 0.25, 0.7, 10/imageArea, 500/imageArea. Given the RBC boundary and a MSER filtered image $F_{MSER}(\mathbf{I})$, the optimizer, using the objective function, looks for the minimal value by changing the shape of the lower boundary. The objective function is defined as:

$$f(\mathbf{s}) = -\frac{G_{CH}(S(\mathbf{s}), F_{MSER}(\mathbf{I}))}{\mathbf{A}_{CH} + \mathbf{A}_{PCH}} \tag{15}$$

It maximizes the ratio of the choroidal area covered in blobs against the total area of the choroid (A_{CH}) and the post choroidal region (A_{PCH}), taken to be an area from the choroid boundary plus 50 pixels, thus it is a constant. For the model initialization the mean of the model was used, which, having a robust objective function, proved to be sufficient. The A-scan optimization seeks black to white transition, given the fact that the choroid is usually darker than the post-choroidal region (though this boundary is often very weak or non-existent):

$$f_A(\mathbf{s}_A) = G_{DW}(S(\mathbf{s}_A), \mathbf{I}) - G_{AW}(S(\mathbf{s}_A), \mathbf{I}) \tag{16}$$

G_{DW} gives the average intensity of the choroidal region, CSI boundary minus 50 pixels, while G_{AW} of the post choroidal region, CSI plus 50 pixels. The final boundary is then filtered using a 1D median filter of size 11 pixels, with the size set heuristically.

7.3.2.1 Results

For evaluation, automatic segmentation results were compared to manual segmentation performed by an experienced examiner. A basic error measure was used, computed for a boundary *i* separately (in this case there is only one layer and two boundaries: RBC and CSI) and from these, error measures were computed for an entire B-scan or for an individual layer, Eq. (9). E_B (Basic) is the basic error measure that defines the number of misclassified pixels.

$$^{i}E_B = \sum_{j=1}^{j=w} \left| yAut_{ij} - yRef_{ij} \right| \tag{17}$$

To compute error for layer *k* separately, only the two boundaries that define a layer are added (offsets) and divided by the sum of the layer area as given by the automatic segmentation (A_A) and the layer area as given by the reference segmentation (A_R), Eq. (12). This is used, instead of summing up across all boundary errors, to normalize for double counting of misclassified pixels, as each layer is bounded by two boundaries.

$$E = \frac{\sum_{i=k}^{i=k+1} {}^{i}E}{A_A + A_R}, 0 < k < b \tag{18}$$

For clinical application, it would be useful for the operator to decide whether the obtained results are reliable. As part of future work, a confidence measure could be introduced, based on the values returned by the objective function after the optimization step. Large values are proportional to the low confidence in the boundary positions determined by the model fitting.

In Table 4 results are presented for the 7 diabetes type 1 eyes. In general, diabetes type 1 choroid is close to the choroidal shape and structure of healthy eyes. Error rates are even lower than what was obtained with retinal segmentation (Kajic et al., 2010). In Fig. 41 a typical B-scan is shown from a diabetes type 1 patient, before and after the A-scan optimization. Notice the good performance even in the central region where the signal is weak.

Table 4 Error values on 7 eyes with diabetes type 1, expressed as percentage of misclassified pixel relative to the layer area. The total average value is computed as an average of all the patient error values, and is independent of the number of B-scans per stack

Error Type	A (100 B-scans)	B (100 B-scans)	C (100 B-scans)	D (98 B-scans)	E (100 B-scans)	F (93 B-scans)	G (100 B-scans)	Total
E_B	14.1	9.2	19.4	15.8	13.4	15.7	16.1	14.8

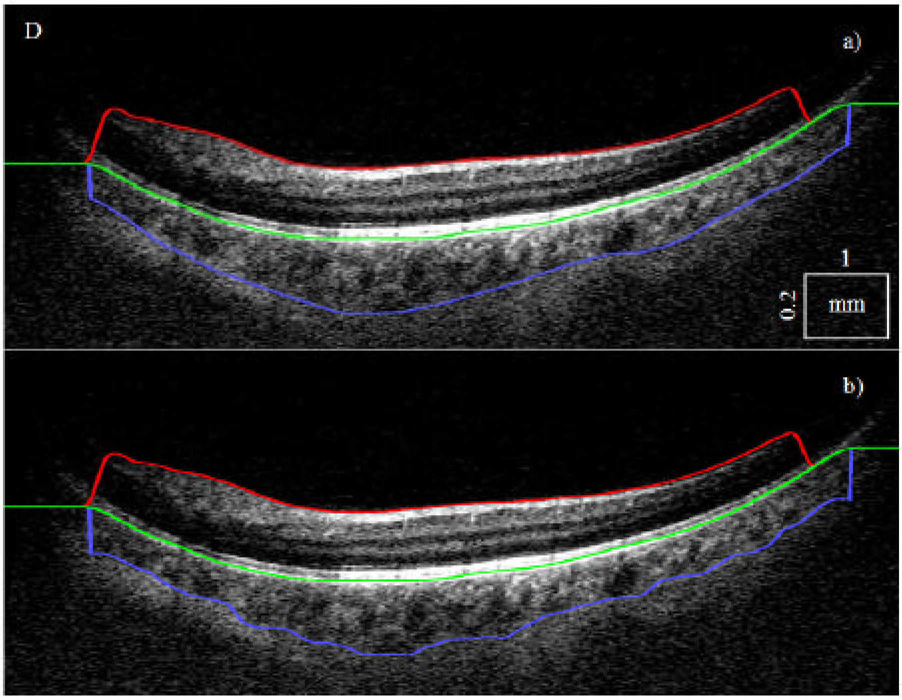


Fig. 41 A slice from a typical diabetes type 1 eye (patient D), before (a)) and after (b)) the A-scan optimization. ILM (red), RBC (green), CSI (purple)

In Table 5, 5 different pathologies are evaluated. The results are, surprisingly, even better than, close to normal, diabetes type 1 eyes (Table 4), especially considering the difficulty of segmenting such varied cases. One likely reason for that is the fact that the five pathologies were pre-processed slightly differently, with an average filtering of neighbouring B-scans performed after the registration (to remove speckle). Another reason is that the manual segmentation was performed in the B-scan range of 20 to 100 (every fourth, 20 B-scans in total), and not 15 to 114 (diabetes type 1 patients, 100 B-scans in total), thus the error rates that are larger in slices (B-scans) close to the start and end of stack, where the signal is lower, contributed to the larger error value of diabetes type 1 data segmentations.

Table 5 Error values on 5 eyes with various pathologies, about 20 slices manually segmented per stack, expressed as percentage of misclassified pixel relative to the layer area. The total average value is computed as an average of all the patient error values, and is independent of the number of B-scans per stack

Error Type	H (20 B-scans)	I (20 B-scans)	J (100 B-scans)	K (20 B-scans)	L (20 B-scans)	Total
E_B	5.6	12.8	10.1	13.0	15.3	11.3

Case H is an eye with a central serous chorioretinopathy. A lesion is present that affects the outer retina above the RBC complex. A deep, difficult to delineate, choroid is another characteristic of the pathology. This eye does not present a serious challenge for RBC segmentation (only slight RPE thinning in the middle), but determining CSI is difficult as the signal becomes very weak due to the thickness of the choroid. In

Fig. 42 a representative B-scan is shown after the final A-scan optimization.

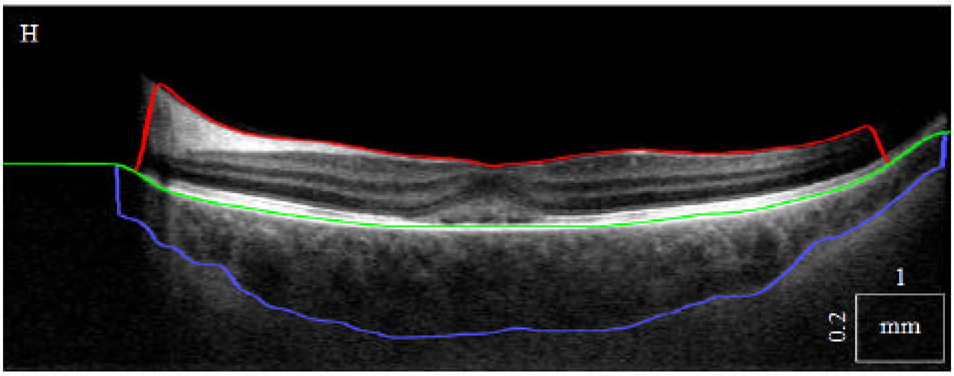


Fig. 42 The algorithm performs well despite the thinning of the RPE in the middle and deep, low signal choroid

Case I is a terminal wet age related macular degeneration (AMD). The RBC complex has an irregular shape and thickness; however the dynamic programming approach that was introduced, performs well. In Fig. 43 a representative B-scan is shown after the final A-scan optimization.

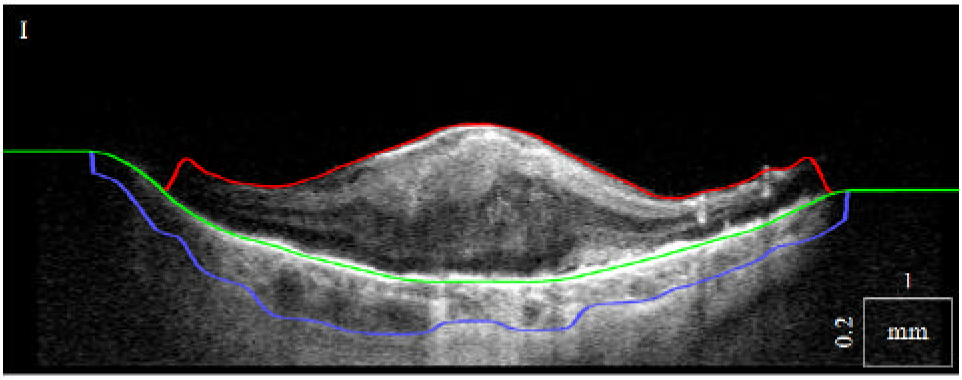


Fig. 43 The choroid is very thin and has a strong variation in contrast

Case J is an eye with RPE atrophy as a result of dry age related AMD with a strong signal in the choroid due to reduced melanin concentration in the RPE. The choroid is extremely thin. The bright area in the choroid presents a challenge for RBC detection as well as for the statistical model (determining CSI), as the choroidal structure is significantly different from a normal choroid. However, as can be observed, the model is general enough to perform well even in that case. In Fig. 44 a representative B-scan is shown after the final A-scan optimization.

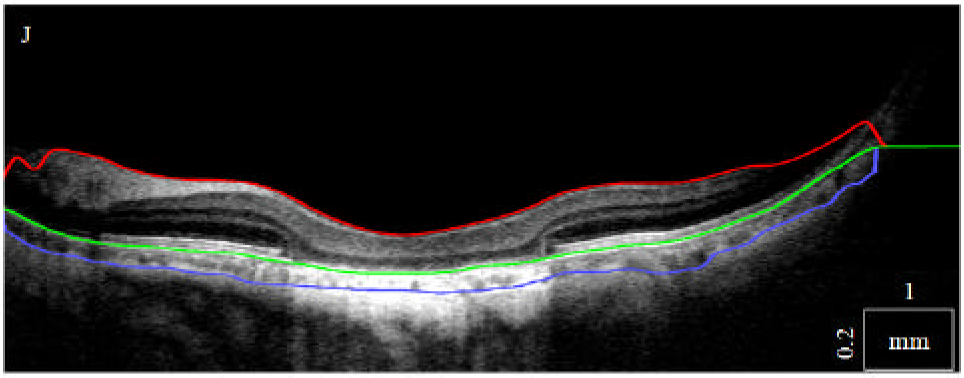


Fig. 44 RPE/RBC and CSI are found despite the extremely thin choroid with a strong signal in the middle as a result of the RPE atrophy

Case K shows an eye with deposits underneath the RPE (drusen). The RBC complex has an unusual inverted shape. A simple parametric approach for representing RBC, such as low order polynomial, would fail in this case. The proposed algorithm still produces reliable results. In Fig. 45 a representative B-scan is shown after the final A-scan optimization.

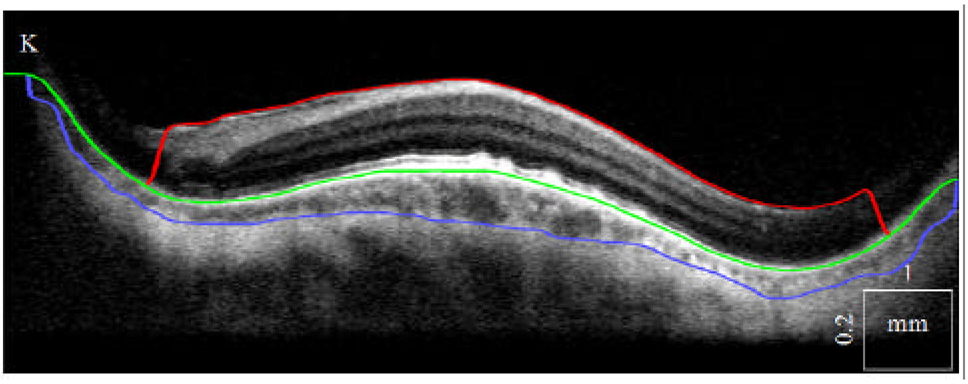


Fig. 45 RPE and RBC have an uncommon "inverted" shape

Case L is another AMD with a very difficult RBC to delineate. In Fig. 46 a representative B-scan is shown after the final A-scan optimization.

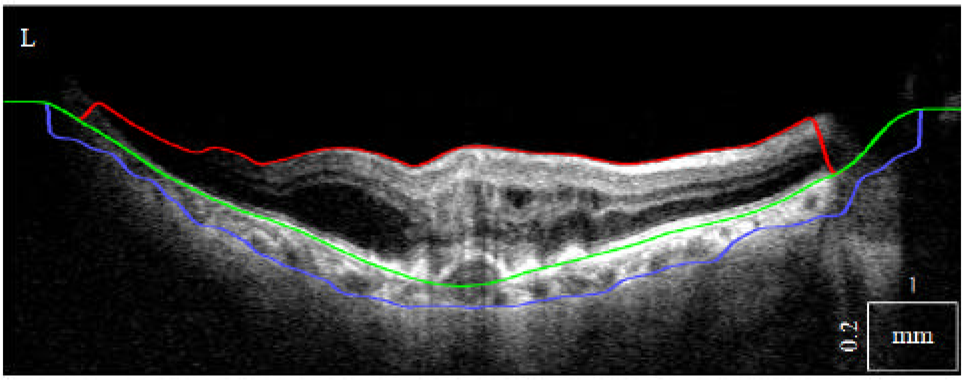


Fig. 46 The algorithm correctly interpolated over the drusen, finding a close approximation of the RBC

In (Kajic et al., 2010) we collected large, manually segmented datasets via Amazon Mechanical Turk (AMT) service. Since the segmentation of choroid, and especially pathologies, required too much expertise to use AMT, there is no inter-observer variability measure. By comparing the obtained error rates to the foveal inter-observer variability it became obvious that the error rates are within that range. The average of the total error values from Table 4 and Table 5 is 13%.

In Fig. 47 six thickness maps are presented. For brevity, only one diabetes type 1 eye is shown, as the other diabetes type 1 eyes are rather similar. However, all 5 pathologies have been included. Due to almost non-existing signal at the start and end of the stacks (volumes) sometimes a few of the first and last B-scans were left out.

