

Towards Visuocomputational Endoscopy: Visual Computing for Multimodal and Multi-Articulated Endoscopy

A Dissertation
Presented to the Graduate School of Informatics
Kyoto University
in Candidacy for the Degree of
Doctor of Philosophy

by
Tuukka Karvonen
Kyoto University, Japan

Supervisor: Prof. Tomohiro Kuroda

Copyright © 2017 by Tuukka Karvonen
Kyoto University, Japan
All rights reserved.

Abstract

Towards Visuocomputational Endoscopy: Visual Computing for Multimodal and Multi-Articulated Endoscopy

Tuukka Karvonen
Kyoto University, Japan

2017

Endoscopy in the medical context aims to provide a view inside the human body. The tool for providing this view is called the endoscope. The principal applications of endoscopy are three-fold: the examination of various parts of the body based on symptoms exhibited, the confirmation of diagnoses through biopsies, and the enabling of surgical operations for treatment of a disease. Various different types of endoscopes exist, and the style of endoscope used is dependent on the clinical target. For example, examining the ear is done with the short-nosed tapered otoscope, while for respiratory tract examination, a long, thin flexible bronchoscope is used. Recently, endoscopy has paved the way for minimally invasive surgery, an approach to surgery that aims to minimize the harm caused to the patient by the surgical procedure. Minimally invasive surgery chiefly differs from traditional open surgery in that surgical tools are inserted through small incisions, rather than the surgeon operating in an opened cavity, and in that vision is provided mainly by the endoscope. Thus, it can also be said that the rising popularity of minimally invasive surgical techniques brings forth a need for better tools for visualization.

Research and development of endoscopes can roughly be divided into two categories: hardware, and computational innovations. The hardware side focuses on developing novel hardware designs related to both the structure of the endoscope and the sensing capabilities. For example, the capsule endoscope used for imaging the digestive tract differs greatly from traditional endoscope shapes. For sensing capabilities, both ultrasound and optical coherence tomography probes that allow for tomographic imaging close to the clinical target have been integrated to endoscopes. The computational side is covered under visual computing, which is an umbrella term covering all aspects of visual information handling, such as image processing, computer vision and computer graphics. Due to the recent advances in both available computing power and visual computing methods, it has found many applications in the field of endoscopy. Through various information processing techniques, visual computing can help provide navigational and clinical decision support to the user, potentially leading to increased accuracy of diagnoses, and reduction in the time taken for endoscopic operations. Visual computing methods can be also applied during

the development of novel endoscopes through methods like computer-aided design. Finally, virtual training systems have recently entered the market, which employ computer graphics to provide an environment for endoscopic and surgical training.

In this thesis, we explore the possibilities for applying visual computation in the field of endoscopy. Due to the two fields having become increasingly intertwined, we propose to call this intersection of the two fields, “visuocomputational endoscopy”. We showcase the challenges and opportunities in this new field through two approaches.

In the first one, we have developed methods for the support of a novel endoscope that combines traditional visible light imaging with optical coherence tomography imaging. We develop volumetric methods to synthesize images from these two imaging modalities, and to assist in clinical decision making in measuring the degree of endolymphatic hydrops, a disorder in which endolymphatic fluid accumulates and causes morphological changes in the inner ear.

In the second approach, we apply visuocomputational methods to aid in the development and usability evaluation of a novel endoscope that is composed of multiple joints allowing for more freedom in viewpoint selection and navigation. We focus on creating an accurate virtual model of the endoscope concept, and then evaluate it through a user experiment. We also propose a highly immersive augmented virtuality platform for evaluation of similar virtual prototypes.

Through the above-mentioned approaches, we validate the visuocomputational endoscopy concept, and show that coupled with hardware advancements it has the power to realize better and more efficient examination, diagnosis, and surgical methods for endoscopy.

Contents

1	Introduction	1
2	Visuocomputational endoscopy	5
2.1	Endoscopy	5
2.1.1	Modern endoscopy	6
2.2	Recent developments	7
2.2.1	Hardware-based developments	7
2.2.2	Computational developments	8
3	Texture mapping for dual-mode en face-OCT endoscope system	11
3.1	Introduction	11
3.2	Optical coherence tomography	12
3.2.1	Interferometric imaging	13
3.2.2	Challenges	16
3.3	Endoscopic and multimodal OCT imaging	19
3.4	Proposed method	19
3.5	Results	20
3.6	Discussion	20
3.7	Conclusion	21
4	Estimating the degree of EH using optical flow	23
4.1	Introduction	24
4.2	Related work	25
4.3	Theory	26
4.3.1	Cochlea	26
4.3.2	OCT	31
4.3.3	Optical flow	31
4.4	Method	32

4.5	Results	34
4.6	Discussion	35
5	Segmentation of OCT images using diffusion maps	41
5.1	Introduction	41
5.2	Background and related work	42
5.2.1	Endolymphatic hydrops	42
5.2.2	Cochlear imaging	42
5.2.3	Segmentation	43
5.3	Proposed system	45
5.4	Results	45
5.5	Discussion	46
5.6	Conclusion	46
6	Semi-Automatic Measurement of EH from Volumetric OCT Images	51
6.1	Introduction	51
6.2	Related work	53
6.3	Methods	54
6.3.1	Preprocessing and segmentation	56
6.3.2	Angle calculation	57
6.4	Validation	58
6.4.1	Initial value generation for DFM method	60
6.4.2	Initial value generation for SLB method	60
6.5	Results	60
6.6	Discussion	62
6.7	Conclusion	64
7	Evaluation of a novel multi-articulated endoscope: proof of concept through a virtual simulation	65
7.1	Introduction	66
7.2	Methods	67
7.2.1	Endoscope structure	69
7.2.2	Controls	70
7.2.3	Evaluation	70
7.3	Results	75
7.4	Discussion	76
7.5	Conclusion	78

8	Augmented virtuality platform for usability evaluation of a novel endo- scope concept	81
8.1	Introduction	81
8.2	Methods	83
8.3	Evaluation	84
8.3.1	Mapping between real/virtual world	84
8.3.2	Accommodativeness	84
8.3.3	Models' suitability for concept	85
8.4	Discussion	85
8.5	Conclusion	86
9	Discussion	91
9.1	OCT project	92
9.2	MAE project	94
10	Conclusion	95
10.1	Contributions	95
10.2	Future direction	96

List of Figures

1.1	Structure and relationship between the chapters of the dissertation.	3
2.1	Structure of a rigid endoscope.	6
2.2	Structure of a flexible endoscope.	7
2.3	Putting different aspects of endoscopic research and development into a context.	9
3.1	Schematic of the en face camera / OCT endoscope.	12
3.2	An example of constructive and destructive interference.	15
3.3	A schematic of Young’s double-slit experiment.	16
3.4	An example OCT B-scan image of a cochlea.	17
3.5	The structure of a typical (time domain) OCT machine.	18
3.6	Structure of the algorithm.	20
3.7	Three phases of the algorithm. Top left: original slice of the OCT volume. Top right: Resulting point cloud before extracting the largest connected component. Bottom: end result, with the top layer extracted and the texture mapped to it.	21
4.1	General structure of the ear in humans.	26
4.2	A microscope image of a mouse cochlea, with all the relevant features labelled.	28
4.3	A detailed cross-sectional view of the structure of the cochlea, showing the small features inside the tunnels. Note that the <i>scala media</i> here is labeled <i>ductus cochlearis</i> , and the Reissner’s membrane is referred to as the vestibular membrane. Image obtained from a public domain version of Gray’s <i>Anatomy of the Human Body</i> [1].	29

4.4	Histological slices imaged by a microscope, obtained from a cochlea with no hydrops (left side) and with severe hydrops (right side). The red area represents the <i>scala vestibuli</i> , the green area the <i>scala media</i> , and the blue area the <i>scala tympani</i> . The colorings were added afterwards to highlight the features. Reissner's membrane is the membrane between the red and the green areas.	30
4.5	Upper: Cochlear structures visible in a histological frame. Lower: an OCT slice of the cochlea, from roughly the same area as the above histological image.	37
4.6	Example of the apparent movement of the structures between two frames. Due to the oblique angle of the Reissner's membrane in comparison to the axes, it exhibits apparent movement when browsing from frame to frame. The optical flow algorithm keeps track of the position of the membrane. . .	38
4.7	An example of an OCT image with the user inputted points and the planes the software calculates	38
4.8	The dimensions of the captured data. X is the amount of axial scans, which in our study was 1024 over 6 mm physical length. Y is the depth of the scan, and Z the amount of slices scanned. Both were set as 256 over 1.5 mm.	39
4.9	Plotting the obtained angle & ratio values in a scatter plot. Two distinct groups can be seen, indicating correlation between the two different measures.	39
5.1	The structure of the cochlea and effect of endolymphatic hydrops on it. . .	43
5.2	Change in the position and structure of the Reissner's membrane between a healthy and an afflicted cochlea.	44
5.3	The flow of the algorithm.	45
5.4	Illustration of selected steps along the process. Red arrow indicates the position of the Reissner's membrane. Note that in (b), the superpixel intensities & structure are also affected by the neighboring slices.	48
5.5	A mesh created from the segmented image, visualized from different directions. Reissner's membrane is highlighted.	49
6.1	Differences between a healthy and afflicted cochlea. Dyed histological slices from mouse cochleae imaged by a microscope. Red color corresponds to <i>scala vestibuli</i> , green to <i>scala media</i> , and blue to <i>scala tympani</i> . A: Sample from a healthy cochlea. B: Sample from an afflicted cochlea.	52

6.2	Structure of the algorithm. Various steps are visualized from reading the OCT images to the final, calculated feature angles. Note that although the shown images are 2-dimensional slices, with the exception of median filtering and the illumination adjustment, all of the processing is done in three dimensions. 1: OCT image after median filtering. 2: Brightness constancy adjustment. 3: Superpixel segmentation. 4: K-means segmentation in diffusion space. 5: Segmenting the two scalae with s/t-cut. 6: Edge pixel labeling. 7: Largest component extraction (with scala vestibuli in red, scala media in green). 8: Wall/membrane segmentation with different colors corresponding to different segments. 9: Fitting planes to individual segments, with plane color corresponding to respective segments. 10: Average planes from step 9, with green corresponding to wall, red to segment.	55
6.3	Features, ground truth values, and estimated features for each dataset. For ground truth values, magenta corresponds to the cochlear wall and teal to the membrane. For estimated features, yellow marks the ground truth, green the DFM method, and red the SLB method.	59
6.4	Results for the evaluation. Top: calculated similarities for the wall feature over the datasets. Middle-top: calculated similarities for the membrane feature over the datasets. Middle-bottom: calculated variances for the wall feature over the datasets. Bottom: calculated variances for the membrane feature over the datasets.	61
6.5	Results averaged over datasets. Average similarities and variances for each method, averaged over the datasets.	62
7.1	Concept behind the movement of the endoscope. As the head of the endoscope is moved through the push/pull and angle adjustments, the rest of the body conforms to the route. This allows the endoscope to both avoid obstacles, and show the target from head on even if it is at an angle. . . .	68
7.2	The test setup. The controller, screen as well as the software are shown. . .	71
7.3	A screen capture of the test, showing the targets the participants were asked to navigate to. Numbers indicate the order in which the targets were shown to each participant.	73
7.4	The course of events for the test for each participant. After testing both of the endoscopes, the participants answered a short questionnaire to evaluate their experience.	74

7.5	A combined scatterplot-boxplot of the angles achieved (left) and times taken (right) per task for each endoscope.	75
7.6	Bland-Altman plots showing the relation between differences and means for the two endoscopes for both angles and times elapsed for the tests. For calculating the differences, a positive value indicates that the score was larger for the rigid endoscope.	76
8.1	Overview of the components of the system.	87
8.2	Top: the virtual stereoscopic view towards the endoscopic target provided by the HMD. Bottom: view in the physical world from the same viewpoint.	88
8.3	The endoscopic view inside the target, with current endoscope shape overlayed.	89
9.1	The four aspects of visuocomputational endoscopy: development of novel endoscopic methods, training for endoscopic procedures, operational support, and decision support.	92

Acknowledgments

First and foremost, I would like to express my gratitude to Prof. Tomohiro Kuroda of the Division of Medical Informatics. Without him originally accepting the responsibility of hosting me while I was doing a short-term exchange program at Kyoto University, this thesis would likely never have come to existence. I would like to further thank him for encouraging me to apply for the doctoral program, for providing guidance throughout it, and for helping me take that step back from the details to see the big picture whenever it was needed. Secondly, I would like to extend my sincere gratitude to the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan for providing the generous scholarship that enabled these studies. For selflessly donating their time and efforts to act as referees for the thesis, and for the feedback they provided, I am grateful to Prof. Tetsuya Matsuda and Prof. Nobuhito Ohte. Without their help the thesis certainly wouldn't be where it is right now.

I am grateful to Dr. Tatsunori Sakamoto and Dr. Toshihiko Sato, the two main enablers and collaborators of the research, for both giving me the opportunity to work on interesting projects, and for their support on the clinical medicine aspects of the research. In a similar vein, Dr. Yosuke Tona helped me a lot in both preparing the cochlea samples for imaging, and for doing the actual imaging part. I would like to thank Dr. Yusuke Muranishi for helping with planning of the MAE, and helping organize the user test for the multi-articulated endoscope, and further would like to extend my thanks to all of the participants of the user test for donating their time. Dr. Christian Vincenot provided fresh perspectives on research, and was a source of knowledge regarding the doctoral course at Kyoto University.

Two people whom I would be amiss not to mention are Drs. Kazuya Okamoto and Naoto Kume, who were there to support me from the beginning in both academic and daily life issues. Dr. Yuki Uranishi provided a lot of advice on image processing and analysis, especially during the early stages of the doctoral program. Dr. Goshiro Yamamoto took a progressively larger role in helping the thesis be finished on time. I would also like to extend my thanks to rest of the Medinfo team for providing a diverse and inspiring environment for doing research, including the faculty members Drs. Masayuki Nambu, Hiroshi Tamura, Hisashi Saito, Genta Kato, Osamu Sugiyama, and Shusuke Hiragi, and the students Samar El Helou, Yu Maruyama, Toshikazu Furuya, Takeshi Imanaka, Misa Esashi, Yuta Fukushi, Kensuke Morris, Tatsuya Miyamoto, Kana Eguchi, Karin Goka, Shoko Ueno, Luciano Santos, Shu Teki, Tomohide Iwao, Jun Hatanaka, Yohei Yamasaki, and Ryutaro Fuji.

This section would also be incomplete without the mention of the Medinfo secretaries, Yoko Hara, Yuko Furusawa, and Kaori Shiomi, who tirelessly helped me with the administrative aspects of progressing on the course, and more than once saved me from being late for a deadline.

I am indebted and eternally grateful for my parents Ari and Kaisu, and brothers Teemu and Tuomas for supporting me with my decisions throughout my life and being there when I needed them. Finally, for my fiancée Misa for her unending support, for being there both for the good and the bad, and for being understanding of the sometimes hectic and unreasonable research life, I extend my deepest gratitude.

Tuukka Karvonen
Kyoto, August 18, 2017

Chapter 1

Introduction

Medical imaging is the application of imaging techniques for medical purposes. In medical imaging, various methods are used to capture and create visual representations of the body for further analysis. A canonical example of such a method are X-rays, which have been used for imaging the human body since the early 20th century. It was put to use almost immediately since it's discovery in 1895 by Wilhelm Conrad Roentgen, in the detection of various types of gastrointestinal cancer. Mammography also sprung up as a diagnostic technique due to the discovery of this important technique [2]. After the discovery of X-rays, various other methods for diagnostic imaging such as computerized tomography (CT) and magnetic resonance imaging (MRI) in the 1970s have taken hold and become commonplace in hospitals. Nowadays it is hard to find a part of body for which examination through diagnostic imaging wouldn't be of help in some way.

Another branch of medical imaging that has its roots even earlier in history is endoscopy, the practice of visualizing the inside of the human body with a specialized device, called the endoscope. The first endoscope was developed in 1806 by Philipp Bozzini when he invented what he called a "Lichtleiter", with the specific purpose of examining the human body. Since then, the endoscope has become an increasingly important tool in not just diagnosis, but also for visualization during surgery. Most of the endoscopes in hospitals today can be classified to either rigid endoscopes, which transmit the light through lenses and rods to a digital camera sensor, or to flexible fibroscopes, which transmit light through optic fibers. Recently, a technique called minimally invasive surgery, or keyhole surgery, has increased in popularity. Instead of large scale invasive operations to access a surgical target, a minimal incision is made, through which tools such as the endoscope can be inserted. Minimizing the incisions results in reduced hospital stays and faster recovery times for the patient.

Visual computing is the umbrella term covering all aspects of computer science dealing

with visual information. Increasingly, fields such as image processing, image analysis, computer graphics, virtual reality and augmented reality have become entangled, and designing solutions for practical real-world issues often involves a plurality of these disciplines. Further, as visual computing methods usually involve interactions with a human to a high degree, as a field it covers human-computer interaction as well. Increasingly, as hardware improvements make it possible to employ different imaging modalities in endoscopic imaging, visual computing methods are used to extract more value out of these systems. From simple image processing techniques such as contrast enhancement to improve the visibility of features of interest, to the application of advanced computer vision and machine learning techniques for automatically recognizing and classifying tumors, the potential for visual computing to help the medical professional in diagnosis is enormous. In addition, visual computing applications to endoscopy are not limited just to image processing and analysis. Various endoscopy trainers have been built which try to accurately simulate the actual endoscopic procedure and allow for on-demand training. Also, augmented reality can be used to overlay organs on the endoscopic image feed to aid in surgical navigation.

In this dissertation we propose a model for applying visual computing techniques to support endoscopy. The model is composed of three parts: low level image processing support, where the visual data from the endoscope is modified to highlight some features, image analysis, where some specific features are analyzed to provide clinical decision support, and computer graphics, where the data is presented to the user in a easily understandable way, or additional information in the form of e.g. navigational support is delivered to the user. We name this intersection between the two fields *visuocomputational endoscopy*. The investigations are done from two distinct perspectives: Firstly, to help in image processing and provide support for clinical diagnosis when using a novel endoscope currently in development that combines traditional digital endoscope imaging with an OCT imaging sensor. Secondly, we apply visual computing techniques to the development and evaluation of a virtual prototype of a novel multi-articulated endoscope. Describing these perspectives from the model's point of view, in the former we deal with image processing and analysis, and in the latter, with computer graphics.

To these ends, in Chapter 2, we look at the background behind the research that was done, and explain briefly the history of endoscopy as well as some modern applications. In Chapter 3 we present a novel method for synthesizing data based on an endoscope currently in development that integrates both a traditional digital endoscope camera sensor as well as an OCT sensor. In Chapter 4, we present the first tool created for measuring endolymphatic hydrops. We go through the methods used, as well as some results obtained by performing measurements with the it. Chapter 5 builds upon the work of Chapter 4, and focuses on

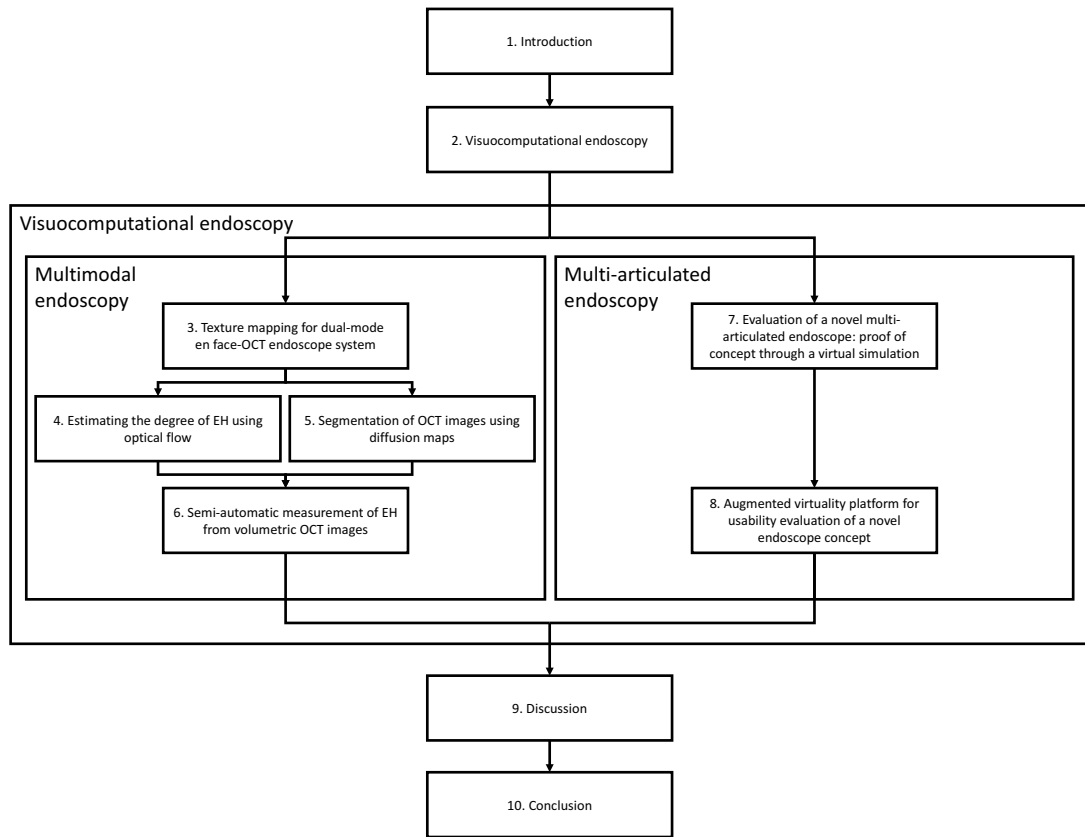


Fig. 1.1: Structure and relationship between the chapters of the dissertation.

asegmentation and visualization method of OCT data, particularly the cochlea and its small features. We explain how the algorithm works and show some example results produced by it. In Chapter 6, the segmentation method is applied, and another, more accurate method for diagnosis of endolymphatic hydrops is presented. Chapter 7 switches the focus to the development of a novel endoscope, and a usability test that was performed to confirm its feasibility. Chapter 8 introduces a platform for the evaluation of such an endoscope in an augmented virtuality environment. In Chapter 9, we discuss about the particular goals of the research, how well we feel we achieved them, what could've been done otherwise and what needs to be done in the future. Finally, in Chapter 10, we conclude the dissertation by summing up the progress made and considering some possible future directions. The structure and relationships between the Chapters can be seen in Fig. 1.1.

Chapter 2

Visuocomputational endoscopy

In this chapter we will go through the background behind the research done. We will start by providing a short history on endoscopy, and explaining conventional endoscopic methods. Then we will show some of the recent developments in the field. Finally, we will take a look at various applications of visual computing in the field of endoscopy.

2.1 Endoscopy

The father of modern endoscopy is considered to be Philip Bozzini, who built upon existing optical technology to create what he called the "lichtleiter", a device that used a candle, mirrors and a cylinder to direct light into the human body, and a lens that reflected the image from the target back to the eye through a separate cylinder. Using various examination shafts, the device could be applied to various cavities including the urethra, the female urinary bladder and gunshot tracts [3].

The device was revolutionary in the way it reflected the light from the imaging target back to the eye through the careful utilization of mirrors. Although it was successfully applied in clinical procedures, it still had a severe flaw in that the field of view was restricted to the width of the tube, a problem that was not solved until 70 years later by a Maximilian Carl-Friedrich Nitze [4], who utilized a wide-angle lens amongst a set of other lenses to visualize a larger portion of the target. Nitze used a platinum wire connected to a battery as the light source.

The next big innovation in rigid endoscopy was done by Harold Hopkins in 1954, who improved on Nitze's optical system by eliminating the lens-to-air interfaces through the use of glass rods that were placed between the lenses inside the endoscope. He also used better lenses at the ends of the endoscope [4]. Through better light transmission, the endoscope provided a brighter, clearer view to the user. The view from the endoscope was bright

enough to be documented by cameras [4]. A few years after this, Hirschowitz published a flexible gastroscope that consisted of fibers through which light was transmitted [5]. These two designs created by Hopkins and Hirschowitz are still the used as the base for manufacturing endoscopes today.

2.1.1 Modern endoscopy

Modern endoscopy has largely two approaches depending on the target of visualization: rigid endoscopy achieved through a rod-lens system, and flexible endoscopy achieved through fiber optics. Rigid endoscopes are mainly used when the target is relatively easily accessible from the point of entry in to the body, such as in nasal endoscopy and thoracoscopy. They offer a stable, clear view on the target, and often contain access channels to facilitate tissue sampling. They can also be operated with one hand. In contrast, flexible endoscopes require two hands, but generally have smaller diameter and have navigational advantages due to their flexibility. Flexible endoscopes are used in colonoscopies for visualizing the colon, as well as for bronchoscopy for visualizing the airways. In both cases the flexibility and length of the endoscope allows for inspection and tissue sampling with minimal invasiveness. In the following, we will explain the principle behind the two approaches to endoscopy.

Rigid endoscopy

The overall structure of a rigid endoscope can be seen in Fig. 2.1. It consists of the eyepiece the user (or a camera) looks through, a light channel to transmit the light from the light source to the target, several rods and lenses to ensure the optical quality, and the objective lens that determines the field of view of the endoscope.

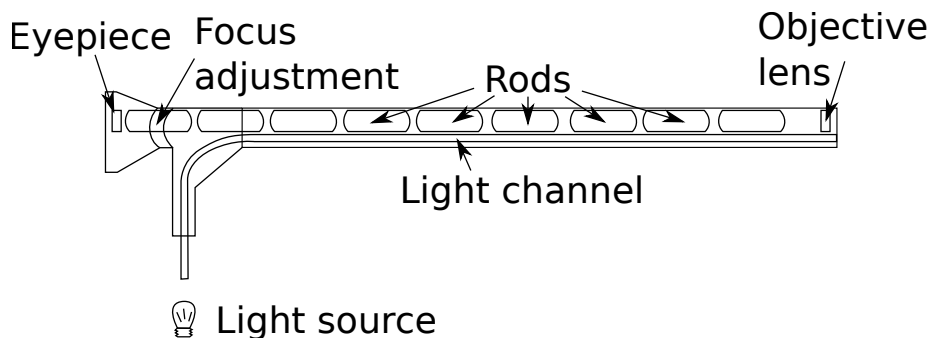


Fig. 2.1: Structure of a rigid endoscope.

Flexible endoscopy

The structure of the flexible endoscope is illustrated in Fig. 2.2. The overall structure is similar to the rigid endoscope, but instead of inflexible rod lenses, fiber optics are used to deliver the image to the user. Fibers of 5-10 μm in diameter are used to relay one "pixel" of information, and a coherent bunch of these fibers form the image. Much like in a rigid endoscope, separate channels are used to relay the lighting to the target. The flexible fibers allow for the distal end to be bent with 1 or 2 degrees of freedom.

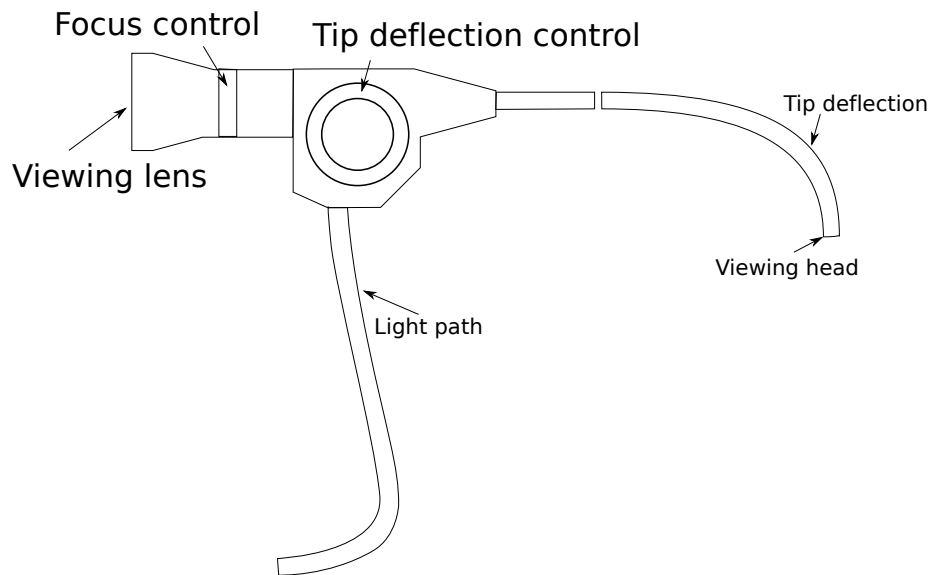


Fig. 2.2: Structure of a flexible endoscope.

2.2 Recent developments

Recent developments in endoscopy can be divided into two broad main categories: innovations based on developments in the hardware used for the endoscopes, and innovations based on the computational methods available that can be applied to the data obtained from those endoscopes. In many cases, progress happens hand-in-hand, by hardware discoveries enabling new ways to capture data from an endoscope, such as in the case of endoscopic ultrasound, and then new algorithmic methods arising which allow us to utilize more complicated data collection methods, such as volumetric ultrasound.

2.2.1 Hardware-based developments

As previously mentioned, the modern endoscope tends to be classified into either of two categories: a rigid, rod-lens type endoscope, or a flexible, fiberoptic endoscope. These

forms continue to be innovated upon, and application-specific variations on the basic ideas exist.

For imaging the digestive tract, capsule endoscopy is a technique where the patient swallows a small capsule housing an imaging sensor, and throughout the capsule's journey through the body images are taken, which can then be used for diagnosis of diseases such as Crohn's disease [6]. To offer a compromise between the stability provided by a rigid endoscope and the flexibility offered by a flexible endoscope, various articulated endoscopes have been developed [7, 8]. Practical implementation is often a challenge due to the constraints given by the imaging target, such as endoscope diameter, and especially for medical imaging the safety constraints, e.g. the reliability of the joint mechanism are an issue.

Another aspect of hardware-based development is using other modalities than the full spectrum of visible light that is usually used to provide the vision. Due to the optical properties of hemoglobin, narrow-band imaging can be utilized to highlight blood vessels. In narrow-band imaging, a specific band of light, commonly blue or green, is projected from the endoscope, and due to the hemoglobin absorbing these wavelengths, the blood vessels appear very dark in contrast to the neighboring tissue [9]. Endoscopic ultrasound [10] allows the ultrasound probe to be placed closer to the visualization target, thus obtaining a clearer view. A similar approach is taken in endoscopic optical coherence tomography (endoscopic OCT) [11], which provides a higher resolution with a narrower field-of-view.

2.2.2 Computational developments

Computational approaches to endoscopy largely focus on two areas: aiding the user in navigation, and helping the user in decision making. Visual computing is often applied here to process the raw data coming from the endoscope. Using visual computing for processing the raw data becomes more important as the amount of data generated by the endoscope increases, for example through usage of multiple modalities, or by using volumetric imaging enabled by e.g. an ultrasound probe. The overarching theme is that of aiding the user to understand the data provided by the endoscope.