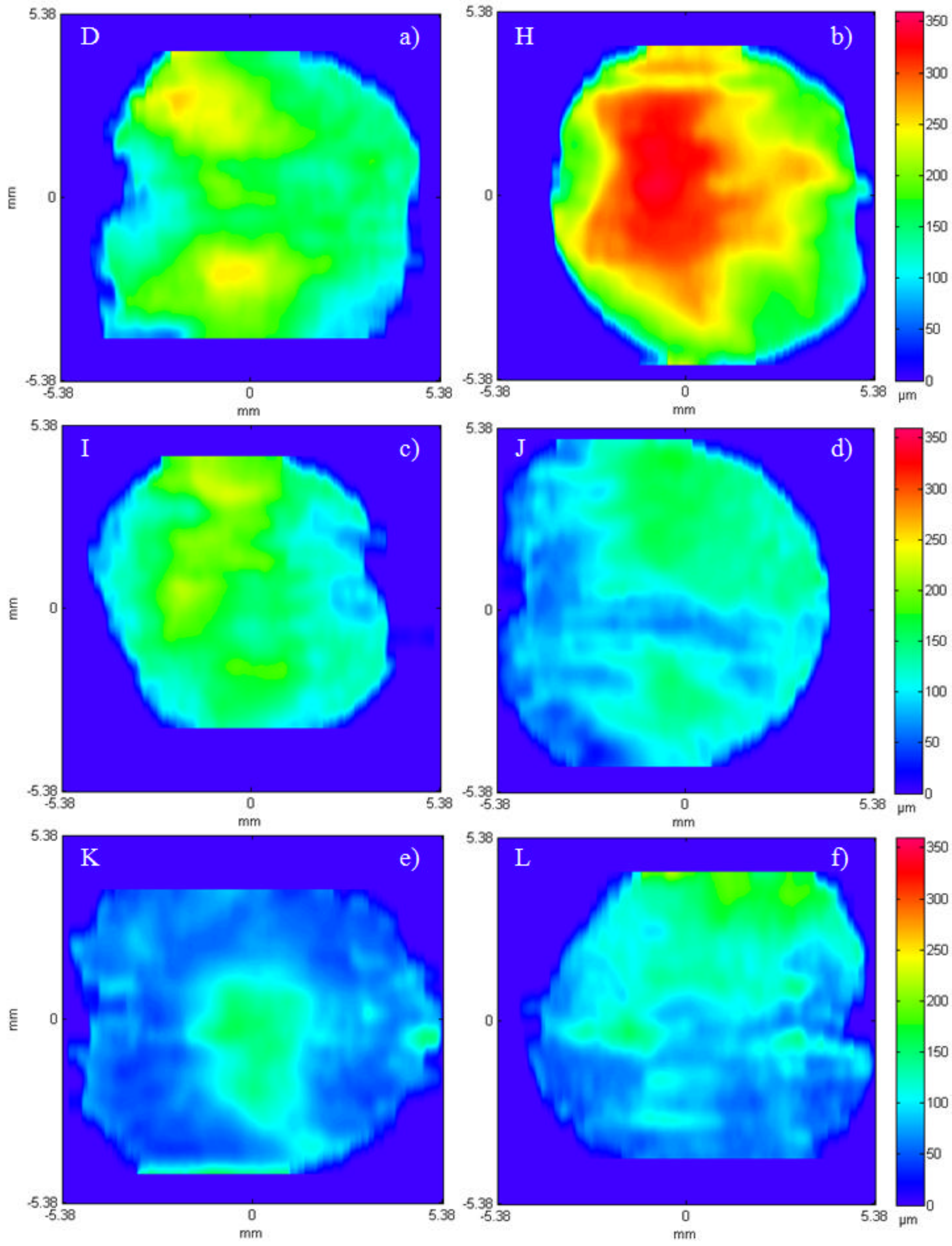


Fig. 47 Choroidal Thickness maps of a diabetes type 1 patient (D, subfigure a)) and 5 pathologies (H-L, subfigures b) to f))

The model based approach uses the variation obtained from the training set and imposes those constraints when segmenting an unseen image. This ensures that the segmentation will likely be close to the ground truth and less sensitive to noise. However, it is extremely important to have a large, representative training set that includes all possible variation. Overall the novel algorithm segments choroid accurately as compared to the human expert segmentation, with the error rate 13%. In case that a precise and robust stack registration algorithm is used in the pre-processing stage, the algorithm could be implemented in full 3D and would be even more robust and accurate.

The current processing time is about 30 seconds per slice, using non-optimized MATLAB implementation. For this application, close to real time processing is not essential as the results obtained are used for subsequent scientific analysis to obtain trends in various choroidal conditions. For clinical use, however, using optimized C implementation and a trivially parallelized approach, this could be brought down to approximately same time for the whole stack, using commodity hardware such as a GPU card.

7.3.2.2 Conclusion

An algorithm for automatic segmentation of the choroid based on a statistical model has been developed. Before applying the statistical model, the top choroidal boundary (RBC) has to be determined, as it used to initialize the model. A novel method was developed for that purpose according to high robustness requirements in order to work well in cases of severe pathologies and signal degradation. It is a shortest path dynamic programming method that was extended to use the path derivative information and edge orientation with the same time complexity. It is also applicable to a general class of computer vision problems.

The choroidal automatic segmentation algorithm has been thoroughly tested and evaluated against a large, manually segmented dataset obtained from a 1060nm OCT system and proved highly robust in wide-field scans even in the presence of artefacts and pathologies. It is the first time that a large, representative choroidal data set has been automatically segmented. The algorithm successfully demonstrated reliable performance under difficult conditions. Clinically, fully automated segmentation of the choroid is essential in making medically useful the possibilities given by the method of high resolution, high speed OCT.

7.4 Conclusions

In this chapter various approaches to segmentation have been presented, some for completeness even though they have been surpassed by the more efficient ones. Thus it is important to emphasize which are the current state of the art both from medical and computer science perspectives.

Detecting top and bottom boundaries using level sets and adaptive thresholding approach worked sufficiently well for many cases, however the method developed for choroidal top and bottom boundary detection outperforms both, while its potential use in a broad range of applications outside the OCT makes it a significant contribution from the computer science point of view as well. It is based on dynamic programming shortest path algorithm that incorporates path derivative and edge orientation criteria (based on directional wavelet analysis), implemented efficiently to preserve the same time complexity O(E) as the standard shortest path approach. The fundamental difference between this approach and others is that the algorithm works with limited path history, instead of only using static weights per edge to determine the distance function. Thus, there is no need for complex and slow workarounds, as when trying to incorporate path information with static weights only.

For the segmentation of inner retinal layers, simpler methods using pure texture classification and adaptive template matching have been proposed, which led subsequently to the development of a unified statistical model that utilizes both shape and texture. From the computer science perspective, two important general novelties with respect to the standard Active Appearance Model (AAM) were introduced: a second term in the optimization function that penalizes large deviations from the three boundaries found by the adaptive thresholding algorithm and the second algorithm stage that refines the model fit for each A-scan independently, giving increased accuracy. Thus, it can be concluded that the algorithm successfully demonstrated reliable performance under conditions which prove extremely challenging for the pre-existing methods. It was used as a basis for the development of the model based choroid segmentation algorithm, proving that indeed it is well suited to the weak and missing data conditions.

8 Pre-Apoptotic Monitoring

Here, a novel ligand-free method for assessment of neuronal viability *in vitro and ex vivo* is described. Using ultrahigh resolution optical coherence tomography (UHR-OCT) to detect changes in the light scattering properties of early stage apoptotic cells, it is shown that an increase in proportion of apoptotic cells *in vitro* can be accurately quantified. Moreover, the data indicates a similar increase in neuronal scatter in retinal explants following axotomy, suggesting that UHR-OCT can be a novel non-invasive technique for the *in vivo* assessment of neuronal health.

Texture inside the pre-apoptotic regions was analyzed by projecting the data into the multidimensional feature space and performing training and analysis in the feature space using Gaussian Mixture Models (GMMs). Analysis was performed on cell cultures grown on glass coverslip (in vitro) by training and distinguishing two groups: healthy and apoptotic. Following the extremely encouraging results (clear separation of the two groups with the distance between them far exceeding the confidence interval) of this proof of principle experiment, the second experiment focused on more demanding pre apoptotic analysis of the retinal explants.

8.1 OCT system

Imaging was performed using a spectrometer-based frequency-domain OCT using a commercially available Ti:Sapphire laser light source (Integral OCT, Femtolasers GmbH, Austria) with a bandwidth of 140 nm a schematic is shown Figure 48 A scanning microscope was used as the sample probe with the OCT scan lens (10x and 5x) (Thorlabs, Newton, New Jersey, USA) with 18 mm and 36 mm working distance, giving an axial resolution of ~4 µm and a transversal resolution of ~6 µm and ~8 µm respectively. The light source was connected to a 90/10 beam splitter (Ipitek Inc, Carlsbad, CA, USA) after which the bandwidth of the interferometer remained at 130 nm. The scanning microscope was connected to the 10% arm of the beam splitter and samples to be imaged were exposed to ~1.4 mW of power only during requisition. Images were recorded using a 12 bit 2048 pixel silicon CCD-camera (AVIVA M2 CL2014-BAO, Atmel, CA, USA) at a line rate of 20 kHz with 50 ms exposure time. Sample and reference arm fibres (SM 650, Fibrecore Ltd, UK) were looped through polarization control paddles to adjust for maximal fringe visibility. A gold coated hollow corner cube reflector mounted on a translation-stage was used in the reference arm and a reflective neutral-density filter-wheel allowed adjustment of the power such that the intensity on the camera was close below the saturation point.

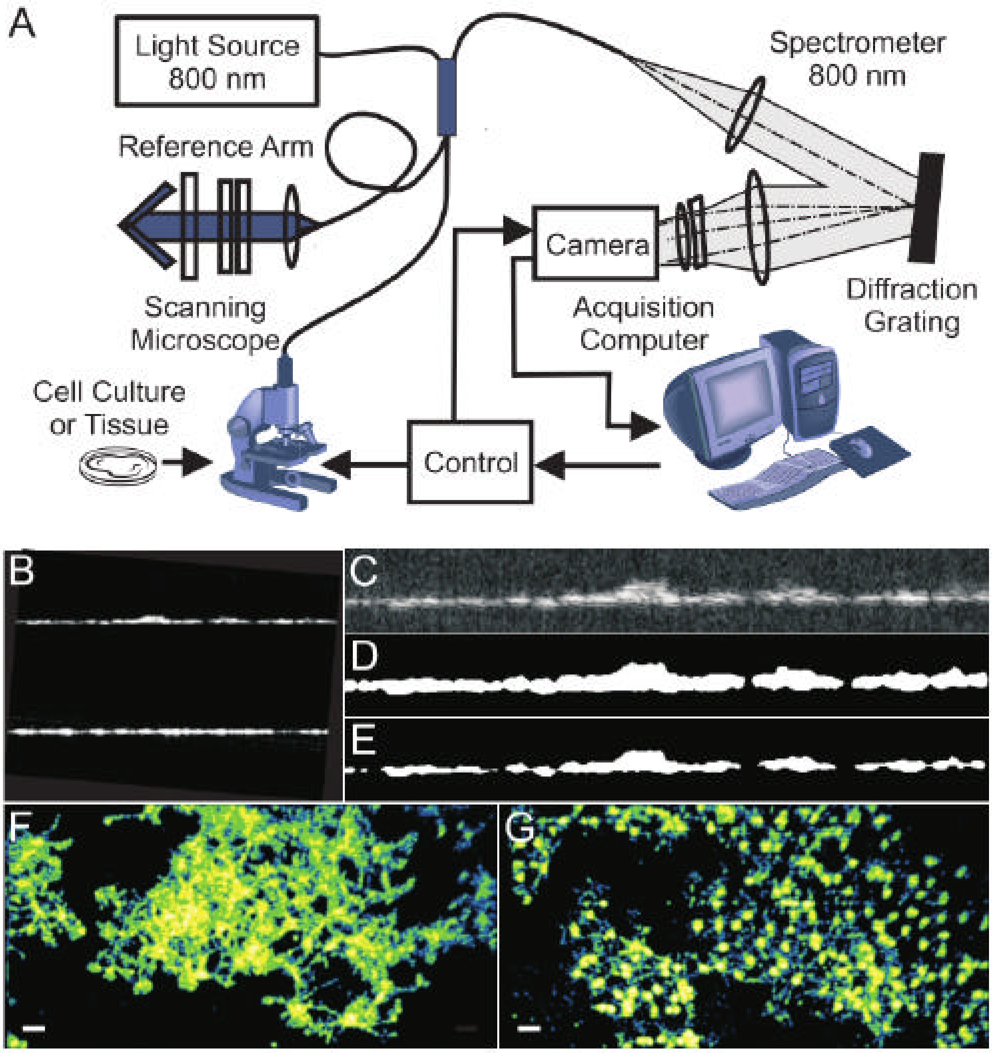


Figure 48 Cell imaging setup and pre processing. **In A the OCT system is shown. B and C (zoomed it) shows the input, the initial mask is shown in D, while the final mask is obtained after morphological operations, E. An example of control cells is shown in F, 24h after the treatment, they contract, G.**

Three different experimental sets of OCT images have been collected and analysed. The first experiment aimed to detect a difference between healthy (control) and apoptotic cell cultures 24 hours after treatment with staurosporine. The second experiment focused on identifying changes during the early stages of apoptosis *in vitro* where each cell culture was imaged at 6 different time points (10 to 60 mins) following treatment with staurosporine. Finally, in the third experiment, *ex vivo* explants where imaged during the first hour of the apoptotic process induced by axotomy.

8.1.1 Identifying the optical signature of apoptotic RGC-5 cells 24h after treatment with staurosporine.

For the first experiment, data were collected from 20 apoptotic and 20 control RGC-5 coverslip cultures whereby each coverslip was imaged at 10 different locations. The quality of all images was then visually inspected and some images were discarded based on technical imperfections. The remaining 120 control and 120 apoptotic 3D images (also referred to as images or image stacks) were selected for analyses. After initial image pre-processing, a fully automated approach was used to locate the cells within the image (the region of interest, (ROI)).

A set of parameters/features that described each ROI in compact form was then defined. In general, the choice of parameters is subjective. In this study, 65 interdependent texture features were defined based on 3D distribution of pixel intensity, using either individually or in combination averages, granulometry measures, and spatial wavelet analyses (as detailed in 8.2.3). A standard feature analysis technique was applied to distinguish between images of healthy versus apoptotic cells. To identify the generalised distribution of features for each type of data (data class) a GMM was used that also allowed for subgroups (clusters) within each data class. All 65 features were used for the analysis and up to four clusters were allowed for each class. As?the basic learning process the exact number of required clusters was then established during the fitting procedure based on Akaike Information Criteria (AIC), that estimates the trade off between accuracy and complexity of the model. Once all the parameters of the model were determined, during the validation stage, a new image-stack was introduced and analysed as a new data point in the multidimensional feature space. The Mahalanobis distance was computed between the new point and the respective centre of each of the earlier defined clusters. The new point was then classified, e.g. assigned to a class separated by the smallest distance.

The cross-validation ("leave-one-out") technique was used to evaluate the statistical accuracy of the classification. Data were divided into subsets, containing 10 to 20 image stacks according to the imaging session (to minimise variability due to the manual adjustment of parameters such as light polarisation during acquisition). During each step of the procedure, one subset was used for validation (each of the images in the subset were classified) while the rest of the data was used for finding the model parameters. The procedure was repeated for each of the subsets and the classification rate defined as a ratio of correct classifications over the total number of classifications.

8.1.2 Identifying the optical signature of early apoptosis in RGC-5 cells

The second experiment was set up to determine whether it is possible to identify optical differences in RGC-5 cells at early stages of apoptosis. For this analysis, data was collected from 34 separate RGC-5 cultures undergoing apoptosis, restricting the analyses to the first hour after application of staurosporine. Some of the images were discarded after initial visual inspection due to poor quality and the feature analysis was performed on 28 cultures, imaged every 10 min from 10 to 60 min post staurosporine treatment (168 image stacks in total, divided into 6 subsets according to the imaging session).