A relatively simple way to aid the user in performing endoscopy is to apply image processing techniques such as contrast adjustment to highlight specific features in the imaging target [12]. To aid in navigation, augmented reality can be used to display nearby non-visible structures on the feed from the endoscope. Feature tracking methods can be used to track the movement of the endoscope, and further to monitor the movement and deformation of organs [13]. Computer-aided clinical decision making is also a target for active research, and can help clinical professionals discover potential malignancies within

the image [14].

Another way to apply visual computation to the field of endoscopy is not through operational support, but by providing virtual environments for training. Often these employ physical tools such as the endoscope and forceps, and present the user a virtual environment where their actions take place. Quantitative evaluation of performance is easy through various metrics that can be collected, and training can be done cost-effectively, irrespective to the time and available resources such as animal carcasses [15].

In a similar sense, visual computing methods are also used in the development of endoscopes. Computer-aided design (CAD) has applications in any kind of industrial design, and for endoscopy it can be used to model endoscope structures also.

Fig. 2.3 shows some of the above-mentioned hardware and computational aspects in the context of endoscopy, as a pipeline from the patient to the computer. It implies that although visuocomputational endoscopy is concerned with computational innovations, it is also highly dependent on the hardware solutions that enable the application of computational methods.

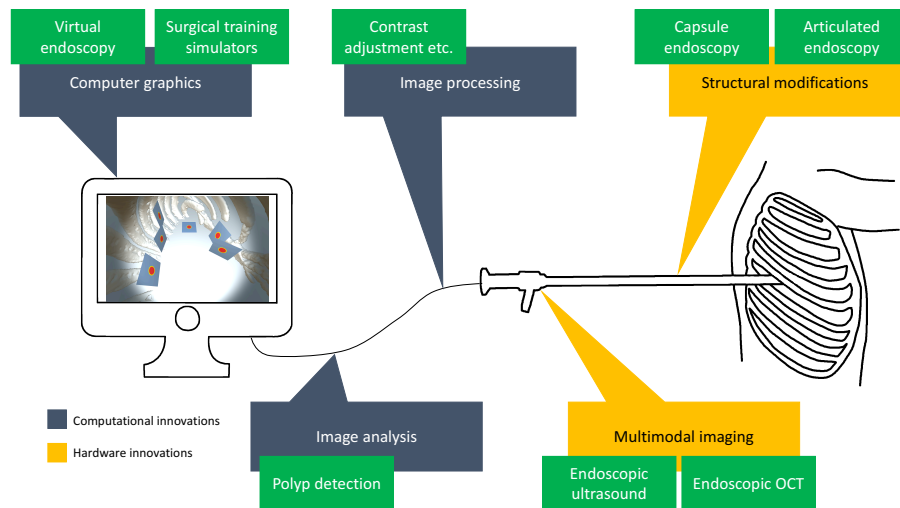


Fig. 2.3: Putting different aspects of endoscopic research and development into a context.

Chapter 3

Texture mapping for dual-mode en face-OCT endoscope system

For minimally invasive surgical operations, endoscopes are used to obtain a view to the surgical target. Traditional endoscopes require the operator to look through a small eyepiece, requiring considerable concentration and taking said concentration away from operation of other surgical equipment such as pincers, scalpels and such with which the actual operation would be done. Modern endoscopes make this operation a bit easier, having integrated digital camera sensors and enabling the endoscopic view to be projected to an external monitor. Still, the field of view presented by the endoscope is fairly narrow, and usually only the immediate surface of the tissue viewed is visible. Currently, a novel endoscope utilizing a technique known as optical coherence tomography is under development. With this tool it will be possible to simultaneously obtain both the conventional en face endoscopic view as well as a view inside the tissue, helping medical professionals avoid possibly dangerous areas of operation and to potentially identify cancerous cells from other benign tissue. We present an algorithm for processing the data produced by such a device that combines both of the imaging modalities and presents the user with the en face image mapped to the topological surface information obtained from the OCT image. This will help the medical professional to more clearly see the structure of the tissue, and thus help with making surgical decisions.

3.1 Introduction

In this paper we will present a system for handling the data from an endoscopic OCT system that combines traditional digital endoscopy and an OCT sensor with the purpose of providing a medical professional additional depth-related data when looking at soft

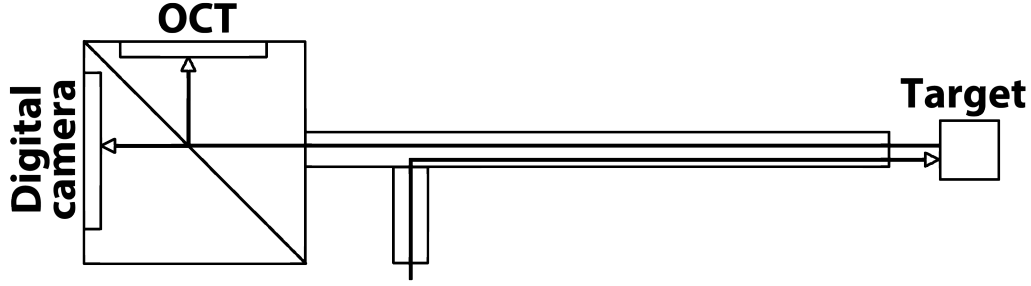


Fig. 3.1: Schematic of the en face camera / OCT endoscope.

tissue, such as the sinuses in rhinology.

When performing surgery, it could be valuable for the surgeon to be able to see beyond the immediate surface of the target tissue. For example in the case of nasal surgery, when operating near the bottom of the brain there is a risk of puncturing the thin bony layer separating the brain from the nasal cavity. In these cases, using a device such as OCT could work as an early warning system that the bony layer is about to be cut into.

Traditional digital rigid endoscopes usually come equipped with a CCD or a CMOS camera sensor. Light is projected from a separate arm to the target, from where the reflecting light is directed to a camera sensor at the other end of the endoscope.

Currently, a rigid endoscope with dual integrated camera sensor and optical coherence tomography sensor is under development. This produces both the volumetric view inside tissue as well as the traditional endoscopic en face view. To produce maximum value for the medical professional using this endoscope, image processing methods can be used to unify the two imaging modalities to for example add topological information to the en face image. The structure of the device can be seen in Fig. 3.1.

In this research the aim was to extract surface topology information from the OCT image, and map the corresponding en face image to that. We use visuocomputational methods to extract the foreground from the background, and then extract the topmost layer from the foreground as a point cloud, and finally map the en face camera image to it.

3.2 Optical coherence tomography

Optical coherence tomography is an imaging method that is comparable to ultrasound, with the distinction that instead of using soundwaves, it uses near-infrared light as the medium for transferring data. Due to the high propagation speed of light, the distance travelled by backscattered light signals cannot directly be measured, so interferometry is

used instead for constructing intensity profiles. OCT is extended from a technique used for measuring the axial length of the eye [16]. In 1991, Huang et al. [17] published an article in *Science*, titled "Optical coherence tomography", which can be considered the beginning of the usage of this technique. From its very inception, it has been applied in the field of ophthalmology, for visualization of cross sections of the anterior segment [18], for diagnosis of retinal diseases such as macular detachment, macular edema and such, [19], and especially for imaging the various layers of the retina [20, 21].

Although the focus of OCT imaging has from the start been on retinal imaging, a plurality of other applications also exist. Due to the relatively small size, the OCT device can be implemented in endoscopy, for example [22]. Other medical OCT applications include studying the differences in optical properties between basal cell carcinomas and healthy skin [23], investigation of contact dermatitis and psoriasis [24], where changes in optical properties between healthy and afflicted skin were also reported, and imaging dental structures [25] for cross-sectional imaging of dental structures. Recently, OCT devices have been used to image cochleae, from imaging microvascular perfusion in mice [26] to imaging the effects of endolymphatic hydrops in cochlea [27]. OCT has also found use in industrial settings, for helping with analysis in subsurface defect detection in ceramic materials [28] and material characterization of epoxy resin compounds, among others.

3.2.1 Interferometric imaging

OCT works on the principles of interference and coherence. Since the speed of light is too high to accurately discern the point an echo originates from via time gating like in ultrasound, more sophisticated methods are required. Thus, a system is devised where we can discern the distance travelled by returning light based on its interference with another, sample light path.

When two light waves are incident on the same point in space, the total intensity at that point is a pointwise sum of the individual waves, dependent on their phase difference. This is governed by the following equation:

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2} \cos \delta.$$

I is the total irradiance at a point P in space the two waves are incident at, I_1 and I_2 are the respective waves, and $\cos \delta$ the *phase difference* between the two waves. Thus, when

$$\delta = 0, \pm 2\pi, \pm 4\pi, \dots$$

$\cos \delta = 1$ and the phase-difference dependent term is maximised. This is referred to as

total constructive interference. Conversely, when

$$\delta = \pm\pi, \pm3\pi, \pm5\pi, \dots$$

the waves undergo *total destructive interference*, and the phase-dependent term becomes zero. An example of both is shown in Figure 3.2. On the left side, the peaks of the red signals match each other, creating the resulting green signal with increased amplitude. On the right side, the upper signal's peaks match the lower signal's valleys, creating the resulting green signal where the two signals have cancelled each other out. Between these terms, depending on the sign of the term e.g. which quartile the cosine term falls on, the waves undergo either *constructive* or *destructive interference* [29]. This was famously visualized by Dr. Thomas Young in the early 19th century, in what later became known as *Young's experiment*. He took a screen otherwise opaque save for a small pinhole in the middle (S_1), placed another screen with two of these pinholes side-by-side a little bit away from the first screen (S_2), and placed a third screen on which the light going through both of these earlier screens was projected (S_3). As the light waves propagate from the two pinholes to the projection screen, depending on the relative distance travelled, they go in and out of phase and form a *fringe pattern* (F) on the wall. Young's experiment and the resulting fringe pattern is visualized in Fig. 3.3 [29].

For a stable interference pattern to form, the two light sources must be *coherent*. In Young's case, the coherency is formed by the first screen, which forms a spatially coherent beam. In the usual case, however, a light source is composed of many different wavelengths of light combining together into white light. Due to the different frequency signals being mixed together, this causes a rapidly varying phase difference, averaging the $\cos\delta$ -term in the previous formula to 0, thus causing no fringe pattern to form. *Coherence length* is defined as the distance over which a signal acts in a coherent way, i.e. is predictable. Only light waves that have travelled equal distances within this coherence length can exhibit the interference patterns. For an ideally monochromatic light source, the coherence length would be infinite, thus allowing the pattern to form over any distance. The broader the bandwidth of the lightsource, the shorter the coherence length becomes. It is this feature of low-coherence light that allows OCT to achieve such high resolutions.

OCT devices use interferometry based on a light beam that is split in two using a 50/50 beam splitter. One of these resulting beams goes to a reference mirror, from where it is bounced back to a photodetector. The other beam is projected on the measured sample, where part of it backscatters and reflects back to the detector. Due to the low coherence light that is used in OCT, interference only occurs between the backscattered light from the

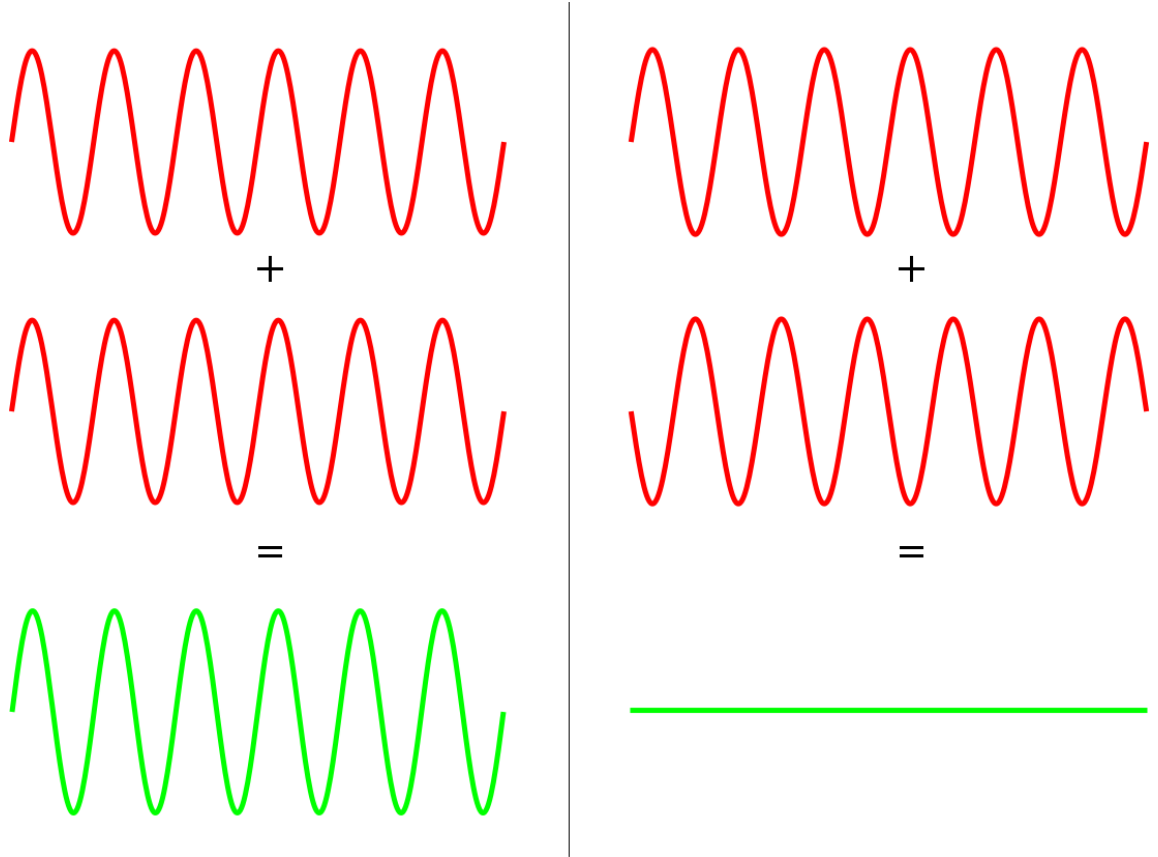


Fig. 3.2: An example of constructive and destructive interference.

sample that has travelled the same distance as the reference arm light. Thus, by moving the reference arm, interfaces at different depths can be imaged. One swipe of the reference arm is called one *axial scan* or A-scan. By moving the detector in a direction perpendicular to the axial scan, one can obtain a B-scan, or a 2-dimensional cross sectional image of the sample. By combining several of these B-scans, a volumetric representation of the data obtained from the sample can be created. A typical OCT B-scan of a cochlea can be seen in Fig. 3.4.

A schematic of an OCT device is presented in Fig. 3.5. This type of OCT device with a moving arm is called time-domain OCT (TD-OCT). The very first OCT devices were like this. However, due to the mechanically moving reference arm, the scan speed of TD-OCT devices is relatively low. The first clinical OCT devices were limited to 400 A-scans per second [30]. In the early 2000s, OCT devices that worked in the frequency domain were first introduced, with much higher scan speeds due to the now fixed reference arm. [31] For frequency domain OCT, either a broadband light source or a swept narrowband light source are used, and accordingly the methods are called spectral-domain OCT (SD-OCT)

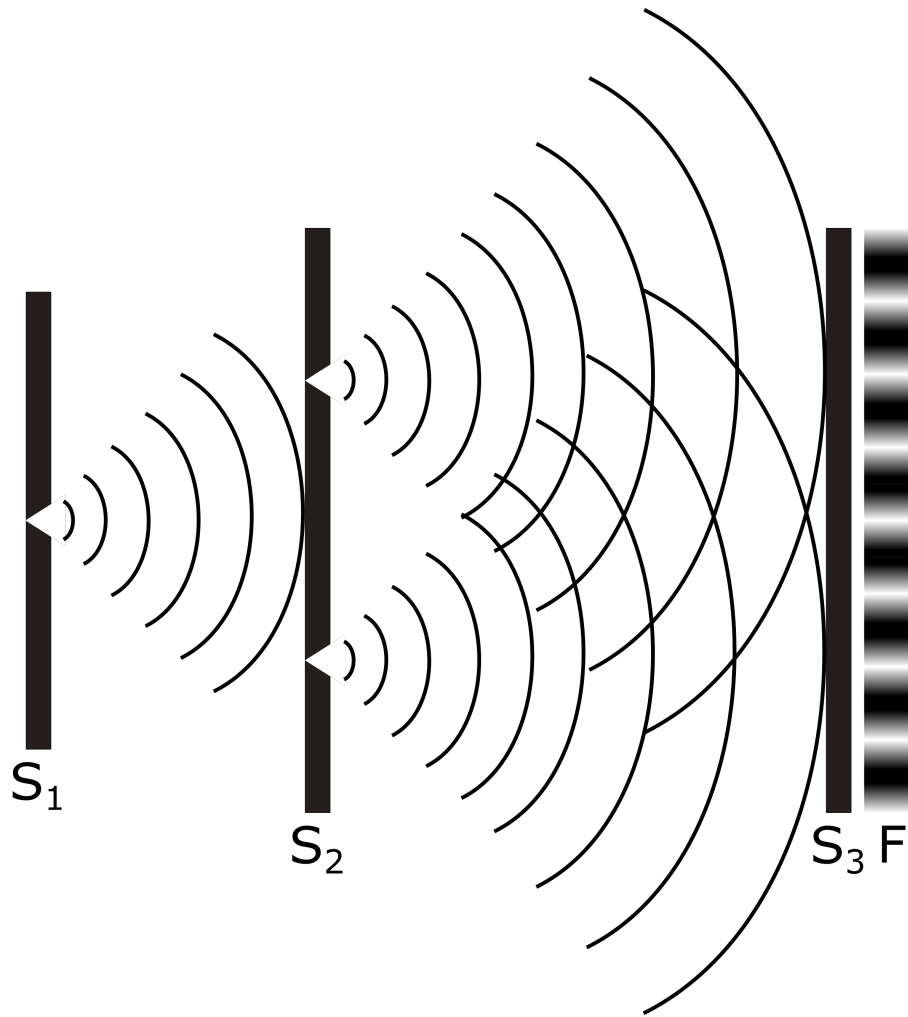


Fig. 3.3: A schematic of Young's double-slit experiment.

or swept-source OCT (SS-OCT). Generally, for both SD-OCT and SS-OCT, the scan rates are in the hundreds of thousands of A-scans per second, enabling reasonable times for obtaining volumetric scans [32, 33].

3.2.2 Challenges

Main challenges of OCT imaging are caused by the random scattering of light. This causes both signal attenuation that is dependent on the depth and substance imaged, as well as noise due to multiple-scattered light.

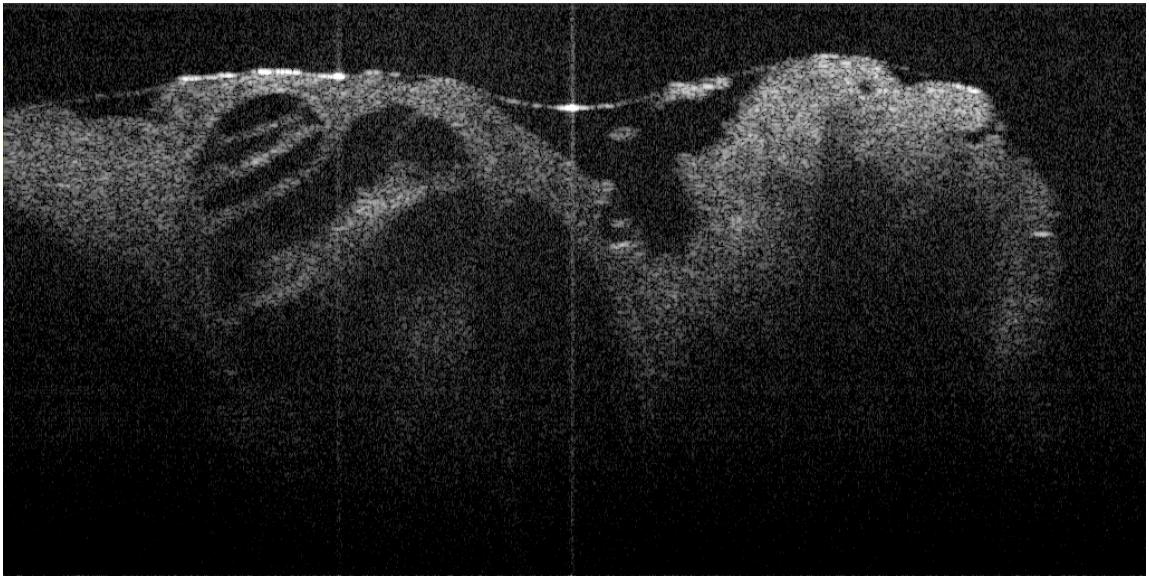


Fig. 3.4: An example OCT B-scan image of a cochlea.

Noise

Optical coherence tomography exhibits a type of noise called speckle noise. It is inherent to the way OCT signal is formed, and it is caused by random delays between backscattered light inside the sampled tissue. As long as at least two individual light beams that return to the detector in the OCT device with a phase difference of less than the coherence length of the signal, they can undergo constructive or destructive interference resulting in the intensity variation within an otherwise uniform tissue sample. The speckle in OCT images can be split into two components: signal carrying speckle, and signal degrading speckle. Signal carrying speckle is formed by single backscattered light arriving to the sensor, while the latter is caused by interference between multiple forward- and backscattered photons that reach the detector. Due to this duality of speckle, it can thus be considered as both the signal and the noise component. [34] Differences in speckle patterns between tissues have been used to segment OCT images. [35,36] On the other hand, the noise component can cause misdiagnoses due to masking relevant image features [37].

Reduction of speckle can be achieved by both software and hardware methods. Hardware methods involve generating uncorrelated speckle patterns and averaging over them. This can be achieved for example by utilizing temporal, spatial, frequential or polarization diversity. [38] The problem in these approaches becomes generating images with varying speckle patterns but with minimal structural changes in the object imaged, as any uncompensated movement of the sample or the OCT device reduces the resolution in the final image [39].

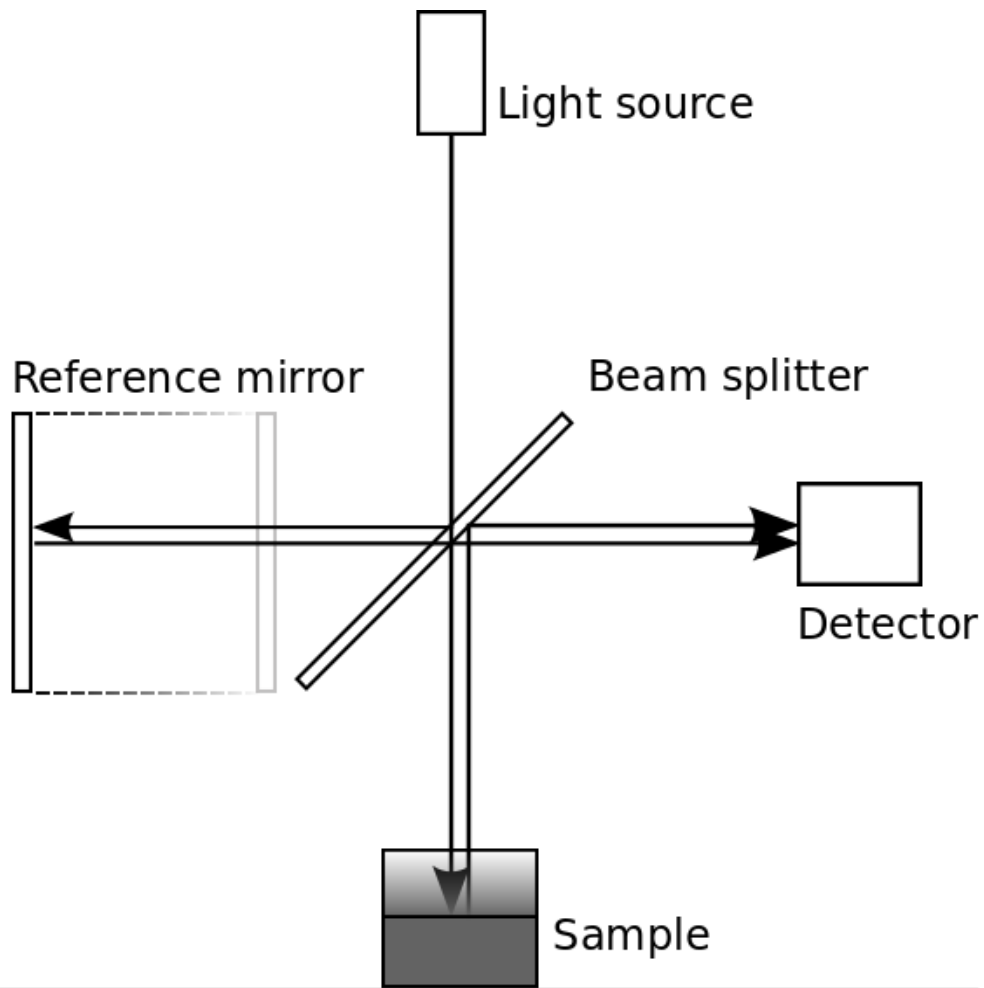


Fig. 3.5: The structure of a typical (time domain) OCT machine.

Software approaches for noise reduction range from spatial averaging over a window [40], wavelet thresholding [41], to anisotropic diffusion [42]. A good overview of different noise reduction methods in general can be found in Szkulmowski et al. (2012) [39].

Signal degradation

Generally the deeper into tissue we look in OCT, the less contrast the images exhibit. As OCT collects as signal the backscattered light that bounces back from the imaged sample to the device, there are mainly two reasons for the loss of contrast. Firstly, as the imaging depth is increased, less and less light actually penetrates deep enough into the tissue, allowing for backscattering. The second reason is due to the different types of light coming back to the sensor. The light can be either of single scattering or multiple scattering type. In single scattering the light has reflected directly back from the sample, and as such due

to the coherence gating likely originates from the point that is being imaged. Multiple scattering refers to the amount of light that has scattered multiple times, bouncing from particle to particle, and eventually due to random chance reaches back to the sensor. The latter type is considered noise and becomes increasingly strong the deeper into tissue is being imaged [43].

3.3 Endoscopic and multimodal OCT imaging

Optical coherence tomography is an imaging method that is analogous to ultrasound imaging, with the exception that instead of using sound waves it uses near-infrared light to capture images. It can obtain high resolution, 2-10 μm volumetric images from up to about 2 mm inside soft tissue. It is mostly applied in the field of ophthalmology for imaging the different layers of the retina, to detect diseases such as macular degeneration. Here, image processing methods for delineating these different layers have been applied, and is a popular field of research. Applied methods include graph-theoretic methods [44] and graph-cuts and dynamic programming [45]. Diffusion maps have also been applied before for OCT image segmentation of retinal layers [46].

There have been numerous applications of endoscopic OCT, such as for gastrointestinal visualization [11] of rabbits via a catheter-endoscope, for real-time imaging of human gastrointestinal system [47], and for thickness measurements and bladder tissue imaging [48]. Multimodal OCT has been achieved with an integrated photoacoustic microscopy-OCT sensor [49]. Recently, a simultaneous chromoendoscopy-OCT system was developed for gastrointestinal imaging [50].

3.4 Proposed method

The structure of the algorithm can be seen in Fig. 3.6. We take the OCT volumetric image, and to allow for fast processing, reduce the image to a quarter of its size using trilinear filtering. For further preprocessing, we do noise reduction using a median filter. Then, to further reduce the processing time without sacrificing much detail, we perform superpixelation using SLIC superpixels [51]. Then using the intensities and coordinates of the superpixels we apply the diffusion map algorithm to obtain the diffusion coordinates, which we cluster into two groups, the background and foreground by k-means clustering. Here we assume that the biggest foreground component corresponds to the main feature, so by extracting only the largest connected component using connected component labeling, and further by going through the resulting volume slice-by-slice, column-by-column, selecting

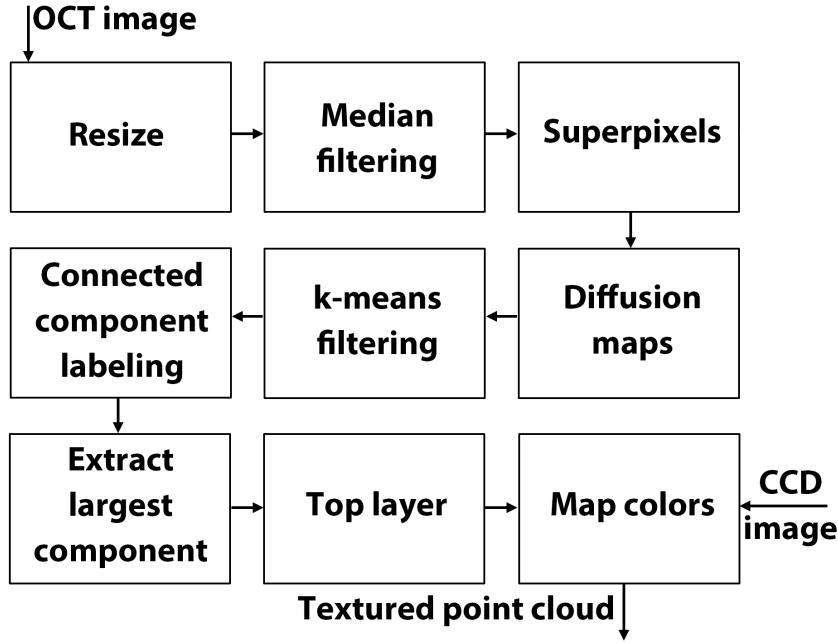


Fig. 3.6: Structure of the algorithm.

only the pixels closest to the endoscope (corresponding to the smallest y-value) we extract the top layer. After this is done, we take the en face image, and match the obtained OCT topological information to the color information from the camera sensor.

3.5 Results

To evaluate the algorithm, we used an OCT volumetric image obtained from a guinea pig cochlea as the target. The OCT system used was a Thorlabs OCS1300SS with a reportex axial resolution of $12\ \mu\text{m}$ over air, and a lateral resolution of $25\ \mu\text{m}$. The imaging area is 5 mm by 3 mm by 5 mm. Furthermore, for the texture we used a white canvas with a red dot in the middle. Figure 3.7 shows three phases of the algorithm, including the end result.