8.1.3 Identifying the optical signature of early apoptosis in explants

The final experiment was designed to identify the early signs of apoptosis in *ex vivo* retinal explants during the first hour post axotomy. Major changes were anticipated in the Inner Plexiform Layer (IPL), where dendritic terminals of RGCs are located, this was based on observations that RGC dendrites undergo a significant amount of remodelling and pruning prior to shrinkage and loss of the cell body (Weber, Kaufman and Hubbard, 1998).

Calcein/PI staining was carried out to identify the presence of live/dead cells within the murine retinal explants.

ROI was identified by a human expert, manually masking several regions within the retina, using a simple graphic interface. The borders of each ROI were then automatically adjusted (eroded) as stated in the methods section. 13 retinas were analysed and 10 ROI sized 128 x 64 x 64 pixels within the IPL of each retina with a depth of 64 slices were collected. This produced 130 image-stacks for each time point that were further divided into 5 subsets according to the timing of image acquisition.

Since feature analysis reliably discriminates healthy and apoptotic RGC5 cells, and that subtle early changes can be identified, the same procedure was then applied to the ROI within intact retinal explants undergoing apoptosis following axotomy (Quigley et al., 1995) .

8.2 OCT image analyses

OCT images were acquired as a stack of 512 slices sized 512 by 1024 pixels (1492 µm by 746µm). Prior to detailed analyses, all OCT images were subjected to standard image processing techniques. In particular, background noise was subtracted from the image; the position of the coverslip holding cell culture was detected by examining the direction of high frequency component in two dimensional frequency space, and all possible tilts were removed by rotating the image; the line scan artefacts were removed using FFT filtering. All image processing was performed in Matlab (Mathworks) using standard functions from image analysis and statistics toolbox.

8.2.1 Masking the Region Of Interest (ROI) for in vitro cell culture

For each 3D image, the intensity values of pixels were plotted as a histogram. The derivative of this distribution determined a heuristic threshold for the selection of pixel intensity values. This approach effectively removed the majority of pixels with a low signal to noise ratio attributed to background noise. All input pixels (Figure 48 B and C) above the threshold were classified as within ROI (Figure 48 D). This approach, however, did not guarantee continuity of ROI, because regions of low signal to noise intensity could also occur within ROI boundaries, producing “holes”. To resolve this issue, basic mathematical morphology techniques were used (first morphological dilation combined with mask erosion, then mask erosion alone; the sphere radius was chosen to be 5 pixels for texture filters and subsequently for morphological operators, that was the estimated bio-physical scale of interest, as it relates to the mitochondria size) to gradually adjust the ROI boundaries, eliminate “holes” and low intensity regions around the border, and produce a homogeneous

3D globule shaped ROI (Figure 48 E). The extension of erosion was set empirically to ensure that any background signals were removed. Finally, any brightness variations within the image (largely dependent on the depth of focus) were corrected using histogram equalization inside the ROI. After the background was removed, it was possible to plot a clear projection of control 0h (Figure 48 F) and apoptotic 24h cells (Figure 48 G).

8.2.2 Masking the Region of Interest for explants

In contrast to homogeneous cell cultures, such as RGC-5 cells, retinal explants are complex multilayer structures. Shadows produced by the adjacent layer and imaging artefacts make automatic ROI detection challenging. Thus, a semi-automatic approach was used with a human expert masking several ROI within the IPL, with a good SNR, using a custom designed graphical user interface (GUI). The borders of each ROI then were automatically adjusted (eroded) similarly as described above.

8.2.3 Feature selection

Each ROI was parameterized using 65 features. First, three standard image filters were applied to calculate the local entropy of the image, local range, and local standard deviation of pixel intensity. For each of these three matrices, four average parameters were calculated across the ROI: mean, median, entropy and standard deviation, hence producing the first 12 features. The feature values were further normalised according to the grey scale resolution of the original image. For example, for the 8 bit image, the entropy value was divided by 8 and standard deviation multiplied by 2.

Next, three co-occurrence matrices were defined over the image with an offset of 5 pixels along one of three (x, y, and z) coordinates respectively. The value of the offset was chosen similar to the size of a single cell. For each of the three matrices four average values were calculated across the ROI, such as contrast, correlation coefficient, energy and homogeneity. This operation provided us with the next 12 features.

Using the threshold that was previously established for the background signals while defining the ROI, the gray scale image was transformed into a binary black and white, and performed simple granulometry analyses taking into account only the size of continuous patches of image, referred to as particles. This histogram, specifying the number of particles of different sizes, provided us with further 12 features (particles up to size 12); the size of the histogram bin was increased (doubled), and another 6 features were extracted. By repeating the same procedure once again, 3 more features were extracted, totalling the number of granulometry features to 21.

To estimate the distribution of spatial frequencies in the image stack, directional wavelet analysis (Kingsbury, 1999) along three axes was performed: x, y, and 45 degree between x and y, concentrating on the three smallest scales of wavelets. For each scale and direction, the mean and standard deviation of coefficient values across the ROI were calculated, adding another 18 features to the feature space.

Finally, 3rd and 4th moments of pixel intensity distribution were calculated across the ROI adding skewness and kurtosis to the feature space.

The resulting feature space is thus comprised of 65 features with high levels of dependence between one another.

8.2.4 Feature Analyses

The core idea behind feature analysis is that objects within images can be easily recognized in a feature space with a smaller dimensionality than the original image, using a GMM to identify feature representation for each of the classes of data (apoptotic cells vs control). The GMM allows for the presence of subpopulations within a class, without requiring information about identity of these subpopulations. It allows up to four subpopulations in GMM (in our methodology) with the exact number being subject to the fitting procedure (training). Taking into account the normal distribution of the data, the best fit was determined in accordance with Akaike Information Criteria (AIC) using the Expectation Maximization (EM):

$$AIC = 2 * k - 2 * \ln(L) \tag{19}$$

where k is the number of subpopulations, and L is the likelihood function characterising the fit of the model with a given set of parameters to the training data.

After defining model parameters during the training stage, the validation stage consisted of presenting a “new” image to be classified using the model parameters. The image is defined as a set of features and its class identity is known but not revealed. To classify the image, the Mahalanobis distance is computed between a new point in the multidimensional feature space X (representing the new image) and the centre of a cluster Y, representing all the images in a given class:

$$d(\mathbf{x},\mathbf{y}) = \sqrt{(\mathbf{x}-\mathbf{y})^T S^{-1} (\mathbf{x}-\mathbf{y})} \qquad (20)$$

where S is the covariance matrix of the cluster.

The new image is assigned to a class separated by the minimal distance. The cross-validation technique (“leave-one-out”) was used to assess how accurately the predictive model differentiates between classes. On each step of cross-validation, all data excluding one of the data subsets were used for finding model parameters. The remaining subset was used for model validation, e.g. all images of this subset were classified (correctly or incorrectly) to one of the classes learned from the training set. Multiple rounds of cross-validation were performed, and the percentage of correctly classified images was estimated.

To visualise multidimensional data, Sammon projection along the x-axis with the largest variance was used. As Mahalanobis distance is only defined for two points within the same class/cluster, the distance between two points from two different clusters was calculated as follows:

$$d(\mathbf{x},\mathbf{y}) = \frac{2}{\frac{1}{d_1} + \frac{1}{d_2}} \qquad (21)$$

where d_1 is Mahalanobis distance computed using the covariance matrix of the first cluster and d_2 is Mahalanobis distance computed using the covariance matrix of the second cluster. Harmonic mean was used to mitigate the impact of large outliers. The Centre of Gravity (COG) plots were produced by finding the mean of each cluster; the confidence intervals were computed, assuming normal data distribution.

8.2.5 Feature Reduction

Feature reduction helps to further decrease dimensionality and speed up the learning process. The feature reduction was performed using sequential forward search and an additional cross-validation (“leave-one-out”) procedure over each training set. In practice, for each step of the analysis one data subset was chosen for validation (classification) and the rest of the data were used for finding model parameters (training) as before. However, now the training set was also subdivided, with one data subset left out for validation of feature selection and the rest of the data used to find model parameters as a function of smaller feature set. The procedure was repeated for each subset from the training set and the multiple sets of “optimal” features were produced. Due to the small number of subsets, “union” was performed on all the relevant features, producing a reduced feature set that was not necessary the optimal. It should be noted that classification was performed on the set that was different and completely independent from the training set and the feature selection, all the images were classified in this remaining data subset using reduced feature set. The whole procedure was repeated with different choices of validation data and classification rate was calculated as before.

Giving the binary nature of classification results, the classification error *p* can be described by the Bernoulli distribution. Bernoulli distribution belongs to a class of exponential distributions, thus the optimal number of features *k* can be estimated using Bayesian Information Criterion (BIC). This is an asymptotic result derived under the assumption that the data distribution is in the exponential family:

$$BIC = k * \ln(n) - 2 * \ln L \tag{22}$$

To use the measure in practice, it is necessary to estimate the likelihood *L.* While standard BIC applications usually assume normal distribution of variable errors/disturbances, in this case it is a Bernoulli distribution again, as the classification error *p* is either 0 or 1 but so are the error disturbances.

We derived the following measure:

$$BIC = n * \ln(p * (1 - p)) + k * \ln(n) \tag{23}$$

where *n* is the number of data sets. During the iterative procedure, all features could then be sorted by their classification contribution and the resulting set could be formed by adding classification features one by one, starting from the most relevant feature, while evaluating the BIC after each addition. The minimal value of the BIC corresponds to the most appropriate data set.

8.3 Results

8.3.1 Identifying the optical signature of apoptotic RGC-5 cells 24h after treatment with staurosporine.

In the first experiment, two distinct classes were identified representing healthy and apoptotic cells respectively. The classification rate was estimated at 95%. To visualise the data, low dimensional projection plots were produced using Sammon projection methods (Sammon, 1969) allowing the conversion of distances from multidimensional to two dimensional space. The resulting distance is dimensionless, and the scale is given in arbitrary units (Figure 49 a), b)). The results were correlated with the tunnel labelling (Figure 49 c)).

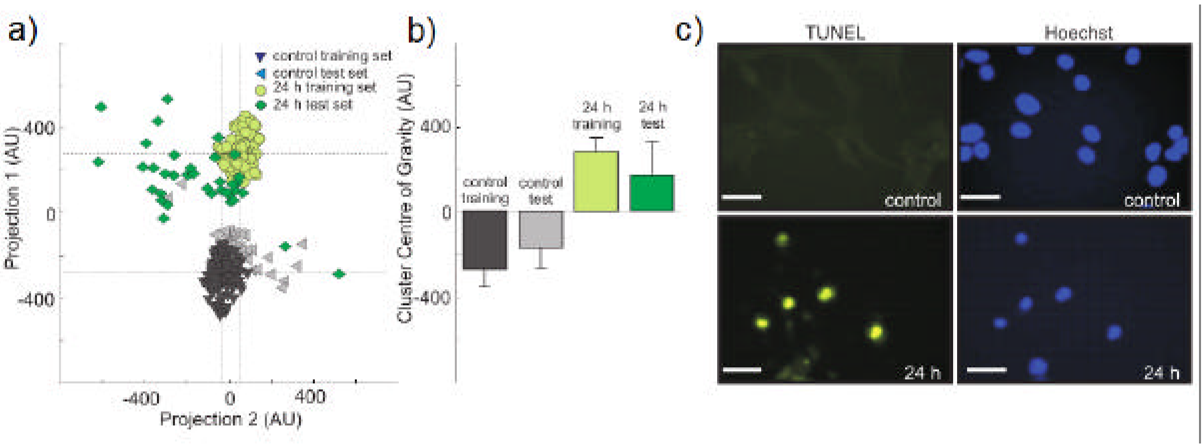


Figure 49 a) and b) show a clear separation of the control and 24h datasets, after the projection into a 2D feature space. Tunnel labelling in c) correlates with the analysis results.

8.3.2 Identifying the optical signature of early apoptosis in RGC-5 cells

The ROIs were automatically detected for all the images as described in the previous section. Due to the increased complexity of the classification task and the smaller data set, the number of independent features had to be reduced. By disregarding the most irrelevant and redundant features, dimensionality was reduced and the learning process was accelerated. Feature reduction was performed using a sequential forward search and an additional cross-validation (“leave-one-out”) procedure for each training set (see details in 8.2.5). Due to the relatively small number of subsets, “union” was performed on all the relevant features. The resulting set of features was then used for validation, which was completely independent in regard to both training and feature reduction. The “union” of the features, however, is not ideal, especially for a large data set. On average 17 best features for each of the validation procedures were used; however, a much smaller number of features could often suffice, although the specific set of highly discriminative features varies. The classifications during the validation stage were, hence, performed with a slightly different set of features for each data subset. As before, the classification was repeated multiple times in order to calculate the classification rate.

There is a progressive right shift in the centre of gravity (COG) of clusters representing cells in the feature space (Figure 50 a) and b)) from the control position towards the apoptotic position. The most dramatic shift occurred during the first 30 mins after the administration of staurosporine. The COG then remained in the same position between 30 and 40 min and shifted again between 40 and 50 min before reaching a plateau.

The classification rate in differentiating 10-20 min from 30-60 min data was 85%. These changes correlated with the MitoTracker Red mitochondrial labelling (Figure 50 c)), which showed a change in the mitochondria morphology post 20 mins of apoptosis induction.

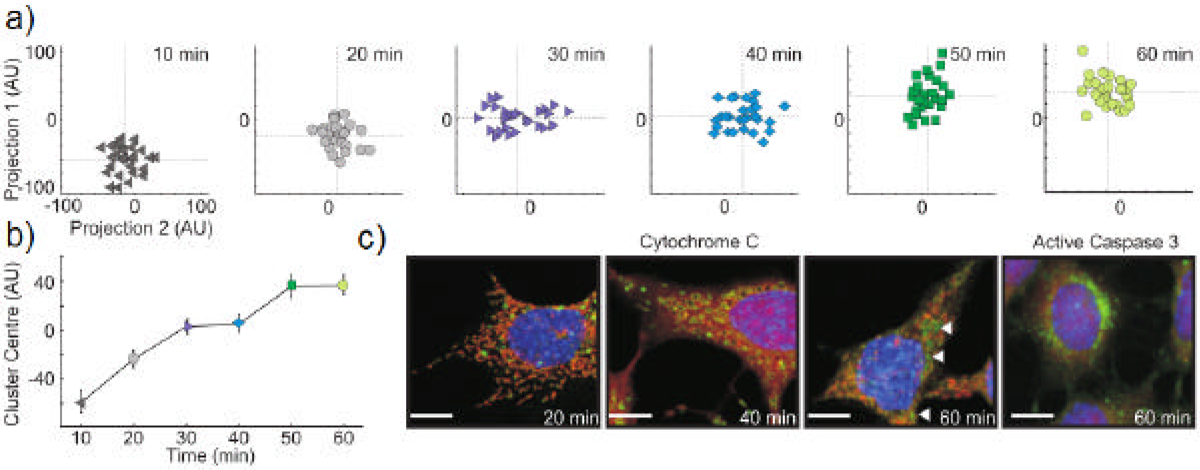


Figure 50 In a) and b), a progressive right shift in the centre of gravity (COG), of clusters representing cells in the feature space, can be observed, from the control position towards the apoptotic position. These changes correlated with the MitoTracker Red mitochondrial labelling, c), which showed a change in the mitochondria morphology post 20 mins of apoptosis induction.

8.3.3 Identifying the optical signature of early apoptosis in explants

Positive Calcein (green) staining with little or no negative PI (red) staining at 23 and 53 min post axotomy indicated that the retinal explants were alive with little or no dead cells within the RGC layer of the retina (Figure 51 a)). Haematoxylin and Eosin histology indicated that the retinal morphology during retinal explant culture did not change.