3.6 Discussion

As can be seen from Fig. 3.7, the curvature of the surface was preserved in the end result. Some of the smaller details, such as the folded over piece of tissue visible in the right side of the original image were lost. This can be explained by the information lost due to the somewhat drastic reduction in resolution as the first step of the algorithm. In the future,

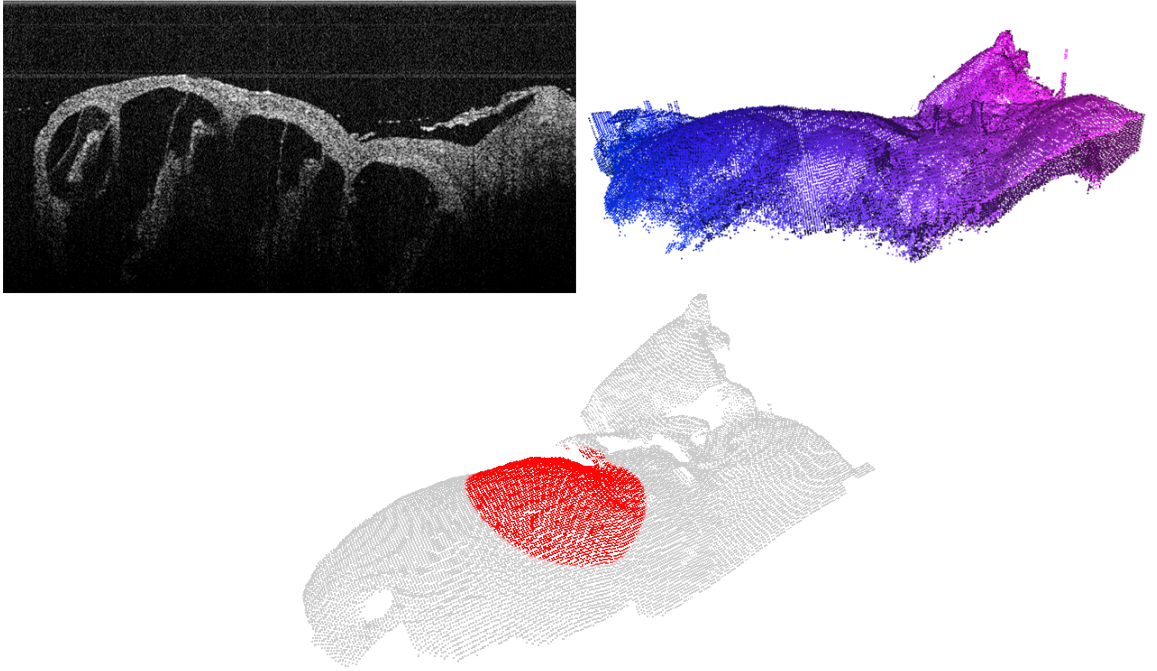


Fig. 3.7: Three phases of the algorithm. Top left: original slice of the OCT volume. Top right: Resulting point cloud before extracting the largest connected component. Bottom: end result, with the top layer extracted and the texture mapped to it.

this step should be replaced with another method that would preserve the details better, such as the superpixel method used later on.

Furthermore, currently the algorithm takes a few seconds to process the data. To achieve true real-time processing, steps to speed up the algorithm should be taken. Considering a constant stream of OCT data, frame-to-frame redundancies could be exploited to reduce the amount of processing for the segmentation.

Considering potential usability issues, the exact way to present the information to the surgeon needs to be considered. Since the conventional way to present the digital endoscopic view is a monitor next to the operating surgeon, methods to show the three-dimensional information using a two-dimensional projection (the monitor) in a way that does not negatively affect the usability of the device should be considered.

3.7 Conclusion

In this paper, we have presented a proof of concept for taking the data from a dual-mode en face camera-OCT endoscope and mapping the camera image to the surface topology information obtained from the OCT volume. This can be used in surgery to give the surgeon more information to work with when considering for example where to perform an

incision.

Chapter 4

Estimating the degree of EH using optical flow

Endolymphatic hydrops is a disorder in which an excessive amount of endolymph fluid causes an increase in the pressure of the endolymphatic system of the inner ear. In cochlea, endolymphatic hydrops can cause stretching in the *scala media*, a tubular construction that, along with two other similar structures, the *scala vestibuli* and *scala tympani*, runs through the cochlea from the base to the top.

Visualizing the enlargement of the *scala media* has traditionally been done by histological study of tissue sections, and the degree can be calculated by comparing the ratio of the area of the *scala media* to the combined total area of the three compartments. However, this requires the obtainment of tissue samples, and thus is not suited for *in vivo* measurements, and is a labor intensive process in general.

Imaging modalities that have been used to measure endolymphatic hydrops include MRI, μ MRI and μ CT. Recently, optical coherence tomography (OCT) has been applied for visualizing cochlea of mice *in vivo*. However, just visually discriminating between healthy and afflicted cochlea, and judging the degree of affliction can be difficult. Thus, we aim to establish a new technique for estimating the degree of enlargement of the *scala media*.

We measured two healthy and two diseased cochlea using conventional area calculation, as well as our method, and present results that show that our technique correlates with established standards for measuring the degree of endolymphatic hydrops in cochlea ($R = 0.977$, $p = 0.5 \times 10^{-7}$).

4.1 Introduction

Ménière's disease is a disease of the inner ear, with symptoms such as hearing loss, vertigo, tinnitus and a sensation of fullness in the ears. Ménière's disease is believed to be caused by *endolymphatic hydrops*, which refers to the accumulation of excess endolymphatic fluid in the inner ear. Genetic knock out of the *Slc26a4* gene is also known to cause endolymphatic hydrops in the inner ear [52].

Traditionally, the diagnosis of Ménière's disease has been through the clinical route, based on observations by a medical professional of the patient's clinical history and symptoms [53]. From histological sections, calculating the degree of endolymphatic hydrops (d_{hyd}) can be done by comparing the size of the *scala media* (A_{med}) to the combined sizes of the three vestibules: the aforementioned *scala media*, the *scala vestibuli* (A_{ves}), and the *scala tympani* (A_{tym}), i.e.

$$d_{\text{hyd}} = \frac{A_{\text{med}}}{A_{\text{ves}} + A_{\text{med}} + A_{\text{tym}}}$$

Recently however, methods for directly imaging the cochlea *in vivo* have emerged. Magnetic resonance imaging (MRI) has been used to image cochlear structure [54]. However, the internal structures including the Reissner's and the basilar membranes have been difficult to visualize with MRI because of their small size and thinness. The use of local contrasting agent has started to enable this visualization [55]. Attempts to visualize cochlear structure *in vivo* using a technique known as optical coherence tomography (OCT) have been made with success [26, 27].

OCT is an imaging technique based on near-infrared light that utilizes low-coherence interferometry and the soft-tissue penetrating properties of near-infrared light to measure the reflectivity profile of different points inside tissue samples. Thus it can be used to rapidly obtain micrometer resolution volumetric images of samples. Obtaining these images using OCT, the angle between the Reissner's membrane and the outer wall of the cochlea (the spiral ligament) can be calculated. This study was conducted to find out if the angle between these two planes could be used as a measurement for the existence of endolymphatic hydrops.

To achieve this, we took cochlea samples from mice with and without the *Slc26a4* gene, obtained images from them using an OCT machine, and did histology study on the cochlea to obtain a ground truth for the severity of endolymphatic hydrops. After that we compared the results to find out if a correlation exists.

4.2 Related work

Endolymphatic hydrops in Ménière’s disease was first discovered independently and simultaneously both by Hallpike & Cairns [56], and Yamakawa [57]. It has since been imaged by μ MRI in [58] as well as MRI in [54, 55, 59]. In [54], the focus was on visualizing fluid progression in the inner ear by injecting gadolinium hydrate diluted with saline in the perilymphatic space. Recently, cochlea have been imaged by OCT devices in [26, 27]. In the former the authors were interested in imaging the microvascular perfusion in mice using OCT, and in the latter visualizing the effects of endolymphatic hydrops in cochlea using OCT. The authors in [59] also proposed a scaling method for assessing the severity of endolymphatic hydrops. Grading was based on the distortion of the Reissner’s membrane, and the relative sizes of the *scala media* and the *scala vestibuli*, with *none* hydrops being observed when the size of the *scala media* is smaller and there is no disruption of the Reissner’s membrane, *mild* hydrops being observed when the Reissner’s membrane is disrupted but the *scala media* is still smaller, and *severe* hydrops when both the membrane is disrupted and the *scala media* is bigger.

Kakigi et al. [60] used isolated and decalcified guinea pig cochlea to obtain full thickness midmodiolar images of cochleae by OCT. With their image, it was possible to apply classical method of calculating the area ratio. However, it is not always possible to obtain such images. Due to the limited imaging depth, normally OCT images include only the lateral area without the modiolus, thus making it difficult to measure the areas of each scala, and to validate if these images are midmodiolar. With our method, only the lateral cochlear areas are required for the calculation and the images do not have to be midmodiolar.

Within the realm of otorhinolaryngology, imaging techniques for diagnosis are increasingly used. MRI and CT scans are common for detecting tumors in cases such as acoustic neuroma [61], dural sinus thrombosis [62] and many others. In ophthalmology, OCT has proven as a superior technique for retinal imaging, such as in detection of macular edema [63] and for detecting macular holes [20], among others. It is also often used for obtaining high resolution cross sectional and volumetric images in eye check ups. Many algorithms have also been developed for automatic retinal layer segmentation, which can help medical professionals in diagnosis, as the amount of data produced by a single volumetric OCT scan can be quite large. A good overview of the different algorithms used for segmentation and their clinical uses can be found in [21].

4.3 Theory

In this section we will go through the structure of the cochlea, as well as an explanation of endolymphatic hydrops. The structure of the cochlea varies depending on the species, and in this study cochlea from mice were used, so in some parts the explanation of the anatomy and the figures will focus on that. After the anatomical background, a short overview of the theory behind OCT is presented, as well as the fundamentals of optical flow.

4.3.1 Cochlea

The ear is the organ used for hearing and balance. It is split into three parts: the external (outer) ear, the middle ear and the inner ear. As sound waves travel inside the outer ear, they vibrate the thin tympanic membrane (eardrum), which transfers the sound waves into mechanical movement. As the eardrum vibrates, it moves the three ossicles, the malleus, incus and the stapes, which in turn transfer this movement to the inner ear through the oval window. The oval window then transfers the mechanical movement into the fluids inside the inner ear. The general structure of the ear is visualized in Fig. 4.1.

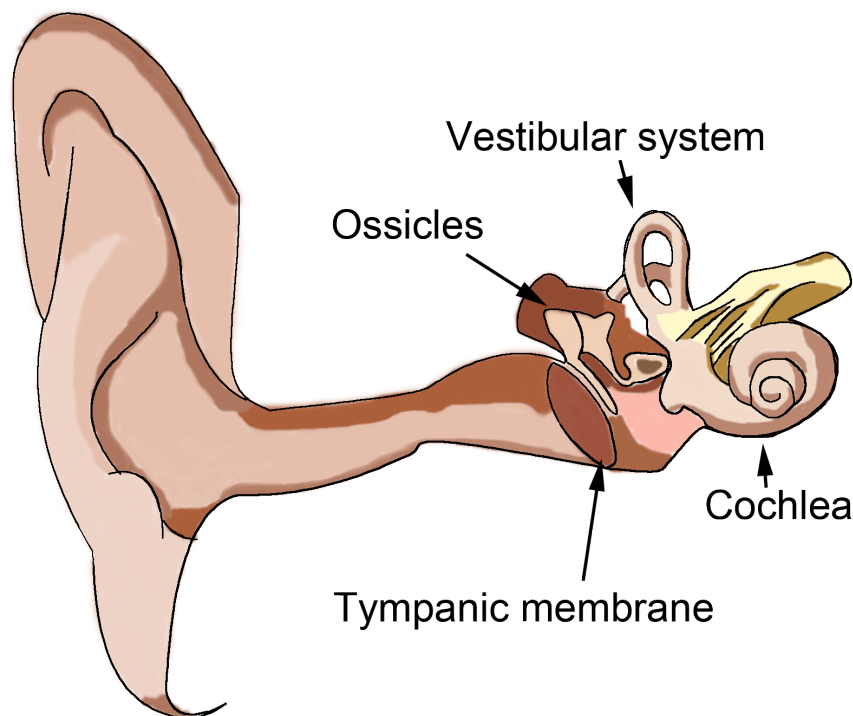


Fig. 4.1: General structure of the ear in humans.

The inner ear in mammals is housed inside the bony labyrinth and is composed of

two parts: the vestibular system, responsible for the sense of balance, and the cochlea, responsible for the sense of hearing. Although the shapes are quite different, both are mainly composed of endolymphatic and perilymphatic fluids, and the circulation of these fluids is essential in guaranteeing their functionality.

Structure

The cochlea is a spiral shell-shaped structure that provides the sense of hearing. In humans, it makes 2.5 turns around its mid-axis, called the modiolus. The amount of turns varies in mammals, for example in mice only making around 1.5 turns and in guinea pig making about 3.5 turns. The general structure of the cochlea is visualized in Fig. 4.3. From the figure it can be seen that inside the cochlea, there are three tunnels that run along all the way from the bottom to the top. These tunnels are called the *scala vestibuli*, *scala media* and *scala tympani*. Sound is transferred to the cochlea by the stapes, a small bone that is connected to the oval window. This transfers the mechanical movement of the ossicles to the fluid inside the cochlea. The oval window is connected to the *scala vestibuli*, and the vibration of the oval window moves the perilymphatic fluid inside the scala. This movement goes through the entire length of the cochlea, transferring at the very tip to the *scala tympani* through which it travels back down the length of the cochlea, vibrating the round window at the end. Depending on the frequency of the noise, at some point through its traversal of the cochlea, the movement is picked up by the inner hair cells of the *organ of Corti*, which convert the movement to electrical impulses that are then sent to the brain stem to be processed. Thus, the frequency range audible to the human ear is laid out along the length of this tunnel [64]. The next subsections will go through some of the important features of the cochlea and how they relate to relaying the sound from the cochlea to the brain. A labelling of these features on a microscope image obtained from the cochlea of a mouse is shown in Fig. 4.2.

The *scala vestibuli* (or vestibular duct) is the topmost of the three cavities inside the cochlea. It houses perilymphatic fluid. The purpose of the *scala vestibuli* is to transfer the sound vibrations entering the ear to the *scala media* through the Reissner's membrane, a thin membrane that separates the two. *Scala vestibuli* is connected to the *scala tympani* at the top of the cochlea, called the helicotrema.

Housing the endolymphatic fluid, the *scala media* (or cochlear duct) is the middle tunnel spanning the cochlea. Unlike its neighbors, the *scala media* is not connected to the other *scalae*. It is separated from the *scala vestibuli* by the Reissner's membrane, and from the *scala tympani* by the basilar membrane. It also houses the organ of Corti.

The *scala tympani* is the lowest of the cavities, connecting to the *scala vestibuli* at the

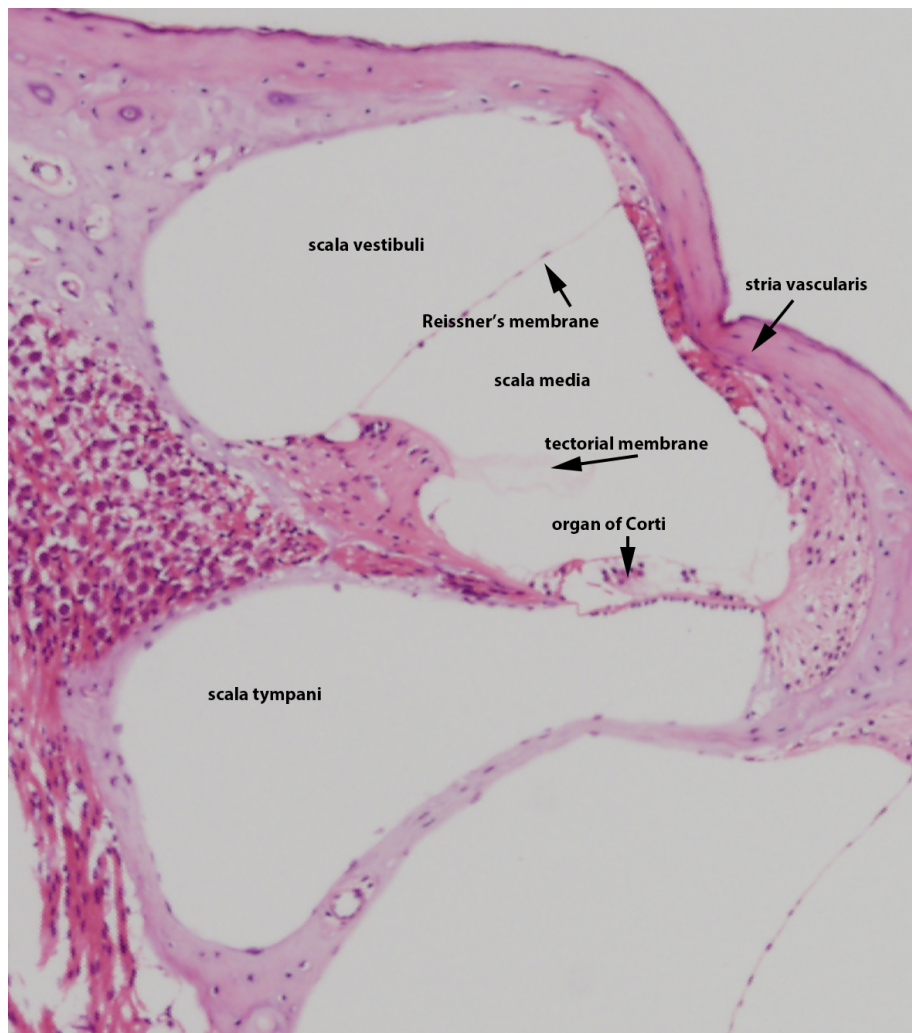


Fig. 4.2: A microscope image of a mouse cochlea, with all the relevant features labelled.

top, and to the round window at the bottom. Together with the scala vestibuli it allows the movement of the perilymphatic fluid inside the cochlea, which is conveyed to the scala media through the basilar membrane.