Similar to the analysis of RGC-5 cultures the centre of gravity clusters representing ROIs in *ex vivo* explants progressively shifted to the right over time (Figure 51 b)). The shift was, however, less pronounced than for the RGC-5 cultures, most likely due to the higher degree of variability of cells within the retinal tissues. Nevertheless, the difference between 15-23 min and 33-53 min post axotomy timepoints was detected in the areas of the IPL with the accuracy of 68%. Visual evaluation of the tissue confirmed that the retina remained transparent throughout the experiment, indicating that the changes observed were not secondary to opacification of the retinal tissue.

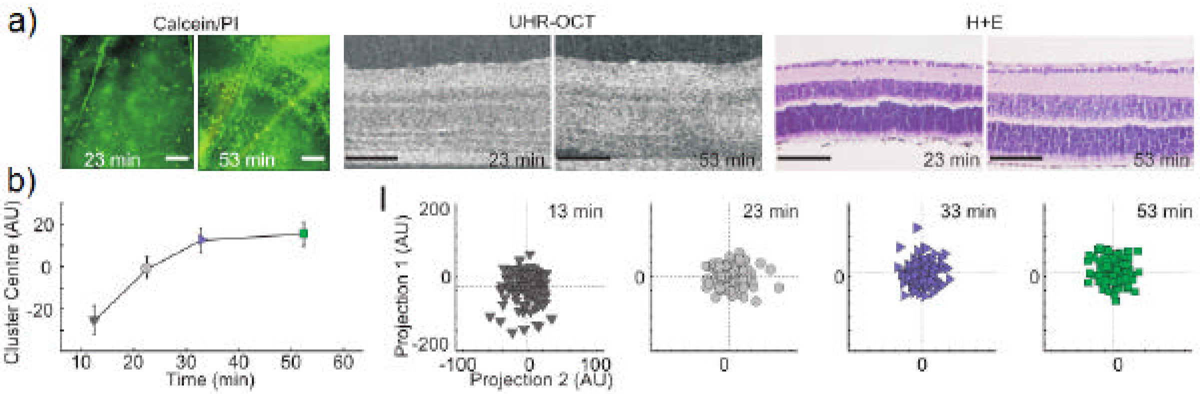


Figure 51 For the explant timeseries, a) shows Positive Calcein (green) staining with little or no negative PI (red) staining, at 23 and 53 min post axotomy, indicating that the retinal explants were alive with little or no dead cells within the RGC layer of the retina. Haematoxylin and Eosin histology indicated that the retinal morphology during retinal explant culture did not change. In b), a shift of COG can be observed, although not as prominent as in the case of the cell timeseries.

8.4 Conclusions

It was shown that texture analysis of retinal tissue imaged by UHR-OCT can identify optical changes associated with cellular events leading to apoptosis. Importantly, these changes were observed in both cell culture and explant systems. The results indicate that UHR-OCT can detect changes in the optical signatures of cells and explants very early after the insult to initiate cell death. In RGC-5, *in vitro* cultures changes were seen as early as 10 mins and in retinal *ex vivo* explants as early as 20 mins, suggesting that UHR-OCT has the potential for *in vivo* diagnostics.

The UHR-OCT derived changes closely match observations from immuno-histochemical analysis reporting morphological changes in the mitochondrial morphology. Whilst it could not be conclusively verified that the optical changes detected, solely derived from mitochondrial morphological changes, recent studies, using a Fourier processing microscope setup, also concluded that the mitochondria generate optical scatter changes during apoptosis (Pasternack, Zheng and Boustany, 2010). The value of this imaging technique in scoring neuronal viability/health is not contingent on knowing which organelles are generating the optic signature.

Binary discrimination of viable and dead cells was evaluated by histochemical staining. Whilst this demonstrated a high degree of accuracy, the conclusion was that the greater value of this imaging technique is in the derivation of quantitative indices of neuronal health. The shift in the centre of gravity of a cluster representing cells in feature space correlated well with the anticipated time course of degeneration within treated cell cultures and axotomised retinal explants. It is possible to overlay these indices on the UHR-OCT images to provide tissue related maps of cellular health.

It was shown that this methodology can be translated from *in vitro* to *ex vivo* retinal explants where early apoptotic changes were detected within the IPL of the retina. The UHR-OCT system is able to image the full depth of the retinal explant and all layers of the retina can be identified, allowing for a pan-retinal analysis of apoptosis. The high axial resolution of OCT is dependent on the spectral bandwidth of the light source and is, effectively, not aperture limited, meaning it does not depend on the objective used in the OCT setup.

This study has demonstrated that UHR-OCT can be used to provide an index of neuronal viability *in vitro* and *ex vivo* and has several advantages over previous apoptosis detection methods, namely it is a non-toxic, ligand-free, non-invasive technique, that can be used in live cell imaging. UHR-OCT is often used in clinical practice to image tissues, therefore it is highly probable that with the development of detection technology, this methodology could be applied to *in vivo* studies. This is a major step forward towards identifying timely therapeutic interventions to reverse or prevent cell loss.

9 Summary

Overall, during the course of the PhD numerous results were obtained that enabled more efficient application of OCT in the existing applications (retinal imaging) and applications to novel fields (pre-apoptotic monitoring). Work on facial dynamics and behaviour transfer was a separate Computer Science department project that produced interesting results on behavioural transfer and general signal processing. Although application wise the main PhD focus and the facial dynamic work did not have any common areas, the methods used were often linked (e.g. machine learning in general). Importantly, some of the developed methods have general use outside the OCT field as well.

The registration results were useful, however more work is possible to improve the elastic (non-affine) registration results and ideally use the independent fundus image in the process as a reference which is free of motion artefacts compared to a fundus image generated from 3D OCT data. Level set method was tried for retinal boundary detection but proved to be too unreliable and slow. The fast adaptive thresholding method was thus developed that is fast and effective. Segmentation results using texture-based segmentation are reasonably accurate when segmenting data that is close to the training data, but behave poorly on the more varied data. The reason for this is the use of the geometrical features in the process that enforces geometry of the training set on target data. If these features are left out, then segmentation results are patchy and geometrically inconsistent, with various layers detected at large offsets from proper locations. Adaptive template matching results were more accurate than the texture based results but are also susceptible to large failures since all the A-scans are segmented independently.

A statistical model based approach was developed that unifies the texture and geometric based approaches under statistical constraints posed and extracted from training data. It performs very well, it was tested on a large independent data set and the results were published in an international journal (Kajic et al., 2010). Two novel methods were added to the standard model approach: the shape prior information extracted by thresholding that increases the overall accuracy (7.2.4.3), and the second 1D refinement step based on the back projection of the training data through the 2D model (7.2.4.3). Both methods can be used in general problems outside the OCT field as well. A similar statistical model was constructed to segment the choroid. It incorporates maximum stable extremal region (MSER) blob detector as the key texture feature.

The top and bottom fovea boundary (ILM and RPE) novel method was developed according to high robustness requirements in order to work well in cases of severe pathologies and signal degradation. It is a shortest path dynamic programming method extended to use the path derivative information and edge orientation with the same time complexity. It is also applicable to a general class of computer vision problems.

Texture analysis was successfully applied to pre-apoptotic monitoring to in vitro and ex vivo (retinal explant) data, justifying future work on in vivo analysis that would allow early diagnosis of some major neurodegenerative conditions. Effective texture features were constructed (e.g. point of view rotationally invariant directional 3D wavelet coefficients) and optimal number of feature criteria derived from Bayes information criteria (BIC) proposed.

The work on facial dynamics accomplished two goals. Using facial curvature information, it was possible to register 3D faces using facial key points (eye pits, nose etc.). The work on visual sequences of various subjects uttering the same sentences led to important results in trustworthiness analysis and additionally produced important improvements to dynamic time warping (DTW) general signal processing algorithm (real dynamic derivative computation and graph based DTW).

The results obtained were presented at various international and local conferences:

- texture based segmentation: oral presentation at SPIE 2008, San Jose, CA, USA
- adaptive template fitting poster presentations at ISIE and ARVO 2008, both Fort Lauderdale, FL, USA
- model based segmentation oral presentation at ECBO 2009, Munich, Germany
- pre-apoptotic monitoring oral presentation at SPIE 2010 San Jose, CA, USA
- facial dynamics poster presentation at Face and Gesture (FG) 2011, Santa Barbara, CA, USA
- various UK based conferences (BMVC London, Aberystwyth, Cardiff etc.).

and were published in international journals and peer reviewed conference proceedings:

- registration work (Povazay et al., 2009)
- model based foveal segmentation (Kajic et al., 2010)
- cell imaging apoptosis related work (Hofer et al., 2011b) and (Hofer et al., 2011a). Also a submission is underway for Nature Methods
- choroid segmentation related (Esmaeelpour et al., 2010) and (Esmaeelpour et al., 2011). An Optics Express submission is underway and another possible computer science journal publication
- facial dynamics related work (Benedikt et al., 2008) and (Aubrey et al., 2011), also another IEEE Signal processing submission is underway

10 Future Work

Based on progress with retinal and choroidal segmentation encourages future work on powerful methods for segmentation of 3D OCT retinal images using statistical models that combine shape and texture information learnt from a large set of training examples. Despite the success of the algorithms developed and presented in this work, there is still an important, and technically extremely challenging, subset of OCT data; severe pathologies, successful segmentation of which would enable therapy monitoring even in the later stages of retinal disease.

Problems arise due to variations in geometry, as layers do not always form a single, continuous region in a 2D slice, but can be fragmented creating many small blobs in a two-dimensional cross-section (e.g. RPE detachment), while even within the same tissue there are structural inhomogeneities due to the different orientations of the tissue. When segmenting layers, the presence of extraneous structures (e.g. blobs of fluid, shadows, vessels) interferes with the segmentation process as well.

These cases are the most challenging aspects of retinal image analysis, well beyond the current state of the art, and new approaches need to be developed. The methods available for segmenting non layer like structures (e.g. subretinal fluid, vessels) use simple thresholding techniques, and are not able to simultaneously perform layer segmentation (Ahlers et al., 2009).

A good solution to the given problems would continue in the spirit of machine learning and use a combination of a statistical model with iterative graph cuts. For the case of binary labelling (e.g. foreground/background), graph cuts provide an efficient (polynomial time) means for finding the optimal image segmentation (with respect to the given cost function) (Boykov, Veksler and Zabih, 2001). However, the choice of cost functions is limited (Kolmogorov and Zabih, 2004), and only approximate solutions are obtained for more than two labels (Boykov et al., 2001), but nevertheless graph cuts are an extremely powerful segmentation method, and have become very popular in the field of computer vision in the last few years.

As mentioned before, graph cuts have already been applied to segment OCT retina images (Garvin et al., 2009), but this work is limited in various respects. First, their images had very good SNR. Second, although the tissue imaged contained pathologies (anterior ischemic optic neuropathy), the structure was identical to normals except for a slight thickening of some layers. Third, they require manually specified heuristic constraints on the layers (smoothness, distance between layers, etc.). Thus, their graph cut approach would not directly extend to more varied pathologies.

Given the inherent difficulty of automatic segmentation it is worthwhile including all available prior information in the process. This has led to the recent trend to introduce spatial priors (the expected appearance of the modelled shape) within the graph cuts framework. One idea (Das et al., 2009) is to constrain the segmented foreground region to be compact, but they required user interaction and also used a simple (limited) definition of compactness. Another similar approach (Gulshan et al., 2010) used a star shape prior, again based on the user specifying one or several key points in the image. The above methods are not suitable for retinal image segmentation since they require a single, generic global shape descriptor, while we are interested in segmenting patchy areas with a certain distribution. Alternatively, it is possible to use a fixed shape template (Freedman and Zhang, 2005), but this is also too limited. The most relevant work for the proposed approach (Zhu-Jacquot and Zabih, 2007) combines a statistical shape and intensity model with graph cuts.

They use four energy terms:

- a single GMM to model the intensity probability distribution of the background
- the standard graph cut spatial homogeneity term
- a fitness between the statistical shape prior and the current segmentation
- the entropy of intensity distributions outside and inside the shape prior, which measures how well the shape prior fits the image and prevents an initial inaccurate segmentation from producing inaccurate shape priors

They iteratively update the GMM and shape prior parameters, followed by performing a graph cut, and repeat these steps until convergence. Another related work (Malcolm, Rathi and Tannenbaum, 2007) on combining a statistical shape prior with graph cuts was used in order to segment a connected shape (non-patchy).

The proposed approach differs from (Zhu-Jacquot and Zabih, 2007) since, given the layer fragmentation, it would have to drop the shape prior term and instead use, for each layer, multiple GMMs to model probability distributions of features which would be learnt from the training set. Fitting the model would involve updating the GMM parameters based on the graph cut segmentation at the previous iteration and would also be constrained by learning from the training set. This would allow a fragmented segmentation which would be nevertheless spatially constrained.

Another issue is that the existing manual labelling program used for the Mechanical Turk interface works with exact boundaries marking different layers. Since the new pathological data would have fragmented layers, a new method would need to be developed using region growing algorithms to ease access for well trained and expert human segmenters to identify the different layers, as the previous interface for users to identify layer boundaries would be unsuitable (i.e. too time consuming).. Therefore a new interface should be designed, based on a semi-automatic approach. The user approximately paints over the areas of interest, while the subsequent automatic stage performs the accurate segmentation based on brightness and some simple texture parameters. The user is then able to correct/refine these results by iterating the process.

As before, evaluation can be done both numerically against the manually segmented layers and clinically by cross-correlating disease staging with other classic methods of evaluation (i.e. fundus diagnostics).

It is important to mention that the developed method would be suitable for a general image segmentation problem and thus have a far wider area of application than just for the OCT field itself.

11 References

Ahlers C, Golbaz I, Einwallner E, Dunavölgyi R, Malamos P, Stock G, Pruente C et al. (2009) Identification of Optical Density Ratios in Subretinal Fluid as a Clinically Relevant Biomarker in Exudative Macular Disease. *Investigative Ophthalmology & Visual Science* 50: 3417-3424.

Ahlers C, Götzinger E, Pircher M, Golbaz I, Prager F, Schütze C, Baumann B et al. (2010) Imaging of the Retinal Pigment Epithelium in Age-Related Macular Degeneration Using Polarization-Sensitive Optical Coherence Tomography. *Investigative Ophthalmology & Visual Science* 51: 2149-2157.

Aubrey A J, Kajic V, Cingovska I, Rosin P L, and Marshall D (2011) Mapping and Manipulating Facial Dyanamics. In: *Int. Conf. on Automatic Face and Gesture Recognition*. Santa Barbara, CA, US.

Balu M, Liu G J, Chen Z P, Tromberg B J, and Potma E O (2010) Fiber delivered probe for efficient CARS imaging of tissues. *Optics Express* 18: 2380-2388.

Banaszek K, Radunsky A S, and Walmsley I A [eds.] (2004) *Blind dispersion compensation for optical coherence tomography*. Lasers and Electro-Optics, 2004. (CLEO). Conference on.

Beichel R, Bischof H, Leberl F, and Sonka M (2005) Robust active appearance models and their application to medical image analysis. *Medical Imaging, IEEE Transactions on* 24: 1151-1169.

Benedikt L, Kajic V, Marshall D, and Rosin P L (2008) Facial Dynamics in Biometric Identification. In: *British Machine Vision Conference*. London, UK.

Blankenberg F G, Katsikis P D, Tait J F, Davis R E, Naumovski L, Ohtsuki K, Kopiwoda S et al. (1998) In vivo detection and imaging of phosphatidylserine expression during programmed cell death. *Proceedings of the National Academy of Sciences of the United States of America* 95: 6349-6354.

Boppart S A, Bouma B E, Pitris C, Tearney G J, Fujimoto J G, and Brezinski M E (1997) Forward-imaging instruments for optical coherence tomography. *Opt. Lett.* 22: 1618.

Boykov Y, Veksler O, and Zabih R (2001) Fast Approximate Energy Minimization via Graph Cuts. *pami* 23: 1222-1239.

Broyden C G (1970) The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations. *IMA Journal of Applied Mathematics* 6: 76-90.

Cabrera Fernández D, Salinas H M, and Puliafito C A (2005) Automated detection of retinal layer structures on optical coherence tomography images. *Opt. Express* 13: 10200.

Cao L, and Fei-Fei L [eds.] (2007) *Spatially Coherent Latent Topic Model for Concurrent Segmentation and Classification of Objects and Scenes*. IEEE 11th International Conference on Computer Vision.

Cauchy A (1847) *Méthode générale pour la résolution des systèmes d'équations simultanées*.

Cereghetti G M, and Scorrano L (2006) The many shapes of mitochondrial death. *Oncogene* 25: 4717-4724.

Chen C H, Pau L F, and Wang P S P (1993) *Handbook of pattern recognition and computer vision*. World Scientific Publishing Co., Inc.

Chinn S R, Swanson E A, and Fujimoto J G (1997) Optical coherence tomography using a frequency-tunable optical source. *Opt. Lett.* 22: 340.

Chiu S J, Li X T, Nicholas P, Toth C A, Izatt J A, and Farsiu S (2010) Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation. *Opt. Express* 18: 19413-19428.

Cootes T F, and C.J.Taylor (2004) *Statistical Models of Appearance for Computer Vision*.

Cootes T F, Twining C J, Babalola K O, and Taylor C J (2008) Diffeomorphic statistical shape models. *Image Vision Comput.* 26: 326-332.

Cordeiro M F, Guo L, Luong V, Harding G, Wang W, Jones H E, Moss S E et al. (2004) Real-time imaging of single nerve cell apoptosis in retinal neurodegeneration. *Proceedings of the National Academy of Sciences of the United States of America* 101: 13352-13356.

Cortes C, and Vapnik V (1995) Support-Vector Networks. *Machine Learning* 20: 273.

Cristinacce D, and Cootes T F (2007) Boosted Regression Active Shape Models. In: *British Machine Vision Conference*.

Das P, Veksler O, Zavadsky V, and Boykov Y (2009) Semiautomatic segmentation with compact shape prior. *Image Vision Comput.* 27: 206-219.

De Souza-Ramalho P, Pereira Da Silva A, Pego P, and Bicho M P (2009) Diabetic retinopathy - Apoptosis and BRB breakdown prevention in early diabetic retinopathy. *Acta Ophthalmologica* 87: 0-0.

Debuc D C, Salinas H M, Ranganathan S, Tátrai E, Gao W, Shen M, Wang J et al. (2010) Improving image segmentation performance and quantitative analysis via a computer-aided grading methodology for optical coherence tomography retinal image analysis. *Journal of Biomedical Optics* 15.

Deng B, Jawerth B, Peters G, and Sweldens W (1993) Wavelet probing for compression based segmentation. In: Laine A F [ed.] *Mathematical Imaging: Wavelet Applications in Signal and Image Processing*. pp. 266--276.

Drexler W, and Fujimoto J G (2008) *Optical Coherence Tomography: Technology and Applications*. Springer.

Dunaief J L, Dentchev T, Ying G-S, and Milam A H (2002) The Role of Apoptosis in Age-Related Macular Degeneration. *Arch Ophthalmol* 120: 1435-1442.

Efros A A, and Freeman W T [eds.] (2001) *Image quilting for texture synthesis and transfer*. Proceedings of the 28th annual conference on Computer graphics and interactive techniques. ACM.

Esmaeelpour M, Povazay B, Hermann B, Hofer B, Kajic V, Hale S, North R V et al. (2011) Mapping choroidal and retinal thickness variation in Type 2 diabetes using 3D-1060nm-OCT. *Invest Ophthalmol Vis Sci*.

Esmaeelpour M, Povazay B, Hermann B, Hofer B, Kajic V, Kapoor K, Sheen N J et al. (2010) Three-dimensional 1060-nm OCT: choroidal thickness maps in normal subjects and improved posterior segment visualization in cataract patients. *Invest Ophthalmol Vis Sci* 51: 5260-5266.

Fabritius T, Makita S, Hong Y, Myllyl R, and Yasuno Y (2009a) Automated retinal shadow compensation of optical coherence tomography images. *Journal of Biomedical Optics* 14: 010503.

Fabritius T, Makita S, Miura M, Myllyla R, and Yasuno Y (2009b) Automated segmentation of the macula by optical coherence tomography. *Optics Express* 17: 15659-15669.

Fan-Yin T (2005) An Intelligent System Approach to Higher-Dimensional Classification of Volume Data. *IEEE Transactions on Visualization and Computer Graphics* 11: 273-284.

Fercher A, Hitzenberger C, Sticker M, Zawadzki R, Karamata B, and Lasser T (2001) Numerical dispersion compensation for Partial Coherence Interferometry and Optical Coherence Tomography. *Opt. Express* 9: 610.

Fercher A F, Mengedoht K, and Werner W (1988) Eye-length measurement by interferometry with partially coherent light. *Opt. Lett.* 13: 186.

Frank S, Gaume B, Bergmann-Leitner E S, Leitner W W, Robert E G, Catez F, Smith C L et al. (2001) The role of dynamin-related protein 1, a mediator of mitochondrial fission, in apoptosis. *Developmental Cell* 1: 515-525.

Freedman D, and Zhang T (2005) Interactive Graph Cut Based Segmentation with Shape Priors. *CVPR*: 755-762.

Garvin M M K, Abramoff M M D, Kardon R R, Russell S S R, Wu X X, and Sonka M M (2009) Intraretinal Layer Segmentation of Macular Optical Coherence Tomography Images Using Optimal 3-D Graph Search. *IEEE Transactions on Medical Imaging* 27: 1495-1505.

Gilboa G, Sochen N, and Zeevi Y Y (2004) Image Enhancement and Denoising by Complex Diffusion Processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26: 1020-1036.

Glocker B, Komodakis N, Tziritas G, Navab N, and N.Paragios (2008) Dense Image Registration through MRFs and Efficient Linear Programming. *Medical Image Analysis*.

Gossage K W, Smith C M, Kanter E M, Hariri L P, Stone A L, Rodriguez J J, Williams S K et al. (2006) Texture analysis of speckle in optical coherence tomography images of tissue phantoms. *Physics in Medicine and Biology* 51: 1563.

Gourley P L, Hendricks J K, McDonald A E, Copeland R G, Barrett K E, Gourley C R, Singh K K et al. (2005) Mitochondrial correlation microscopy and nanolaser spectroscopy - New tools for biophotonic detection of cancer in single cells. *Technology in Cancer Research & Treatment* 4: 585-592.

Green D R, and Reed J C (1998) Mitochondria and apoptosis. *Science* 281: 1309-1312.

Grulkowski I, Gora M, Szkulmowski M, Gorczynska I, Szlag D, Marcos S, Kowalczyk A et al. (2009) Anterior segment imaging with Spectral OCT system using a high-speed CMOS camera. *Opt. Express* 17: 4842-4858.

Gulshan V, Rother C, Criminisi A, Blake A, and Zisserman A (2010) Geodesic Star Convexity for Interactive Image Segmentation. *CVPR*.

Guyon I, and Elisseeff A (2003) An introduction to variable and feature selection. *J. Mach. Learn. Res.* 3: 1157-1182.

Hazaveh K (2003) Image Processing Techniques for Noise Removal, Enhancement and Segmentation of Cartilage OCT Images – A Review and Critique. In: *2nd International Conference for Upcoming Engineers*.

Hazaveh K, and Raahemifar K (2003) Local discriminant basis algorithm-a review of theory and applications in signal processing. *Circuits and Systems, 2003. ISCAS '03. Proceedings of the 2003 International Symposium* 4: 341-344.

Hee M R, Puliafito C A, Duker J S, Reichel E, Coker J G, Wilkins J R, Schuman J S et al. (1998) Topography of diabetic macular edema with optical coherence tomography. *PubMed* 105: 360-370.

Herz P, Chen Y, Aguirre A, Fujimoto J, Mashimo H, Schmitt J, Koski A et al. (2004) Ultrahigh resolution optical biopsy with endoscopic optical coherence tomography. *Opt. Express* 12: 3532.

Hofer B, Povaz ay B, Hermann B, Rey S M, Kajic V, Tumlinson A, Powell K et al. (2011a) Artefact reduction for cell migration visualization using spectral domain optical coherence tomography. *J Biophotonics* 4: 355-367.

Hofer B, Považay B, Unterhuber A, Wang L, Hermann B, Rey S, Matz G et al. (2011b) Fast dispersion encoded full range OCT for retinal imaging at 800 nm and 1060 nm. 18: 4898-4919.

Huang D, Wang J, Lin C P, Puliafito C A, and Fujimoto J G (1991) Micron-resolution ranging of cornea and interior chamber by optical reflectometry. *Lasers Surg. Med.*

Huber R, Wojtkowski M, Taira K, Fujimoto J, and Hsu K (2005) Amplified, frequency swept lasers for frequency domain reflectometry and OCT imaging: design and scaling principles. *Opt. Express* 13: 3513.

Ibanez L, Schroeder W, Ng L, and Cates J (2005) *The ITK Software Guide, Second Edition*.

Jan-Mark G, and Arnold W M S (2005) A Six-Stimulus Theory for Stochastic Texture. *Int. J. Comput. Vision* 62: 7-16.

Jiao S, Knighton R, Huang X, Gregori G, and Puliafito C (2005) Simultaneous acquisition of sectional and fundus ophthalmic images with spectral-domain optical coherence tomography. *Opt. Express* 13: 444.

Jiao S, Wu C, Knighton R W, Gregori G, and Puliafito C A (2006) Registration of high-density cross sectional images to the fundus image in spectral-domain ophthalmic optical coherence tomography. *Opt. Express* 14: 3368.

Joachim W (1999) Coherence-Enhancing Diffusion Filtering. *Int. J. Comput. Vision* 31: 111-127.

Kajic V, Povazay B, Hermann B, Hofer B, Marshall D, Rosin P L, and Drexler W (2010) Robust segmentation of intraretinal layers in the normal human fovea using a novel statistical model based on texture and shape analysis. *Opt Express* 18: 14730-14744.

Karbowski M, and Youle R J (2003) Dynamics of mitochondrial morphology in healthy cells and during apoptosis. *Cell Death Differ* 10: 870-880.

Karmarkar N (1984) A new polynomial-time algorithm for linear programming. In: *Proceedings of the sixteenth annual ACM symposium on Theory of computing*. ACM.

Kingsbury N (1999) Image processing with complex wavelets. *Philosophical Transactions of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences* 357: 2543-2560.

Kirk W G, Cynthia M S, Elizabeth M K, Lida P H, Alice L S, Jeffrey J R, Stuart K W et al. (2006) Texture analysis of speckle in optical coherence tomography images of tissue phantoms. *Physics in Medicine and Biology*: 1563.

Kohli P, and Torr P (2007) Dynamic Graph Cuts for Efficient Inference in Markov Random Fields. *IEEE Trans. Pattern Anal. Mach. Intell.* 29: 2079-2088.

Kolmogorov V, and Zabih R (2004) What Energy Functions Can Be Minimized via Graph Cuts? *pami* 26: 147-159.

Komodakis N, Tziritas G, and Paragios N (2008) Performance vs Computational Efficiency for Optimizing Single and Dynamic MRFs: Setting the State of the Art with Primal Dual Strategies. *Computer Vision and Image Understanding*.

Koozekanani D, Boyer K, and Roberts C (2001) Retinal thickness measurements from optical coherence tomography using a Markov boundary model. *Medical Imaging, IEEE Transactions on* 20: 900.

Kulkarni M D, Thomas C W, and Izatt J A (1997) Image enhancement in optical coherence tomography using deconvolution. *IEE Electronics Letters* 33: 1365-1367.

Leitgeb R, Hitzenberger C, and Fercher A (2003) Performance of fourier domain vs. time domain optical coherence tomography. *Opt. Express* 11: 889.

Lemaur G, Drouiche K, and DeConinck J (2003) Highly regular wavelets for the detection of clustered microcalcifications in mammograms. *Medical Imaging, IEEE Transactions on* 22: 393.

Levenberg K (1944) A method for the solution of certain non-linear problems in least squares. *Quarterly Journal of Applied Mathmatics* 2: 164-168.

Lexer F, Hitzenberger C K, Fercher A F, and Kulhavy M (1997) Wavelength-tuning interferometry of intraocular distances. *Appl. Opt.* 36: 6548.

Li L, Socher R, and Fei L (2009) Towards total scene understanding: Classification, annotation and segmentation in an automatic framework. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on* 0: 2036-2043.

Malcolm J, Rathi Y, and Tannenbaum A (2007) Graph Cut Segmentation with Nonlinear Shape Priors. *icip*: 365-368.

Marks D L, Oldenburg A L, Reynolds J J, and Boppart S A (2003) Digital Algorithm for Dispersion Correction in Optical Coherence Tomography for Homogeneous and Stratified Media. *Appl. Opt.* 42: 204.

Matas J, Chum O, Urban M, and Pajdla T (2004) Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing* 22: 761-767.

Mattson M P (2000) Apoptosis in neurodegenerative disorders. *Nat Rev Mol Cell Biol* 1: 120-129.

Mayer M A, Hornegger J, Mardin C Y, and Tornow R P (2010) Retinal Nerve Fiber Layer Segmentation on FD-OCT Scans of Normal Subjects and Glaucoma Patients. *Biomed. Opt. Express* 1: 1358-1383.

McGowan D (2006) Pruning processes. *Nat Rev Neurosci* 7: 685-685.

Mishra A, Wong A, Clausi D A, and Fieguth P W (2010a) Quasi-random nonlinear scale space. *Pattern Recognition Letters* In Press, Corrected Proof.

Mishra A K, Fieguth P W, and Clausi D A (2010b) Decoupled Active Contour (DAC) for Boundary Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33: 310-324.

Molnár J, Chetverikov D, Cabrera DeBuc D, Gao W, and Somfai G (2011) Layer extraction in rodent retinal images acquired by optical coherence tomography. *Machine Vision and Applications*: 1-11.

Mourant J R, Freyer J P, Hielscher A H, Eick A A, Shen D, and Johnson T M (1998) Mechanisms of light scattering from biological cells relevant to noninvasive optical-tissue diagnostics. *Applied Optics* 37: 3586-3593.

Mujat M, Chan R, Cense B, Park B, Joo C, Akkin T, Chen T et al. (2005) Retinal nerve fiber layer thickness map determined from optical coherence tomography images. *Opt. Express* 13: 9480.

Nocedal J, and Wright S J (1999) *Numerical Optimization*. Springer.

Ojala T, and Pietikäinen M (1997) Unsupervised texture segmentation using feature distributions. In: Del B [ed.] *Image Analysis and Processing*. Vol. 1310. Springer Berlin / Heidelberg, pp. 311-318.

Palade G E (1953) An electrone micrpscope study of the mitochondrial structure. *Journal of Histochemistry & Cytochemistry* 1: 188-211.

Pasternack R M, Zheng J Y, and Boustany N N (2010) Optical scatter changes at the onset of apoptosis are spatially associated with mitochondria. *Journal of Biomedical Optics* 15.

Perkon I, Košir A, Tasic J, and Diamond M (2010) Whisker detection as a shortest path problem. In: *Visual Observation and Analysis of Animal and Insect Behavior*.

Perona P, and Malik J (1990) Scale-Space and Edge Detection Using Anisotropic Diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.* 12: 629-639.

Podoleanu A, Charalambous I, Plesea L, Dogariu A, and Rosen R (2004) Correction of distortions in optical coherence tomography imaging of the eye. *Physics in Medicine and Biology* 49: 1277.

Povazay B, Bizheva K, Unterhuber A, Hermann B, Sattmann H, Fercher A F, Drexler W et al. (2002) Submicrometer axial resolution optical coherence tomography. *Opt. Lett.* 27: 1800.