The basilar membrane is the part separating the scala media from the scala tympani. It is a structure that is narrow and stiff at its base and becomes wider and thinner towards the apical end. Due to this structure, when a stimulus is applied the maximum vibration of the membrane varies as a function of the frequency, that is, depending on the frequency of the sound entering the ear the basilar membrane vibrates strongly at differing positions. The strongest response to high frequencies are obtained at the basal end, and to lower frequencies at the apical end [65].

Located inside the scala media, the organ of Corti is responsible for converting the

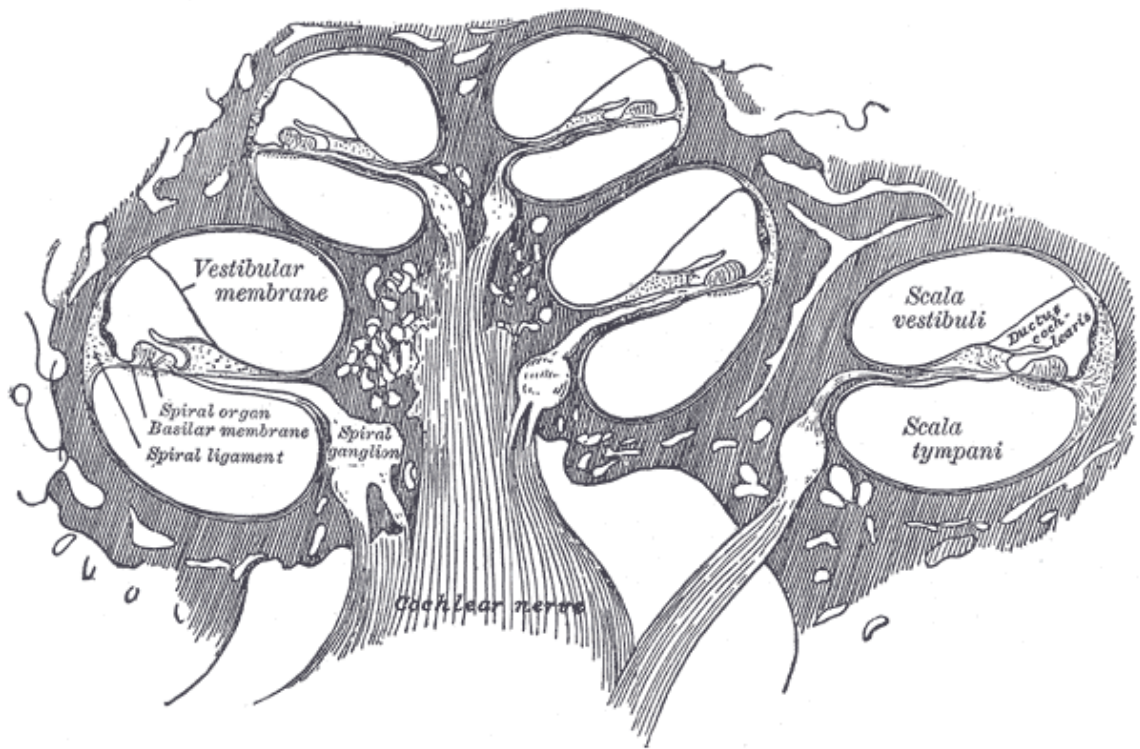


Fig. 4.3: A detailed cross-sectional view of the structure of the cochlea, showing the small features inside the tunnels. Note that the *scala media* here is labeled *ductus cochlearis*, and the Reissner's membrane is referred to as the vestibular membrane. Image obtained from a public domain version of Gray's *Anatomy of the Human Body* [1].

movement of the fluids inside the cochlea to the electrochemical signals that are then transferred to the brain through the auditory nerve. As the basilar membrane vibrates, small hair cells that are attached to both the organ of Corti and the tectorial membrane on top of it translate the vibration into shearing motion. This bends the stereocilia, organelles that protrude from the hair cells, causing changes in voltage, thus transducing the energy caused by motion into electrical signals sent to the brain [65].

The Reissner's membrane separates the scala vestibuli and scala media. Along with the separation of these two areas, it functions to allow the transfer of sodium ions from scala media to scala vestibuli [66].

Endolymphatic hydrops

Endolymphatic hydrops is the accumulation of endolymphatic fluid inside the inner ear. In the cochlea, it accumulates inside the *scala media*, causing an imbalance in the pressure between the scalae distorting the delicate features around it and impairing the function

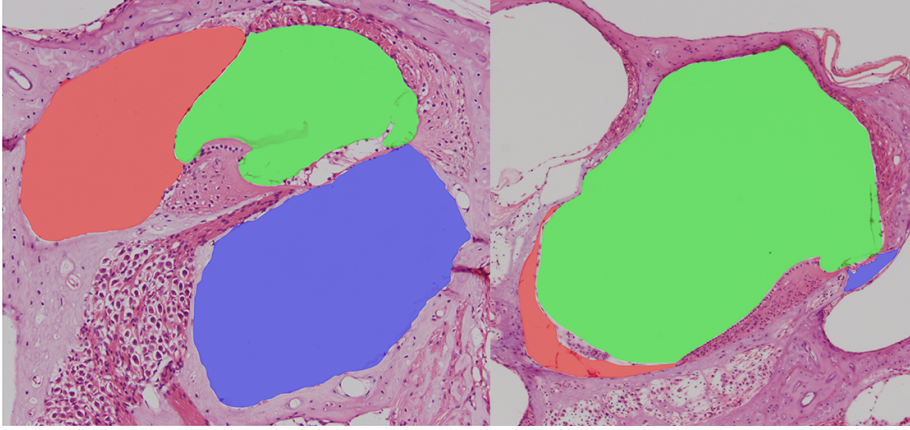


Fig. 4.4: Histological slices imaged by a microscope, obtained from a cochlea with no hydrops (left side) and with severe hydrops (right side). The red area represents the *scala vestibuli*, the green area the *scala media*, and the blue area the *scala tympani*. The colorings were added afterwards to highlight the features. Reissner's membrane is the membrane between the red and the green areas.

of the cochlea. This buildup can happen both in the cochlea and in the vestibular system. The cause for endolymphatic hydrops is unknown, but theories for the cause include constrictions in blood vessels, viral infections and genetic variations that cause abnormalities [67].

As can be seen in Fig. 4.4, the accumulation of the fluid stretches and pushes out the Reissner's membrane, a thin membrane that separates the *scala media* from the *scala vestibuli*. This distortion is the basis for measuring the degree, as the more enlarged the *scala media* becomes, the larger the ratio between its circumference and the total circumference of all the three *scalae* becomes. Nakashima et al. [59] proposed a three-staged system for grading endolymphatic hydrops in cochlea: *none*, if there is no displacement of the Reissner's membrane, *mild*, if there is displacement of the membrane, but the size of the *scala media* is less than or equal to the size of the *scala vestibuli*, and *severe*, if the size exceeds that of the *scala vestibuli*.

Ménière's disease (MD) is a disease of the inner ear. It manifests itself through four common symptoms: vertigo, experienced from minutes to hours at a time, tinnitus, hearing loss and a feeling of pressure inside the ear. It was first described by Prosper Ménière's in 1861 [68]. Later on, many investigations have been done to find the source of the disease, and it has become largely associated with endolymphatic hydrops [56, 57, 69, 70].

MD is typically diagnosed clinically through the symptoms present, and thus it can be difficult to diagnose, especially when not all of the symptoms are present. Due to this, its prevalence is also hard to measure. Many studies have been made on the incidence

and prevalence of Meniere’s disease, with prevalence numbers ranging from 21 / 100 000 persons in Japan [71] to 513 / 100 000 persons in Finland [72]. One of the most recent and biggest studies, performed by Harris and Alexander in 2010 and based on 60 million patients, puts the prevalence of MD to about 200 cases / 100 000 patients in the United States [73].

When the symptoms are sufficient to point towards a diagnosis of MD, clinical techniques such as an audiometric assessment or various nystagmographical tests can be performed to further validate the diagnosis. Other techniques that may provide to be useful include electrocochleography and vestibular evoked myogenic potential [53, 74].

Due to endolymphatic hydrops being the main histopathological correlate to MD [53], it can be surmised that the ability to visualize the condition of the inner ear to detect any abnormalities would be valuable in the diagnosis. Indeed, such attempts have been made and endolymphatic hydrops has been visualized by using techniques such as MRI [54], μ MRI [58] and optical coherence tomography (OCT) [26, 27].

4.3.2 OCT

OCT is an imaging method that utilizes the near-infrared light spectrum and a phenomenon known as interference to capture three-dimensional images from optical scattering media. OCT is analogous to ultrasound, in that the reflection of the light wave is used to measure the distance of the interface from which it reflected. The main difference is that with light the propagation speed is too high to measure the time it took for the light to bounce back. Instead, the phase difference of the light, and the fact that the coherency length of the light source is very short, are used to filter out other light that arrives at the detector. This incoherent, i.e. wide-spectrum light, allows the obtainment of micrometer-resolution images from samples. Traditionally, OCT machines have worked in the time domain, by scanning over a single column of the tissue by moving a reference mirror. This approach with its moving parts is relatively slow, however, and recently frequency domain OCT (FD-OCT) devices have been manufactured, which allow for real-time capturing of volumetric images with voxel counts in the tens of millions.

4.3.3 Optical flow

Optical flow, in a general sense, is the apparent movement of objects in a scene. In image processing, it is the movement of pixels between two sequential images. In this sense, it is often used to track some features over time or space. The general optical flow algorithm is based on two constraints: the brightness constancy constraint [75]:

$$E(x, y, t) = E(x + \delta x, y + \delta y, t + \delta t),$$

which essentially confirms that between two sequential samples of a scene (images), the brightness of a certain feature stays unchanged. Another form for the equation is as follows:

$$\nabla E \times \mathbf{v} + E_t = 0,$$

where ∇E is the spatial intensity gradient, and $\mathbf{v}$ the image velocity. This equation has two unknowns and thus cannot be solved as is. Different approaches for calculating optical flow thus introduce additional constraints to the equation. The method used in our approach is known as the Lucas-Kanade method [76], which calculates optical flow locally in a window, overdetermining the above equation, and using the least squares principle to calculate the best optical flow vector. Because it calculates the movement in a local neighbourhood, it is less sensitive to noise than some other optical flow calculation methods. In particular we use an iterative version of the method that utilizes image pyramids and affine transformations for more robust tracking [77]. It was selected over feature- and pure pixel based methods because of the relative robustness to noise (calculates movement in local patches) and large scale movements.

4.4 Method

Five months old laboratory mice were used for cochlea harvesting. In total, two mice had their cochlea harvested, one of which was morphologically normal (*Slc26a4*^(+/-)), and the remaining presenting endolymphatic hydrops (*Slc26a4*^(-/-)). Cochlea from both ears were used. After harvesting the cochlea, they were encased in a 4% paraformaldehyde solution overnight, after which they were immersed in 10% ethylenediaminetetraacetic acid (EDTA) for 7 days for decalcification. After that they were imaged using the OCT machine. Following, they were encased in paraffin solution for a week and histological slices were obtained. 7-10 slices per cochlea were obtained to ensure that the measuring would be a success. The slices were 4 μm thick with 20 μm spacing. Histological slices were stained using hematoxylin and eosin, and imaged using a microscope (Olympus BX50). From the images obtained, the areas of the particular compartments of the cochlea were calculated using ImageJ software [78] by hand-drawing borders around the three structures, and having the software calculate the areas. The experiment was approved by the animal experiment ethics committee at Kyoto University prior to conducting the experiments

(Approval number: MedKyo15155). An example that shows both the OCT image as well as the dyed histological slice from roughly the same area is shown in Fig. 4.5.

From the OCT volumetric image series, the angle facing the *scala media* between the Reissner's membrane and the spiral ligament was calculated using custom software. To obtain a more reliable result for the angle, as a preprocessing step the algorithm applies a bilateral filter to remove noise in the image. Bilateral filter was selected because it is good at reducing noise while maintaining the edges present in the image. Then, the software prompts the user to draw points on a slice corresponding to the two different structures. From these points, using an optical flow algorithm, the software calculates the movement of the points over several frames to both directions along the Z-axis (refer to Fig. 4.8). An example of the movement of the points and the tracking by optical flow is illustrated in Fig. 4.6.

This mitigates the effect a single slice has on the angle. As the angle can be hard to see, two rules for the point selection were made: 1) the connection between the Reissner's membrane and the spiral ligament has to be clearly visible, and 2) since the exact point of attachment can be hard to discern, the point of steepest angle around the area is used as the point between the two planes. The slice as close as possible to the center of the cochlea was used, to ensure that all of the structures would be visible, and that the angles would be as close to right as possible. Two planes are then fitted using all of the points, and the angle between the planes is used as the measurement of endolymphatic hydrops. An illustration of the user inputted points as well as the calculated lines can be seen in Fig. 4.7. From the slices, the ratio between the area of the *scala media* and the total area of the three vestibules was calculated. Using these two measures, the angle and the ratio were compared to find out if there exists a correlation between the two variables. Also, using the calculated ratio, the severity of hydrops was assigned by using the scale mentioned in section 4.2.

To summarise, the workflow is as follows:

1. Load the data.
 2. Navigate to the center slice of the data.
 3. Assuming the structures are clearly visible, mark the points on the two different structures.
 4. Calculate the positions on the n adjacent slices.
 5. Fit two planes on these points using a least squares plane fitting algorithm.
-

The size & resolution of the OCT images obtained are illustrated in Fig. 4.8. The machine used was a Thorlabs OCS1300SS OCT system, with a central wavelength of 1300 nm, a spectral bandwidth of 100 nm, and an average output power of 10 mW. The axial scan rate of the machine is 16 kHz. Thus it can obtain the measured 1024 column by 256 slide volumes in about 16 seconds.

4.5 Results

The results of the measurements are shown in table 4.5, and further visualized in Fig. 4.9. The data is comprised of three sets of measurements made by different researchers, as well as the calculated averages and standard deviations. Two of the researchers mentioned were qualified medical professionals, and one was familiar with the proposed method. The grades are assigned according to the method explained in section 4.2. As can be seen, there is a clear difference between the ratios of the cochlear ducts, as well as the angles, between the healthy and the diseased cochlea. Furthermore, two clear groups can be seen in the scatter plot, indicating that there is indeed correlation between the angle and the ratio.

	Ratio (%)	Angle (deg)	Assigned grade
NO left	0.258	109.239	<i>none</i>
NO right	0.270	106.195	<i>none</i>
HY left	0.953	150.214	<i>severe</i>
HY right	0.932	148.891	<i>severe</i>
NO left	0.242	116.913	<i>none</i>
NO right	0.267	107.191	<i>none</i>
HY left	0.946	152.723	<i>severe</i>
HY right	0.923	146.321	<i>severe</i>
NO left	0.233	108.279	<i>none</i>
NO right	0.253	99.675	<i>none</i>
HY left	0.952	150.239	<i>severe</i>
HY right	0.938	142.022	<i>severe</i>
NO mean & st. dev.	0.254 ± 0.014	107.915 ± 5.554	
HY mean & st. dev.	0.941 ± 0.012	148.402 ± 3.758	

Using the Kruskal-Wallis test to check if the angles between the normal and the hydrops groups differs gives $\chi^2 = 8.3077$ with $p = 0.004$, indicating that the mean ranks between the two groups are different. This suggests that there is a significant difference with the angle between the cochlea with and without hydrops.

Calculating the Pearson correlation coefficient using the values provided gives $R = 0.977$ with $p = 0.5 \times 10^{-7}$, indicating that there exists a statistically significant correlation between the angle of the Reissner's membrane, and the degree of endolymphatic hydrops

as measured by the ratio of the different *scalae*. However, with the low amount of samples caution should be exercised when interpreting the numbers.

4.6 Discussion

This research shows that the angle between the Reissner's membrane and the outer cochlear wall can be used as a metric for measuring the existence of endolymphatic hydrops using OCT imaging. Traditionally, measuring it *in vivo* has been difficult. Measuring the degree of endolymphatic hydrops with MRI needs to be carried out with the aid of some shadowing agent, which carries the risk of nephrogenic systemic fibrosis [79] and relies heavily on the clinical expertise of the operating physician to determine the locations and sizes of the various structures inside cochlea. With OCT the structures are also more clear. However, completely imaging the cochlear shape with OCT is difficult because of the rapid drop off in intensity as deeper structures are imaged. Combining OCT with our technique, these measurements are made possible by replacing the traditionally used metric of the ratio with a metric more useful concerning the limitations of OCT. Together with previous research done on *in vivo* OCT imaging of the cochlea, this suggests the viability of *in vivo* measurement of endolymphatic hydrops. The fact that OCT does not penetrate hard tissue such as bone very well implies that completely non-invasive visualization cannot be done, and that surgical procedures are needed to gain access to the cochlea.

Despite the correlation present in the results, more data should be collected for more robust testing. Due to the relative simplicity in obtaining mice with and without severe endolymphatic hydrops, the two groups were fairly well separated, as can be seen from Fig. 4.9. Indeed, using the scale mentioned in section 4.2, the measurements would be classified either as *none* or as *severe*, depending if they were done on mice with the gene intact, or removed, respectively. Thus, imaging various degrees of hydrops would be fruitful for seeing in what way the data is distributed between these two extremes.

From the data obtained, and as can be seen from Fig. 4.5, it can be difficult to discern exactly where the Reissner's membrane actually connects to the spiral ligament. In our study, this was taken into account by starting the two different planes from the sharpest point in the inner curve, but cannot completely be alleviated as the problem raises from the limited resolution & noise floor of the images. Also, due to the curvature of the Reissner's membrane, the angle depends somewhat on how long of a stretch of the membrane one uses as an input to the algorithm. This is because the algorithm fits a plane to the points. This could be alleviated by trying to fit a parabolic surface instead.

Atrophy of the lateral wall also affects the angle measured, thus modifying the ratio

between the angle and the area. Properly determining the effect of lateral wall atrophy on the ratio will require more measurements. In microscope images of the cochlea, the surface of the Reissner's membrane is often not completely smooth, but exhibits irregularities and is sometimes completely separated from one side. It is likely that this is caused by the histological slicing, rather than being actually present in the cochlea before separation. The method presented using OCT and angle calculation is less disruptive to the cochlea, thus in effect giving a higher chance for the membrane to be of normal condition. However, the case where the membrane is disrupted but continuous is a limitation of our method, as the angle is difficult to calculate but area calculation should still be achievable.

More robust handling of noise could also be attempted, as optical flow is considerably sensitive to it. Either a noise model that takes better in to account the multiplicative nature of the noise in OCT images could be developed, such as the one described in [80], or a method for tracking the positions that is less sensitive than optical flow to noise could be implemented. The tradeoff for these potentially better performing algorithms likely is speed however. The relatively substantial amount of data that OCT produces makes it costly to do any kind of complicated processing.

Optical flow as a method for object tracking has been used widely and is regarded as a reliable method. Extensive evaluations on the accuracy of different optical flow algorithms (including the one we used) can be found in [81]. Validation of the accuracy of the optical flow method as pertains to the angle measurements was considered out of scope for mainly two reasons: 1) the accuracy was deemed sufficient by the doctors that used the application, and 2) accurately measuring the accuracy is in itself a problem with no clear solution, as calculating the movement by hand is very time consuming and will not necessarily produce the right answer.

Future work could also include automatically detecting the position of the Reissner's membrane. This would allow for completely automatic measurement of endolymphatic hydrops from OCT images.

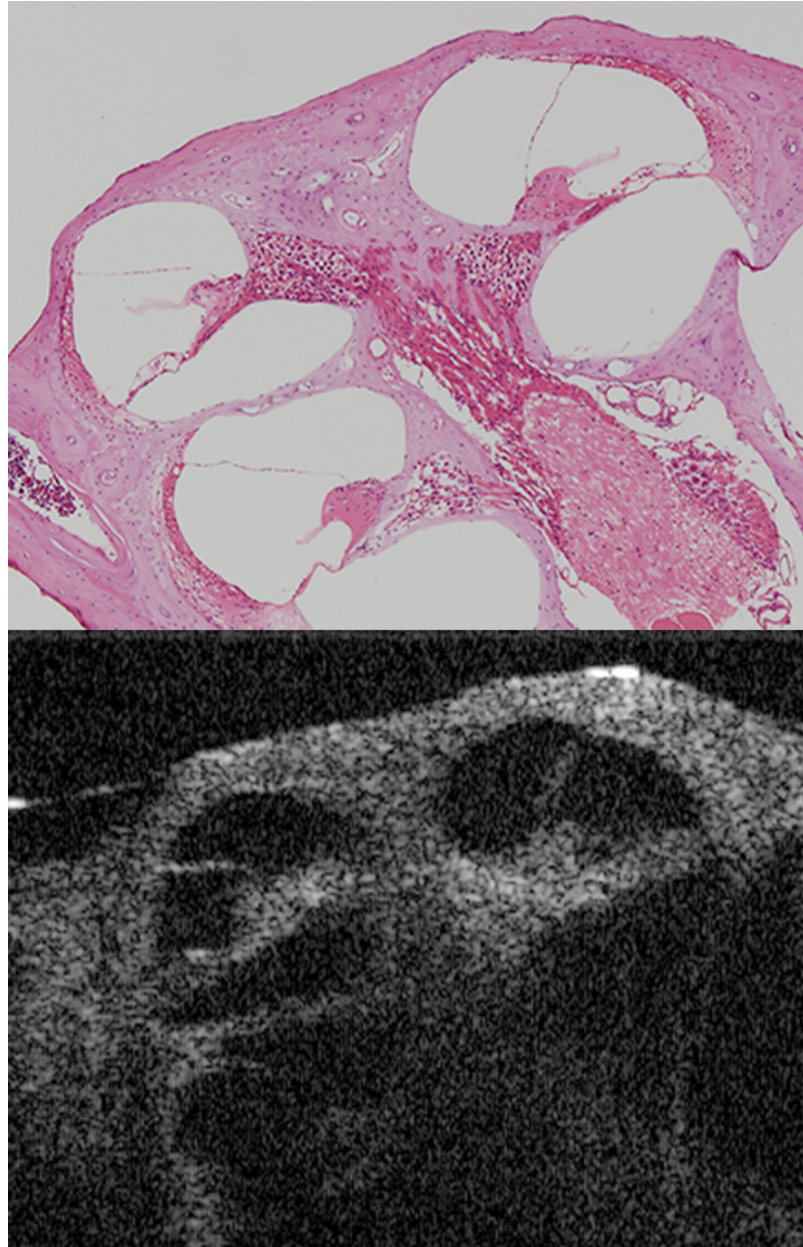


Fig. 4.5: Upper: Cochlear structures visible in a histological frame. Lower: an OCT slice of the cochlea, from roughly the same area as the above histological image.

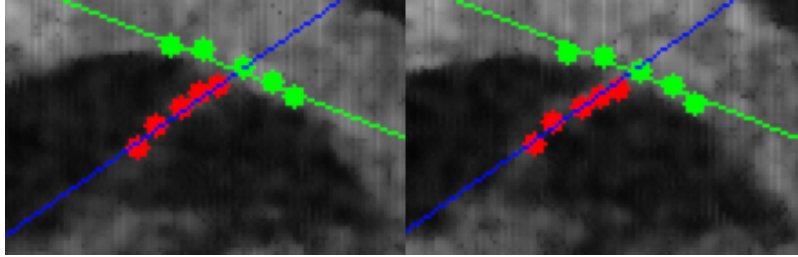


Fig. 4.6: Example of the apparent movement of the structures between two frames. Due to the oblique angle of the Reissner's membrane in comparison to the axes, it exhibits apparent movement when browsing from frame to frame. The optical flow algorithm keeps track of the position of the membrane.

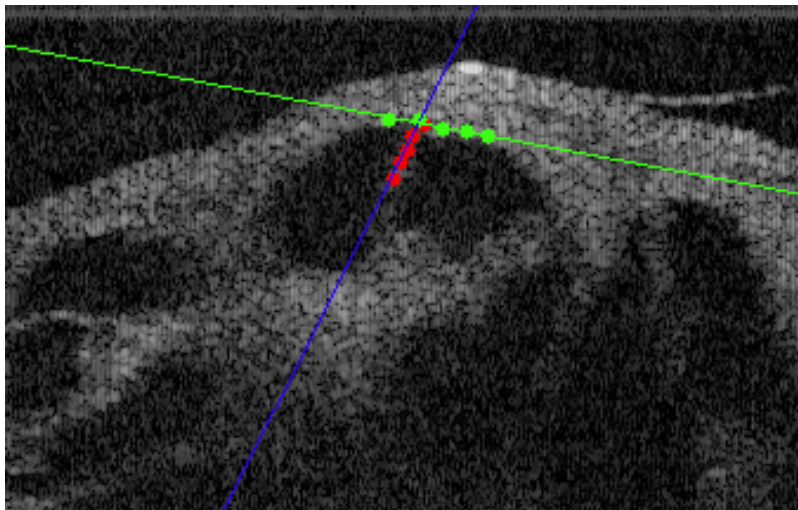


Fig. 4.7: An example of an OCT image with the user inputted points and the planes the software calculates

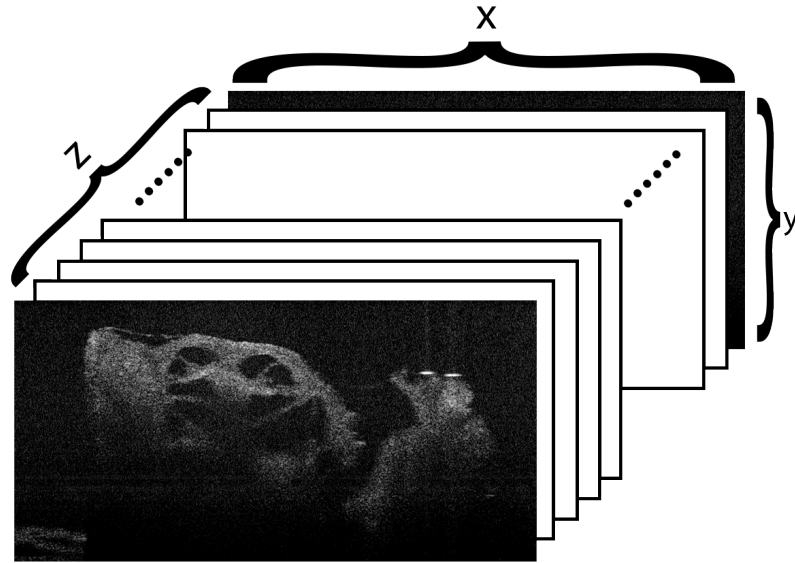


Fig. 4.8: The dimensions of the captured data. X is the amount of axial scans, which in our study was 1024 over 6 mm physical length. Y is the depth of the scan, and Z the amount of slices scanned. Both were set as 256 over 1.5 mm.

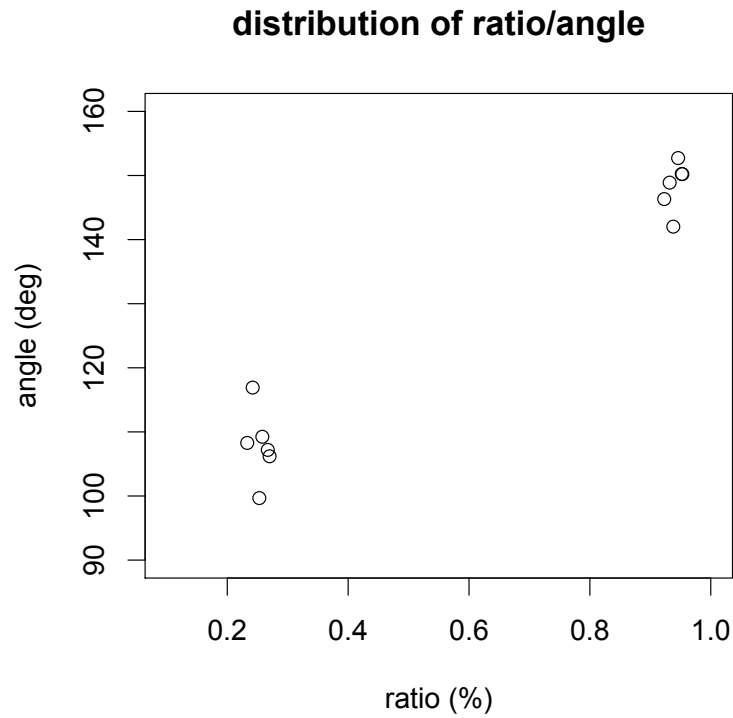


Fig. 4.9: Plotting the obtained angle & ratio values in a scatter plot. Two distinct groups can be seen, indicating correlation between the two different measures.

Chapter 5

Segmentation of OCT images using diffusion maps

5.1 Introduction

Meniere’s disease is a disorder of the inner ear that can manifest with the following symptoms: spinning, hearing loss, vertigo and a feeling of fullness in the ear. It is widely believed to be caused by endolymphatic hydrops, an accumulation of endolymphatic fluid in the inner ear. The accumulation increases pressure in the cochlea which causes a variety of symptoms including loss of hearing. Diagnosis has traditionally been done by looking at the symptoms exhibited [53], but other indirect methods such as vestibular evoked myogenic potential (VEMP) and electrocochleogram (ECoG) exist. Direct measurement of endolymphatic hydrops has been achieved by MRI, and a suitable grading system based on the relative sizes of the tunnels inside has been proposed. However, in traditional MRI the resolution places a limitation on the diagnostic capabilities, and especially small structures inside the cochlea such as the Reissner’s membrane are difficult to visualize. Recently, a high resolution imaging method called optical coherence tomography (OCT) has been used to image cochlea *in vivo*, and a novel technique for measuring the degree of endolymphatic hydrops based on the structures inside it has been proposed.