Povazay B, Hermann B, Hofer B, Kajic V, Simpson E, Bridgford T, and Drexler W (2009) Wide-field optical coherence tomography of the choroid in vivo. *Invest Ophthalmol Vis Sci* 50: 1856-1863.

Prilloff S, Fan J Y, Henrich-Noack P, and Sabel B A (2010) In vivo confocal neuroimaging (ICON): non-invasive, functional imaging of the mammalian CNS with cellular resolution. *European Journal of Neuroscience* 31: 521-528.

Pshenichny B N, and Danilin Y M (1983) *Numerical Methods in External Problems*. Mir Publishers, Moscow.

Puliafito C A, Hee M R, Lin C P, Reichel E, Schuman J S, Duker J S, Izatt J A et al. (1995) Imaging of macular diseases with optical coherence tomography. *Ophthalmology* 102: 217.

Quigley H A, Nickells R W, Kerrigan L A, Pease M E, Thibault D J, and Zack D J (1995) Retinal ganglion-cell death in experimental glaucoma and after axotomy occurs by apoptosis. *Investigative Ophthalmology & Visual Science* 36: 774-786.

Reichstein D, Ren L Z, Filippopoulos T, Mittag T, and Danias J (2007) Apoptotic retinal ganglion cell death in the DBA/2 mouse model of glaucoma. *Experimental Eye Research* 84: 13-21.

Robertson J D, Orrenius S, and Zhivotovsky B (2000) Nuclear events in apoptosis. *Journal of Structural Biology* 129: 346-358.

Rossant F, Ghorbel I, Bloch I, Paques M, and Tick S (2009) Automated segmentation of retinal layers in OCT imaging and derived ophthalmic measures. In: *Proceedings of the Sixth IEEE international conference on Symposium on Biomedical Imaging: From Nano to Macro*. Boston, Massachusetts, USA, 1700225: IEEE Press, pp. 1370-1373.

Rottey S, Loose D, Vakaet L, Lahorte C, Vermeersch H, Van Belle S, and Van De Wiele C V (2007) Tc-99m-HYNIC Annexin-V imaging of tumors and its relationship to response to radiotherapy and/or chemotherapy. *Quarterly Journal of Nuclear Medicine and Molecular Imaging* 51: 182-188.

Rottey S, Slegers G, Van Belle S, Goethals I, and Van de Wiele C (2006) Sequential Tc-99m-hydrazinonicotinamide-annexin V imaging for predicting response to chemotherapy. *Journal of Nuclear Medicine* 47: 1813-1818.

Sammon J W (1969) A nonlinear mapping for data structure analysis. *Ieee Transactions on Computers* C 18: 401-&.

Sarunic M V, Yazdanpanah A, Gibson E, Xu J, Bai Y, Lee S, Saragovi H U et al. (2010) Longitudinal study of retinal degeneration in a rat using spectral domain optical coherence tomography. *Opt. Express* 18: 23435-23441.

Schmitt J M (1998) Restoration of Optical Coherence Images of Living Tissue Using the CLEAN Algorithm. *Journal of Biomedical Optics* 3: 66.

Schmitt J M, Xiang S H, and Yung K M (1999) Speckle in Optical Coherence Tomography. *Journal of Biomedical Optics* 4: 95.

Schmitz-Valckenberg S, Guo L, Maass A, Cheung W, Vugler A, Moss S E, Munro P M G et al. (2008) Real-time in vivo imaging of retinal cell apoptosis after laser exposure. *Investigative Ophthalmology & Visual Science* 49: 2773-2780.

Scholz M (2005) Non-linear PCA: a missing data approach. *Bioinformatics* 21: 3887-3895.

Scholz M, Fraunholz M, and Selbig J (2007) Nonlinear Principal Component Analysis: Neural Network Models and Applications.*Principal Manifolds for Data Visualization and Dimension Reduction*. pp. 44-67.

Selesnick I W, Baraniuk R G, and Kingsbury N G (2005) The Dual-Tree Complex Wavelet Transform. *IEEE Signal Processing Magazine* 22: 123-151.

Sorzano C O S, Thevenaz P, and Unser M (2005) Elastic registration of biological images using vector-spline regularization. *Biomedical Engineering, IEEE Transactions* 52: 652.

Spall J C [ed.] (1997) *Accelerated second-order stochastic optimization using only function measurements*. Decision and Control, 1997., Proceedings of the 36th IEEE Conference on.

Swanson E A, Izatt J A, Hee M R, Huang D, Lin C P, Schuman J S, Puliafito C A et al. (1993) In vivo retinal imaging by optical coherence tomography. *Opt. Lett.* 18: 1864.

T.F.Cootes, Edwards G J, and C.J.Taylor (2001) Active Appearance Models. *IEEE PAMI* 23: 681-685.

Terreno E, Castelli D D, and Aime S (2010) Encoding the frequency dependence in MRI contrast media: the emerging class of CEST agents. *Contrast Media & Molecular Imaging* 5: 78-98.

Thevenaz P, and Unser M [eds.] (1996) *A pyramid approach to sub-pixel image fusion based on mutual information*. Image Processing, 1996. Proceedings., International Conference on Image Processing.

Thornberry N A (1997) The caspase family of cysteine proteases. *British Medical Bulletin* 53: 478-490.

Thornberry N A (1998) Caspases: key mediators of apoptosis. *Chemistry & Biology* 5: R97-R103.

Timo O, Matti P, inen, Topi M, and enp (2002) Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* 24: 971-987.

Tolliver D, Koutis Y, Ishikawa H, Schuman J S, and Miller G L [eds.] (2008) *Unassisted Segmentation of Multiple Retinal Layers via Spectral Rounding*. ARVO.

van der Meer F J, Faber D J, Aalders M C G, Poot A A, Vermes I, and van Leeuwen T G (2010) Apoptosis- and necrosis-induced changes in light attenuation measured by optical coherence tomography. *Lasers in Medical Science* 25: 259-267.

Vedaldi A, and Fulkerson B (2008) *VLFeat: An Open and Portable Library of Computer Vision Algorithms* [Online]. Available at: http://www.vlfeat.org/ [Accessed:

Vercauteren T, Perchant A, Malandain G, Pennec X, and Ayache N (2006) Robust mosaicing with correction of motion distortions and tissue deformations for in vivo fibered microscopy. *Medical Image Analysis* 10: 673.

Vermeer K A, van der Schoot J, Lemij H G, and de Boer J F (2011) Automated segmentation by pixel classification of retinal layers in ophthalmic OCT images. *Biomed. Opt. Express* 2: 1743-1756.

Vermes I, Haanen C, and Reutelingsperger C (2000) Flow cytometry of apoptotic cell death. *Journal of Immunological Methods* 243: 167-190.

Vermes I, Haanen C, Steffensnakken H, and Reutelingsperger C (1995) A novel assay for apoptosis - flow cytometric detection of phosphatidylserine expression on early apoptotic cells using fluorescein-labeled annexin-v. *Journal of Immunological Methods* 184: 39-51.

Viola P, and Jones M (2001) Rapid object detection using a boosted cascade of simple features. *CVPR*.

Wang I J, and Spall J C [eds.] (1998) *A constrained simultaneous perturbation stochastic approximation algorithm based on penalty functions*. Intelligent Control (ISIC), 1998. Held jointly with IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA), Intelligent Systems and Semiotics (ISAS), Proceedings.

Weber A J, Kaufman P L, and Hubbard W C (1998) Morphology of single ganglion cells in the glaucomatous primate retina. *Investigative Ophthalmology & Visual Science* 39: 2304-2320.

Weissman J, Hancewicz T, and Kaplan P (2004) Optical coherence tomography of skin for measurement of epidermal thickness by shapelet-based image analysis. *Opt. Express* 12: 5760.

Westphal V, Rollins A, Radhakrishnan S, and Izatt J (2002) Correction of geometric and refractive image distortions in optical coherence tomography applying Fermat?s principle. *Opt. Express* 10: 397.

Wong A, Mishra A, Bizheva K, and Clausi D A (2010) General Bayesian estimation for speckle noise reduction in optical coherence tomography retinal imagery. *Optics Express* 18: 8338-8352.

Wong E, Kumar V, Howman-Giles R B, and Vanderheyden J L (2008) Imaging of Therapy-Induced Apoptosis Using Tc-99m-HYNIC-Annexin V in Thymoma Tumor-Bearing Mice. *Cancer Biotherapy and Radiopharmaceuticals* 23: 715-725.

Wright A J, Poland S P, Girkin J M, Freudiger C W, Evans C L, and Xie X S (2007) Adaptive optics for enhanced signal in CARS microscopy. *Optics Express* 15: 18209-18219.

Yang Q, Reisman C A, Wang Z, Fukuma Y, Hangai M, Yoshimura N, Tomidokoro A et al. (2010) Automated layer segmentation of macular OCT images using dual-scale gradient information. *Opt. Express* 18: 21293-21307.

Yazdanpanah A, Hamarneh G, Smith B, and Sarunic M (2009) Intra-retinal Layer Segmentation in Optical Coherence Tomography Using an Active Contour Approach. In: *Proceedings of the 12th International Conference on Medical Image Computing and Computer-Assisted Intervention: Part II*. London, UK, 1691261: Springer-Verlag, pp. 649-656.

Yoo B, and Pagel M D (2008) An overview of responsive MRI contrast agents for molecular imaging. *Frontiers in Bioscience* 13: 1733-1752.

Yu K, Ji L, Wang L, and Xue P (2001) How to optimize OCT image. *Opt. Express* 9: 24.

Yun S H, Tearney G, de Boer J, and Bouma B (2004) Motion artifacts in optical coherence tomography with frequency-domain ranging. *Opt. Express* 12: 2977.

Yung K M, Lee S L, and Schmitt J M (1999) Phase-Domain Processing of Optical Coherence Tomography Images. *Journal of Biomedical Optics* 4: 125.

Zawadzki R J, Choi S S, Jones S M, Oliver S S, and Werner J S (2007) Adaptive optics-optical coherence tomography: optimizing visualization of microscopic retinal structures in three dimensions. *J. Opt. Soc. Am. A* 24: 1373.

Zhu-Jacquot J, and Zabih R (2007) Graph Cuts Segmentation with Statistical Shape Priors for Medical Images. *IEEE Conf. on Signal-Image Technologies and Internet-Based System*: 631-635.

Zvyagin A V, Blazkiewicz P, and Vintrou J (2005) Image reconstruction in full-field Fourier-domain optical coherence tomography. *Journal of Optics A: Pure and Applied Optics* 7: 350.

12 List of Figures

13 List of Abbreviations

AAM	Active Appearance Model
ACO	Ant Colony Optimization
AMD	Age related Macular Degeneration
AMT	Amazon Mechanical Turk
ASM	Active Shape Model
BFGS	Broyden-Fletcher-Goldfarb-Shanno
BIC	Bayesian Infromation Criteria
CARS	Coherent Anti-Stokes Raman Spectroscopy
CCD	Charge-Coupled Device
ChT	Axial Choroidal Thickness
CL	Connecting Cilia
CMOS	Complementary Metal–Oxide–Semiconductor
COG	Centre Of Gravity
CSI	Choroidal-Sclera Interface
CSLO	Confocal Laser Scanning Ophtalmoscopy
DE	Differential Evolution
DL	Direct Logarithm
DTCW	Dual Tree Complex Wavelet
DTW	Dynamic Time Warping
ELM	External Limiting Membrane

FFT	Fast Fourier Transform
GA	Genetic Algorithm
GCL/IPL	Ganglion Cell Layer and Inner Plexiform Layer
GMM	Gaussian Mixture Model
GNA	Gauss-Newton Algorithm
HIT	Human Intelligence Task
ICON	In Vivo Confocal Neuroimaging
IE	Information Expansion
INL	Inner Nuclear Layer
IS	Inner Segment
LBP	Local Binary Patterns
LDB	Local Discriminant Basis
LMA	Levenberg-Marquardt Algorithm
MCMC	Marcov Chain Monte Carlo
MD	Minimum Distortion
MRI	Magnetic Resonance Imaging
MSER	Maximum Stable Extremal Region
NaN	Not a Number
NFL	Nerve Fibre Layer
NN	Neural Network
OCT	Optical Coherence Tomography

ONH	Optic Nerve Head
ONL	Outer Nuclear Layer
OPL	Outer Plexiform Layer
OS	Outer Segment
PCA	Principal Component Analysis
PELM	Post-ELM
PSF	Point-Spread Function
QA	Quantum Annealing
RAAM	Robust AAM
RBC	RPE/Bruch's membrane/choriocapillaris complex
ROI	Region Of Interest
RPE	Retinal Pigment Epithelium
SA	Simulated Annealing
SLO	Laser Scanning Ophtalmoscope
SNR	Signal-to-Noise Ratio
SPSA	Simultaneous Perturbation Stochastic Approximation
SQP	Sequential Quadratic Programming
SVM	Support Vector Machine
TL	Truncation Logarithm
TM	Thickness Map
TMD	Truncated MD

UHR	Ultra High Resolution
ZAP	Zero Adjustment Procedure

Mapping and Manipulating Facial Dynamics

Andrew J. Aubrey, Vedran Kajić, Ivana Cingovska, Paul L. Rosin and David Marshall

Abstract— **This paper describes a novel approach to building models of temporal dynamics for facial animation with applications in performing perceptual testing of trustworthiness. A vital component of the system is a method to bring two image sequences into temporal alignment. Our approach is to project the two sequences into face space (built using shape models [1]) and apply dynamic time warping (DTW). However, the variability in the sequences causes the standard DTW algorithm to perform poorly on our data, and so we have overcome this by extending DTW in the following ways: 1) the signal magnitudes are augmented by incorporating derivatives [2], and a scheme for estimating weights in the cost function is proposed, 2) the set of sequences is used to build a graph, with nodes representing sequences and edges indicating the cost of applying the extended DTW to align pairs of sequences; better alignments between sequences can now be found by traversing the minimum cost path through the graph.**

Once all signals are aligned to a common temporal reference it is straightforward to map the temporal dynamics from one face to another. A remapped face is synthesised using the new trajectory in face space to drive an active appearance model [1]. Furthermore, the common temporal reference allows us to build a statistical model of the dynamics. This can be used to both identify dynamics of interest and also to manipulate the dynamics, e.g. to reduce or exaggerate facial dynamics.

I. INTRODUCTION

Facial animation is becoming an increasingly important area in computer graphics as various media (movies, television, video games) look to produce more and better computer generated animation. Much of the early work focused on generating a face that looked as realistic as possible, but recent emphasis is directed at making these faces move in a realistic manner.

Due to recent advances in computer graphics it is becoming increasingly common for psychologists to use computer generated stimuli to conduct perceptual experiments since they can be more easily controlled and manipulated than direct video footage.

The goal of this work is to enable psychologists to analyse the trustworthiness of a speaker as determined by the perception of their facial dynamics. For this a system is required that allows the dynamics of a speaker to be modified. In our work we require a system that allows automatic manipulation of a speaker's dynamics. We also require a system that both retains the shape and appearance characteristics of a subject, and also modifies the speech of a subject so that after modification the audio and video remain synchronised.

The necessity for life-like animation requires the use of performance driven animation [3]. Current techniques do not easily allow manipulation of the timing of facial dynamics, nor can they be easily used to synchronise the audio and video data.

Previous work on performance driven animation has included work by Vlasic et al [4], Zhang et al [5] and Chang and Ezzat [6]. Statistical models such as the Active Appearance Model (AAM) are widely used due to their ability to encode a high level of information using only several parameters. Cosker et al [7] describe how AAMs can be used to transfer expressions between AAMs and morph-target based models by identifying key regions of the face and manipulating these regions independently of one another. Theobald et al [8] and De la Hunty et al [9] both describe how AAMs can be used for real time expression transfer. In [8] the authors drive the AAM mean of one model with the modes of variation from another model. While this produces very realistic animations, it does not allow any control over the dynamics. The method in [9] is an extension of an early version of the method in [8] and suffers the same drawback.

This paper addresses those issues using a framework based on dynamic time warping (DTW) and shape and appearance models. Firstly, we explain our novel approach for extending existing DTW methods. We then, using shape models to capture the face dynamics, show how the trajectories of shape models from a reference video and target video are aligned using DTW. The aligned shape trajectories are used to resynthesise a new video sequence with the desired dynamics. Our method differs from previous work as it allows us to manipulate the temporal dynamics in a more sophisticated manner. As well as simply mapping temporal dynamics, we can exaggerate them, compute mean dynamics, identify typical and atypical dynamics etc. The alignment found using DTW can also be used to warp the audio to ensure synchrony with the video. It should also be pointed out that none of the methods cited above allow simultaneous warping of the speech as our method does.

II. DYNAMIC TIME WARPING

Dynamic time warping (DTW) is a method to find the optimal non-linear alignment between two time series signals. It achieves this by warping (compressing or expanding) the time base of two signals. The classic DTW algorithm is as

This work was supported by One Wales Research Institute of Visual Computing (RIVIC)

A. Aubrey, P. Rosin and D. Marshall are with the Visual Computing Group, School of Computer Science and Informatics, Cardiff University, Wales, UK. {`A.J.Aubrey, Paul.Rosin, Dave.Marshall`} `@cs.cardiff.ac.uk`

V. Kajić is with the School of Optometry and Vision Sciences, Cardiff University, Wales, UK. `KajicV1@cardiff.ac.uk`

I. Cingovska took part in this work while at the School of Computer Science and Informatics, Cardiff University as part of the International Association for the Exchange of Students for Technical Experience (IAESTE) program.

follows [10]. Given two sequences s_1 and s_2 of length M and N respectively, a warping path W can be found that defines a mapping (alignment) between s_1 and s_2. To align the two sequences an M by N distance matrix must first be constructed where the (i^{th}, j^{th}) element of the matrix contains the distance between the i^{th} point in s_1 and the j^{th} point in s_2. Typically the Euclidean distance is used.

The warping path W is a contiguous set of matrix elements defined as:

$$W = w_1, w_2, \ldots, w_k, \ldots, w_K \quad \max(M,N) \leq K \leq n+m-1 \tag{1}$$

where K is the length of the warp path and $w_k = (i,j)_k$. The aim of DTW is to find the path W through the distance matrix with the least cost. To this end the following constraints are employed:

- Monotonicity: the indices i and j are monotonically spaced in time.
- Continuity: if $w_k = (i,j)_k$ then $w_{k+1} = (i',j')_{k+1}$ where $i'_{k+1} - i_k \leq 1$ and $j'_{k+1} - j_k \leq 1$. This limits (in time) the distance between the aligned points.
- Boundary: The warping path must start and finish at diagonally opposite corners of the distance matrix, $w_1 = (1,1)$ and $w_K = (M,N)$.

There are many warping paths that satisfy the above constraints, but the path of interest has the minimum warping cost:

$$C_{DTW} = \sum_{k=1}^{K} D(i_k, j_k) \tag{2}$$

Using dynamic programming [10], the path with the minimum cost can be found by building an accumulated cost matrix C which is defined as follows:

$$C(i,j) = D(i,j) + \min \left\{ \begin{array}{c} C(i,j-1) \\ C(i-1,j) \\ C(i-1,j-1) \end{array} \right\} \tag{3}$$

The above equation controls the step size of DTW, and modification of (3) produces a different alignment.

While the classic DTW algorithm performs well for signals that are similar except for local accelerations and decelerations in the time axis, Keogh and Pazzani [11] describe how it is prone to failure when the two sequences differ in amplitude. Other typical errors are when a small portion of one signal maps to a large portion of the other signal or when a rising edge is incorrectly mapped to a falling edge because they are close in time.

A. DTW variation

To improve the accuracy of the warping path there have been many proposed variations to the classic algorithm; several are discussed by Muller [10]. Typically, one or more of the following variations are used.

Globally constraining the domain of the warp path can speed up the DTW computation. Also, it will limit the degree of certain failures but may not prevent them. The Sakoe-Chiba band [12] and the Itakura parallelogram [13] are the two most widely used global constraints. A weighting factor can be introduced to (3) to bias the direction of the warp path. This can prevent the many to one mapping situation, but a bad choice of weights can force an incorrect path to be taken. The step pattern (3) can also be modified and is one of the more popular methods of varying the DTW algorithm due to the flexibility it allows.

The above DTW variations are not always effective, and it may be necessary to incorporate additional information. To this end Keogh and Pazzani [11] proposed the Derivative DTW (DDTW) algorithm. They calculate the first derivative of the sequences which provides shape information of the signals. Their experiments show that the DTW and DDTW have similar performance with signal distortion confined to the time axis. However, DTW performs poorly when there is a distortion in the amplitude of the signals, whereas the DDTW algorithm performs significantly better.

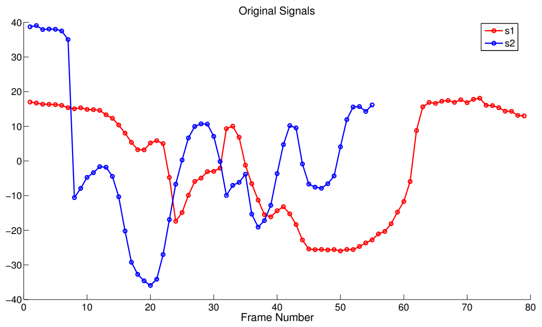
According to Benedikt et al [2], DDTW is highly sensitive to noise, so they expand on the work in [11] by proposing a Weighted Derivative DTW (WDTW) algorithm. The WDTW is a weighted combination of classic DTW and DDTW, and they also use the local second derivative of the signals. The use of the second derivative allows information on acceleration to be included in the calculation of the alignment. Their experiments show the WDTW produces a better alignment than relying on just the signal magnitudes or derivatives alone. However, the values of the weights were carefully chosen manually as the magnitudes of the derivatives and the original signal are not in the same range.

Figure 1 shows the results of applying the classic DTW algorithm and the WDTW algorithm to a pair of signals. The original signals are shown in Fig 1(a), and they are taken from our data set to highlight why existing methods may not give the desired results. Fig 1(b) is the result of applying the classic DTW algorithm to the signals, and Fig 1(c) is the result of applying the WDTW algorithm. It is clear that both methods have to some degree aligned the two signals, however, this has been achieved at the cost of creating periods where there is no amplitude variation. This repeating of a value for several samples may not be an issue for some applications. However, as we want to resynthesise video with realistic dynamics this a problem as the face will appear frozen for several frames which appears unrealistic.

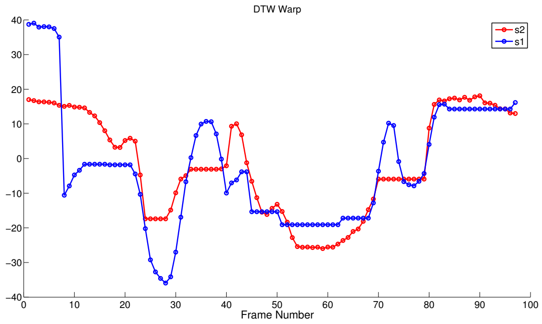
In our work we use the WDTW algorithm as the basis of our DTW framework but contribute the following novel extensions. The first is a simple method for automatically calculating the weight values. Secondly, we describe a graph based DTW algorithm that, given a set of signals to be aligned, can find the best mapping between two signals by using intermediate steps of first mapping to other signals in the set and then to the target signal.

B. Automatic Weight Selection

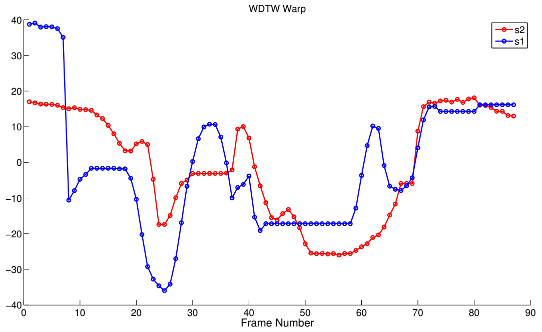
Weighted derivative DTW (WDTW) is simply a weighted combination of the DTW and DDTW methods. It is also possible to include higher derivatives, but as in [2] we limit ourselves to using first and second derivatives only. Mathematically this is expressed as:



(a) Original signals



(b) Signals warped using DTW



(c) Signals warped using WDTW

Fig. 1. Plots of, (a) two signals from our dataset, (b) the signals warped using the classic DTW algorithm and (c) the signals warped using the WDTW from [2].

$$C_W = w_0 \times C_0 + w_1 \times C_1 + w_2 \times C_2 + \ldots + w_q \times C_q \tag{4}$$

where C_0 is the distance matrix found using the classic DTW method, C_q is the distance matrix found using the q^{th} derivative and w_q is the weight associated with the q^{th} derivative.

Each pair of signals to be warped requires the weight values for each of the distance matrices in (4), the wrong values of these weights can lead to a bad alignment by including too much or too little of the derivatives. Without loss of generality, the weight w_0 is chosen to be 1.

To find values for w_1 and w_2 the following is used. Given two signals to warp s_1 and s_2 of length n and m respectively, treat each signal independently, in this case work with s_1 first:

1) First calculate s'_1 and s''_1, the first and second order derivatives respectively.
2) Randomly permute s_1, s'_1 and s''_1 to obtain s_{1P}, s'_{1P} and s''_{1P}.
3) Find w_1 and w_2 using the following ratios:

$$w_1 = \frac{\sum (s_{1P} - s_1)^2}{\sum (s'_{1P} - s'_1)^2} \tag{5}$$

and

$$w_2 = \frac{\sum (s_{1P} - s_1)^2}{\sum (s''_{1P} - s''_1)^2} \tag{6}$$

This results in a set of weights for sequence s_1, $\{w_{s_1 0}, w_{s_1 1}, w_{s_1 2}\}$. The same procedure can be applied to sequence s_2 to obtain the set of weights $\{w_{s_2 0}, w_{s_2 1}, w_{s_2 2}\}$. Using either of these sets of weights risks biasing the WDTW algorithm in favour of one of the signals. To counter this we find the mean set of weights:

$$w_0 = w_{s_1 0} = w_{s_2 0} = 1 \tag{7}$$

$$w_1 = E(w_{s_1 1}, w_{s_2 1}) \tag{8}$$

$$w_2 = E(w_{s_1 2}, w_{s_2 2}) \tag{9}$$

where E is the expectation value.

The above is easily extended for multi-dimensional signals, by summing the numerator and denominator of (5) and (6) over each dimension. For example, for (5) we would have:

$$w_1 = \frac{\sum \sum_{r=1}^{R} (s_{1rP} - s_{1r})^2}{\sum \sum_{r=1}^{R} (s'_{1rP} - s'_{1r})^2} \tag{10}$$

where R is the number of dimensions. A similar expression is obtained for (6).

III. GRAPH BASED DTW (GWDTW)

The signals we wish to warp are not guaranteed to have a high correlation in shape and will have different amplitudes, thus it is quite possible for the previously mentioned DTW methods to fail to align the signals correctly. As mentioned in [11] the amplitude difference will have a significant impact on the "correctness" of the warp. If however there exists a set of sequences $S = \{s_1, s_2, \ldots, s_Y\}$, it may be possible to correctly warp s_1 to s_x by first warping via s_y where $1 \leq x, y \leq Y$.

If the signals are treated as nodes and the warping between two signals as edges the set of signals and their respective warping can be treated as a graph (see Figure 2). This may be viewed as a complete graph as all nodes are connected to each other. It is therefore possible to align any two signals in the set. Each pair of aligned signals will have an associated error (the mismatch between the resulting aligned signals) and this is used as a weight value for the graph edges. Figure 2 is an example graph containing four signals represented as nodes and the associated edge weights given by the signal mismatch error values. The total error of the warping path W is used as the edge weights for pairs of signals. i.e $e = w_K$.

To find the best warping path between two signals it is simply a case of finding the shortest path through the graph. As the error values are non negative, Dijkstra's shortest path algorithm [14] can be applied. The path with the smallest total error e is chosen as the warping path. Typically in our small dataset we find path lengths of two or three.

IV. FACE APPEARANCE MODEL

To transfer the facial dynamics from one face sequence to another we first capture the shape and appearance of each face using the standard active appearance model (AAM) developed by Cootes et al [1]. To obtain the shape information, salient feature points on the face must be chosen and tracked through each frame in the video sequence. This is a highly time consuming process if done manually, therefore, an automatic method is sought. It is possible to use make up or markers placed on the subject's face prior to capture to track facial features, however, we wish to use the captured textures for re-synthesis, so it is preferable to avoid post processing the images to remove unwanted make up and markers. There are methods that can automatically track landmarks, but while they are adequate for rigid motion, they are still not fully reliable for highly non-rigid movement, particularly in tracking the lips as the lips are highly deformable.

A. Landmarking a video sequence

To automatically landmark a sequence of images it is typically required to manually landmark an image in the sequence (usually the first frame) and then automatically track these landmarks through the image sequence. However, this can prove to be unreliable for the lips, especially the outer and inner edge of the lower lip due to the nature and amount of deformation they undergo whilst speaking or performing facial expressions. To overcome this we use a semi-automatic framework based around groupwise registration of the frames in each sequence. To landmark a video sequence the following steps are taken. First the groupwise registration algorithm of Sidorov et al [15] is applied which performs non-rigid alignment of every image in a sequence. The result of this is a mean image, where the sharper the mean image is the better the registration and a set of deformation maps $\{\mathscr{D}_n; n = 1, \ldots, N\}$, where N is the length of the sequence. The deformation map describes the pixel mapping from a single image to the mean. Therefore, we only need to landmark a single frame (the mean image) and use the set of deformation maps to reverse map the landmarks onto the original images. Figure 3 contains example mean images from groupwise registration of one sequence in our dataset. The left image is the mean image from the first groupwise iteration. The blurring around the edge of the face and the mouth region highlights the amount of respective rigid and non-rigid motion in the sequence. The right image is the final groupwise output. The image is visibly sharper, hence, landmarks placed in this image will be mapped to the correct positions in the original set of images with little to no manual correction.

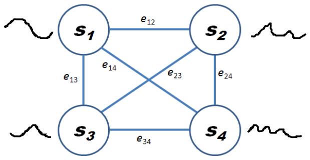


Fig. 2. Example graph, where nodes $s_1, \ldots, s_4$ are the signals and the edges e_{ij} are the error values obtained from warping signals i and j.

Fig. 3. Mean images from the first (left) and last (right) iteration of groupwise registration from one sequence from the dataset

While we have found this technique significantly reduces the manual placement of landmarks, we find that some frames occasionally need to be manually corrected. This is because groupwise registration requires the deformation in the image set to be diffeomorphic. When working with faces, blinking or mouth opening will violate this rule. The sequences we use have the subjects speaking, and through all sequences we find that only several frames require manual correction (where the sequence length varies between 60 and 100 frames), so there is still a significant saving in manual landmarking.

B. Active Appearance Models

Active appearance models (AAMs) [1] are a joint statistical model of shape and colour values (texture), where a single appearance parameter defines a corresponding texture and shape vector. The shape model is obtained from the Cartesian coordinates of the landmarks. For a single image, $\Phi(j)$ is the vector of landmark coordinates, and the collection of vectors $\{\Phi(j)\}, 1 \leq j \leq N$ describe the shape variation over the set of images N. As we wish to re-synthesise colour video, the textures from within these shapes are described by the texture model Γ. By applying principal component analysis (PCA) to the shape and texture data separately, the statistical shape and texture models are obtained. The shape model can be expressed as:

$$\Phi = \overline{\Phi} + \mathbf{P}_{\Phi}\mathbf{b}_{\Phi} \tag{11}$$

and the texture model is expressed in a similar format:

$$\Gamma = \overline{\Gamma} + \mathbf{P}_{\Gamma}\mathbf{b}_{\Gamma} \tag{12}$$

where $\overline{\Phi}$ and $\overline{\Gamma}$ are the mean shape and texture vectors, $\mathbf{P}_{\Phi}$ and $\mathbf{P}_{\Gamma}$ are matrices formed from eigenvectors, and $\mathbf{b}_{\Phi}$ and $\mathbf{b}_{\Gamma}$ are the corresponding eigenvalues, typically referred to as shape and texture parameters respectively. Varying the shape and texture parameters $\mathbf{b}_\Phi$ and $\mathbf{b}_\Gamma$ allows any of the original images to be approximated, or indeed new images to be synthesised. The advantage of using shape and appearance models is that the shape and texture parameters allow the high dimensional space within which the images lie to be represented by only a few parameters.

To construct a full AAM we combine $\mathbf{P}_\Phi$ and $\mathbf{P}_\Gamma$ and apply PCA once more to obtain a combined shape and appearance parameter. The number of shape and texture parameters retained can be chosen by either retaining the several highest shape and texture values, or by retaining a percentage of their total energy.

To align the dynamics of a reference and target sequence we first align the trajectories of the shape parameters of the reference to the target. A new sequence is resynthesised by using the shape parameters of the aligned reference sequence to correctly manipulate the appearance model of the reference sequence.

V. DATA COLLECTION AND EXPERIMENTS

The data used in the following experiments was recorded using an interlaced digital video camera at 25fps and the subjects were filmed under the same lighting conditions. Each subject was asked to recite the sentence “Once upon a time”, and to also elongate one word in the sentence. We recorded seven sequences and denote these as $s_1, s_2, \ldots, s_7$.

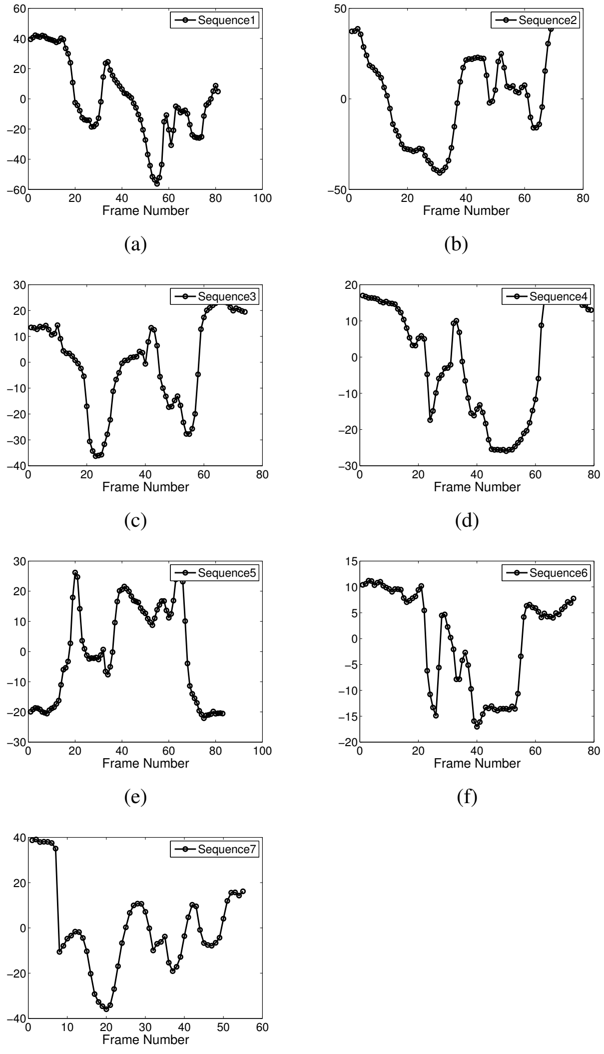
To build the shape and appearance models we used the semi-automatic procedure described in Section IV to annotate each frame with 49 landmarks. For each subject we retained 95% of the variance in the shape model and for the appearance model we retained 99% of the total variance.

The purpose of this work is to transfer the dynamics of one subject’s facial actions onto another subject who is performing a similar set of facial actions. The idea being that one subject inherits the timing of the dynamics of the other subject. Each of the shape parameters will have a trajectory through the image sequence, so for actors performing similar actions there should be similarities between the shape trajectories. This is shown in Figure 4, which depicts the trajectories of the principal shape component from the actors reciting the same phrase.

It should be noted that even though the shape parameters are multi-dimensional, it is trivial to adapt the classic DTW, its variants and our GWDTW technique to be used with such signals.

A. Experiments

One issue with using shape models is ensuring the principal components (modes of variation) of different subjects are correlated. By this we mean that the first t modes correspond to the same features in all subjects. It would be useless to have the first 2 modes of subject one corresponding to vertical and horizontal motion of lips and the first 2 modes of subject two corresponding to eye motion. To ensure correspondence we manually inspect the first t modes of each subject to ensure they correspond.



(a) (b) (c) (d) (e) (f) (g)

Fig. 4. Plots of the first mode of variation of the shape model from each of the subjects in the dataset. The plots (a) to (g) are the sequences 1 to 7 respectively.

In the following experiments we do not show the results of using the classic DTW algorithm to align the sequences, but only consider the WDTW and our proposed GWDTW method. To test the accuracy of the method we start by first comparing the result of warping sequence seven (s_7) to sequence three (s_3) with the weights $w_0, w_1, w_2 = 1$. Figures 5 and 6 show the results of aligning the signals using the WDTW and GWDTW methods respectively. In Figure 5 s_3 is represented by the blue curve and s_7 by the red. It can be seen that until around frame 30, the WDTW method appears to have aligned the signals correctly, but from thereon the trajectories are not aligned.

Using our GWDTW method we find that there is not a satisfactory direct warp from s_7 to s_3. Instead s_7 is first warped to s_4 and then warped to s_3. Figure 6 depicts the result of warping s_7 to s_4 with the blue curve and the result of warping from here to s_3 with the red curve, which is in

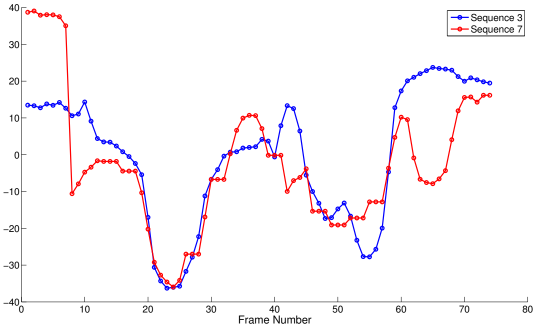


Fig. 5. Sequence s_7 warped to s_3 using WDTW with the weights fixed at $w_0 = w_1 = w_2 = 1$.

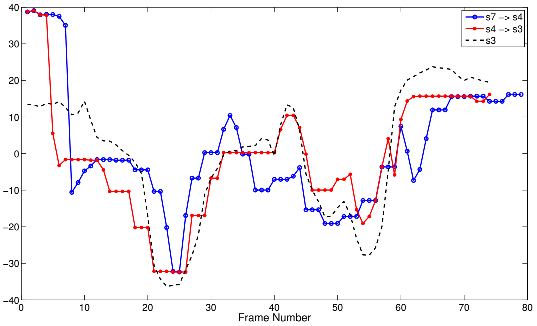


Fig. 6. s_7 warped to s_3 using our GWDTW method with the weights fixed at $w_0 = w_1 = w_2 = 1$.

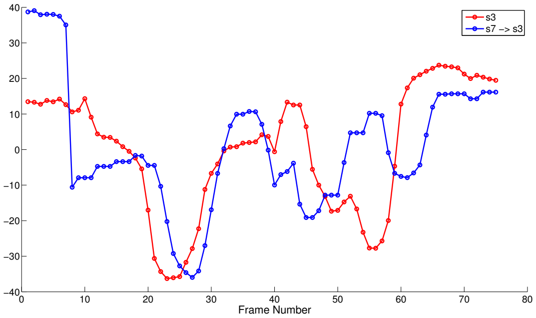


Fig. 7. s_7 warped to s_3 using WDTW with the weights w_1 and w_2 calculated automatically.

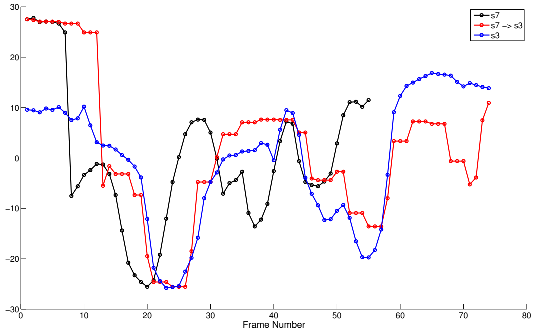


Fig. 8. s_7 warped to s_3 using our GWDTW method with the weights w_1 and w_2 calculated automatically.

fact s_7 but with the time base of s_3. The dashed black curve is s_3 shown for comparison with the warped version of s_7. It is clear from these results that in this instance a better warp is achieved by using the intermediate step of warping via s_4.

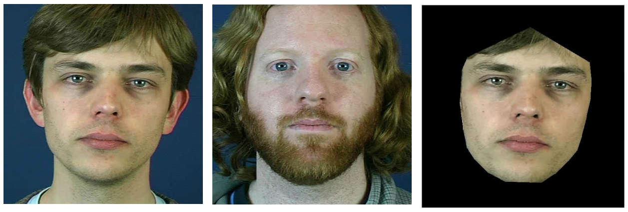
To test the automatic weight selection scheme described in Section II-B we automatically calculate the weights as described and then apply them to the WDTW and GWDTW methods. The results are shown in Figures 7 and 8 respectively. We find that the GWDTW method still provides a better alignment than the WDTW. Also it should be noted that the path GWDTW takes through the graph has now been altered (because of different weights) and now maps directly from s_7 to s_3.

Figure 9 contains example frames from two original sequences, s_7 and s_2, and the resulting resynthesised sequence when s_7 is warped to s_2. The actors are saying the phrase “Once upon a time”, however, s_2 is elongating the word “once” while s_7 is not emphasising any words. Because of this the sequences are of different lengths, s_7 has 55 frames and s_2 has 68 frames. Applying our GWDTW method to the shape parameters of s_7 and s_2 allows us to produce a resynthesised sequence for s_7 which now has the same temporal dynamics as s_2 and is also of the same length. Frame 15 (Figure 9(b)) indicates the opening of the mouth for the word “once” in both, but by frame 33 (Figure 9(c)) s_7 is beginning to pronounce the word “upon”. In the resynthesised sequence, the dynamics of s_7 have now been aligned with s_2.

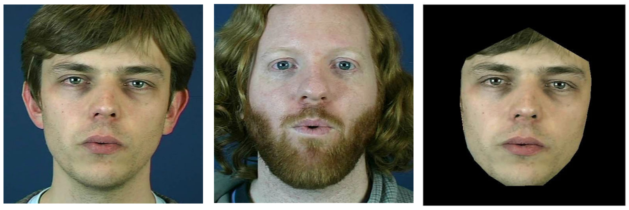
VI. Conclusion

In this paper we have described a novel approach for building models of temporal dynamics. Our application is to generate stimuli, consisting of videos of talking heads, for experiments to investigate how the perceived trustworthiness of individuals is affected by alterations to their temporal dynamics. Image sequences were represented as trajectories in face space, and aligned with GWDTW, our improved version of dynamic time warping. GWDTW was shown to perform better than the standard DTW, and could successfully map the dynamics of one speaker onto another.

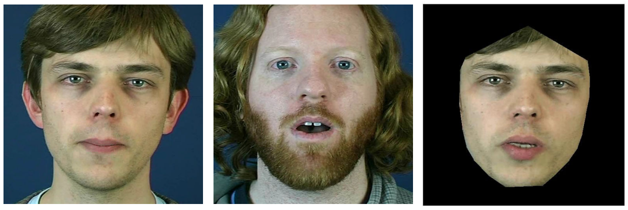
Current work is looking at the warping of audio, using the mappings determined from the alignment in face space. Early work has shown promising results. Our method will then be used to generate stimuli for the initial perceptual experiments. Finally, we plan to build and exploit PCA models of temporal dynamics for further perceptual experiments.



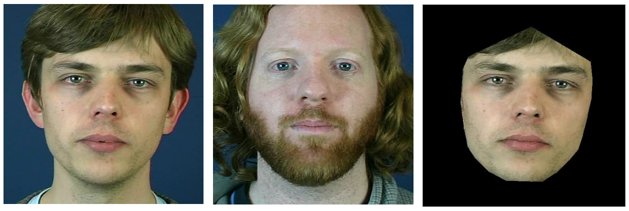
(a) Frame 1



(b) Frame 15



(c) Frame 33



(d) Final frame

Fig. 9. Frames from two original sequences and the resulting resynthesised sequence. Left column s_7, Middle column s_2, Right column the resynthesised sequence of s_7 warped to s_2.

REFERENCES

[1] T. Cootes, G. Edwards, and C. Taylor., "Active appearance models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 681–685, 2001.

[2] L. Benedikt, D. Cosker, P. L. Rosin, and D. Marshall, "Assessing the uniqueness and permanence of facial actions for use in biometric applications," *IEEE Transactions on Systems, Man and Cybernetics, Part A*, vol. 40, no. 3, pp. 449–460, 2010.

[3] L. Williams, "Performance-driven facial animation," in *SIGGRAPH '90: Proceedings of the 17th annual conference on Computer graphics and interactive techniques*, 1990, pp. 235–242.

[4] D. Vlasic, M. Brand, H. Pfister, and J. Popović, "Face transfer with multilinear models," *ACM Transactions on Graphics*, vol. 24, no. 3, pp. 426–433, 2005.

[5] Q. Zhang, Z. Liu, B. Guo, D. Terzopoulos, and H. Shum, "Geometry-driven photorealistic facial expression synthesis," *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 1, pp. 48–60, 2006.

[6] Y. Chang and T. Ezzat, "Transferable videorealistic speech animation," in *SCA '05: Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on Computer animation*, 2005, pp. 143–151.

[7] D. Cosker, R. Borkett, D. Marshall, and P. L. Rosin., "Towards automatic performance-driven animation between multiple types of facial model," *Computer Vision, IET*, vol. 2, no. 3, pp. 129 –141, Sept 2008.

[8] B. Theobald, I. Matthews, M. Mangini, J. Spies, T. Brick, Z. Ambadar, J. Cohn, and S. Boker., "Mapping and manipulating facial expression," *Journal of Language and Speech*, vol. 52(2/3), pp. 369–386, 2009.

[9] M. de la Hunty, A. Asthana, and R. Goecke., "Linear facial expression transfer with active appearance models," in *2010 International Conference on Pattern Recognition*, 2010.

[10] M. Muller, *Information Retrieval for Music and Motion*. Springer, 2007.

[11] E. Keogh and M. Pazzani, "Derivative dynamic time warping," in *First SIAM International Conference on Data Mining*, 2001.

[12] H. Sakoe and S. Chiba., "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 26(1), pp. 43–49, 1978.

[13] F. Itakura., "Minimum prediction residual principle applied to speech recognition," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 23(1), pp. 67–72, 1975.

[14] E. W. Dijkstra, "A note on two problems in connection with graphs," *Numerische Math*, vol. 1, pp. 269–271, 1959.

[15] K. Sidorov, S. Richmond, and D. Marshall., "An efficient stochastic approach to groupwise non-rigid image registration," in *Proceedings of Computer Vision and Pattern Recognition (CVPR)*, Miami, USA, June 2009, pp. 2208–2213.