In this paper, we propose a method for extracting the 3D structure of the cochlea from volumetric grayscale OCT images, with the goal of facilitating the automatic measurement of endolymphatic hydrops using OCT.

5.2 Background and related work

Traditionally the only way to visualize endolymphatic hydrops has been through histological sectioning. Recently images of varying resolution have been obtained through MRI, CT and OCT creating a larger need for analysis of the phenomenon. This section will go through the physiological background behind endolymphatic hydrops, as well as a short description of OCT and related segmentation methods.

5.2.1 Endolymphatic hydrops

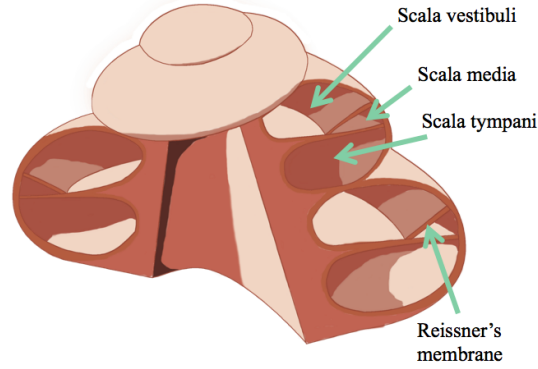
The cochlea is one half of the inner ear responsible for the sense of hearing. It comprises three tunnels: the *scala vestibuli*, the *scala media* and the *scala tympani*. Before sound reaches the cochlea, it is converted from pressure waves to mechanical motion in the middle ear. Then, as the mechanical motion is transferred to the inner ear it causes movement in the perilymph, a fluid housed inside the *scala vestibuli* and *scala tympani*. This also causes movement in the endolymph, housed by the *scala media*, which excites hair cells that then send electrical impulses to the brain, thus creating the sense of hearing. Endolymphatic hydrops is the excess accumulation of this endolymph, which causes interference in the cochlear function and the distortion of the Reissner's membrane, a membrane of about 12 μm in thickness [82] that separates the *scala media* from the *scala vestibuli*. The structure of cochlea and the effect of endolymphatic hydrops on it are visualized in Figs. 5.1.

5.2.2 Cochlear imaging

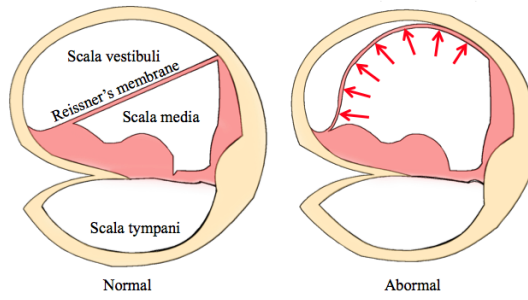
Endolymphatic hydrops in cochlea has been recently imaged by MRI [54] as well as OCT [60] [27]. A grading method for endolymphatic hydrops for MRI based on the area of compartments inside the cochlea was proposed by Nakashima et al. [59].

Segmentation of cochlea using computed tomography (CT) has been done semi-automatically using level sets [83]. An automatic approach using μCT for building models and then applying that model to conventional CT was explored by Noble et al. [84]

OCT is an interferometric technique that is analogous to ultrasound, with the exception that it uses near-infrared light for the measurements, and instead of using propagation time as measure for distance, interference between two light beams is used. One of the most popular applications of OCT is in retinal imaging, where it is used for detection of disorders such as macular degeneration. Segmentation of retinal OCT images has been done by various techniques including a graph-theoretic approach for layer segmentation [44], graph-cuts and dynamic programming [45] and coarse-grained diffusion maps [46]. In the latter, the authors used varying shapes for clusters of pixels in constructing the kernel, and men-



(a) The structure of the cochlea.



(b) View of cochlea displaying the effect of endolymphatic hydrops. Red arrows indicate the pressure that the accumulated endolymph applies on the membrane.

Fig. 5.1: The structure of the cochlea and effect of endolymphatic hydrops on it.

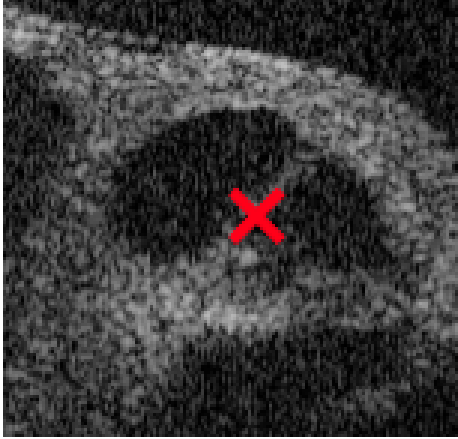
tioned the possibility of using Normalized Cuts for partitioning the image pre-application of diffusion maps. For OCT imaging of the cochlea, due to the limited penetration of the near-infrared light it is not feasible to expect that all of the three tunnels of cochlea can be visualized at the same time. Thus, a different metric for the degree is needed. One such metric could be based on the change in angle between the outer wall of the cochlea and the Reissner's membrane. This is visualized in Figs. 5.2. Previously, a semi-automatic method based on user input of the position of the two features on a 2D slice has been proposed [85].

5.2.3 Segmentation

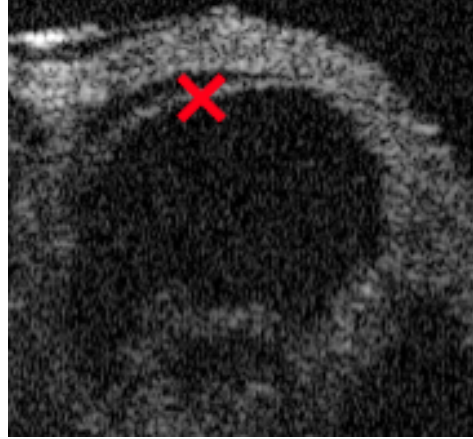
Superpixels are a group of methods for preprocessing high volume data, with the goal of grouping pixels with similar features together in meaningful groups. The motivation for doing this can be as a preprocessing step to reduce computational complexity, or for segmentation purposes. Various methods for superpixels have been proposed, including Normalized Cuts [86] and watershed [87]. Simple linear iterative clustering (SLIC) superpixels is a method based on k-means clustering that iteratively assigns data points to

the nearest superpixel based on distance metrics. It is a relatively recent method that performs well against other state-of-the-art superpixel methods [51]. The algorithm has one parameter, n , for the desired amount of superpixels. Using n and the dimensions of the provided data, it places initial superpixel centers at regular intervals. Then, using an iterative algorithm, it calculates the distance of pixels around the superpixel centers, and based on the distances, assigns each pixel to the closest superpixel and updates the superpixel center location. The execution stops when a number of iterations is reached.

Diffusion maps are a non-linear dimensionality reduction method based on the concept of diffusion. As a method it was first introduced by Coifman & Lafon [88]. Diffusion maps attempts to uncover the original lower-dimensional manifold that the data is sampled from, and thus can provide more accurate distance measures between data points. Diffusion maps is based on a kernel that defines connectivity in the data. By using the kernel a normalized graph Laplacian matrix is formed, and the eigendecomposition of the matrix corresponds to the coordinates of the data in diffusion space. The underlying idea is that by using few dimensions in the diffusion space we can closely approximate the original distribution of the data. It has been used for various purposes including edge-aware image editing [89], learning visual vocabulary [90], and retinal image segmentation [46].



(a) A healthy cochlea. Note the relative straightness of the Reissner's membrane, marked with a red X.



(b) An afflicted cochlea. The Reissner's membrane has been distorted.

Fig. 5.2: Change in the position and structure of the Reissner's membrane between a healthy and an afflicted cochlea.

5.3 Proposed system

The structure and flow of the algorithm is described in Fig. 5.3. The algorithm starts by reading the OCT data that the user has obtained from the OCT device. After reading the data, there are two major, distinct steps before the segmented binary image is created: the preprocessing step, and the clustering step. The aim of the preprocessing step is to reduce the amount of data we need to process by diffusion maps, and to obtain a sort of preliminary segmentation of the data. The processing starts with median filtering, where the images are smoothed a little bit to tend the superpixels towards more convex shapes. After smoothing the superpixels are created. This is done to speed up the algorithm while preserving the local structures inside the image. After superpixel creation, we form the diffusion map. For this, we use a Gaussian kernel with the Euclidean distances between superpixel intensities as the distance metric. Based on our attempts, using only the first eigenvector of the diffusion matrix seems to be enough to organize the data into the foreground and the background. Applying k-means clustering with 2 clusters on this eigenvector then gives us the superpixels that belong to the cochlea. Finally, we use the labels assigned to each pixel to re-expand the superpixels to have the original volume, thus obtaining the volumetric binary image with foreground corresponding to the cochlea and background to everything else.

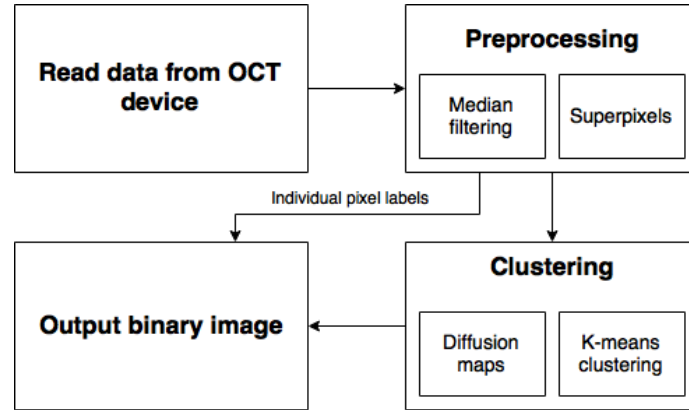


Fig. 5.3: The flow of the algorithm.

5.4 Results

In our study, the data is composed of 512 slices of 1000x512 pixel grayscale images, obtained by a Thorlabs OCS1300SS OCT system. The system has a reported axial resolution of 12 μm over air, and a lateral resolution of 25 μm . The imaging area was 5 mm by 3 mm by 5

mm. The sample cochlea used in all OCT images is from a guinea pig. In other studies, we have also successfully imaged and delineated the Reissner’s membrane from other guinea pig cochleae, as well as from mice. In Fig. 5.4 slices from three parts of the process are shown. The first slice shows the raw OCT 2-dimensional data that is obtained from the machine. The second is the corresponding slice from the superpixel image. The third one is the final segmented binary image. The arrows point to the Reissner’s membrane, and as can be seen, it has mostly been preserved in the segmented image. In Figs. 5.5 a 3D mesh generated by iso2mesh [91] in MATLAB and visualized by Meshlab software can be seen. The mesh is shown from 3 different sides, including the top view, bottom view and a view from the side. On the bottom view the Reissner’s membrane is clearly visible.

5.5 Discussion

As can be seen from the results, the method preserves the most important detail: the structure of the membrane in a way that the angle can be calculated. This suggests that the data can be used for calculation of the severity of endolymphatic hydrops from obtained OCT images. In future work, this automatic calculation of the angle from the produced data should be considered to obtain a full framework for grading endolymphatic hydrops. The current progress shows that creation of such a framework is feasible.

As a first step, the superpixel creation is important in order to preserve the fine structures present, but at the same time over-segmentation of e.g. the Reissner’s membrane can cause discontinuity due to the local variance in intensity levels. This discontinuity can be seen in Fig. 5.4c. To reduce this, a more advanced noise reduction method as a pre-processing step could be considered.

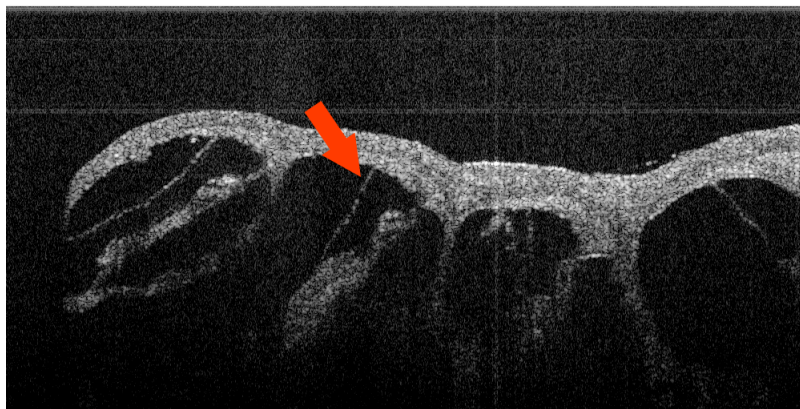
Better methods for visualization of the data could be applied. The volumetric image to mesh function provided by iso2mesh is somewhat clumsy for our particular application, and much smoother meshes could be achieved by customized processing of the data.

Although the focus is on measuring endolymphatic hydrops, the proposed method is in no way restricted to that, and can in fact be used for other data sets that exhibit similar characteristics e.g. noisy volumetric data with relatively fine structures.

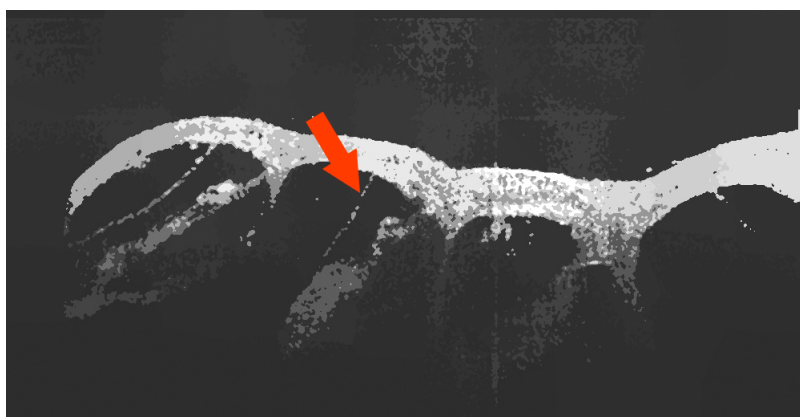
5.6 Conclusion

We demonstrated a method for segmentation of cochlear OCT images that preserves the minute structures present, such as the Reissner’s membrane. Future work includes fine-tuning the algorithm, and building on the presented method to allow for the use of au-

tomatic grading of endolymphatic hydrops. Furthermore, although the focus of the work is on cochlear OCT images for grading endolymphatic hydrops, other applications for the method should be considered.



(a) Original 2D slice showing the four turns of the cochlea.

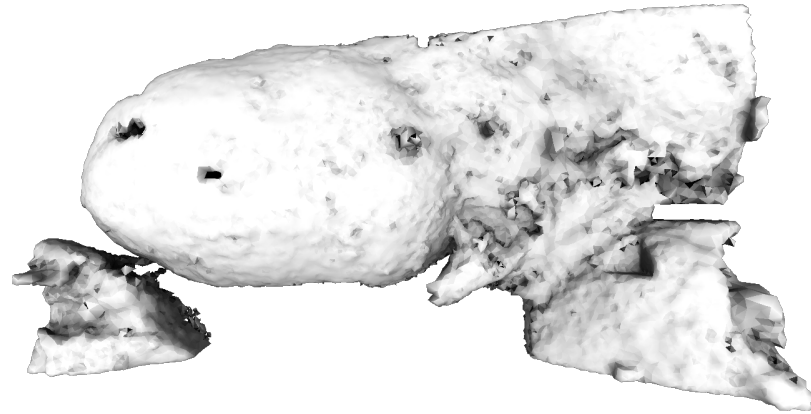


(b) The 2D section from the 3D superpixel structure corresponding to (a).

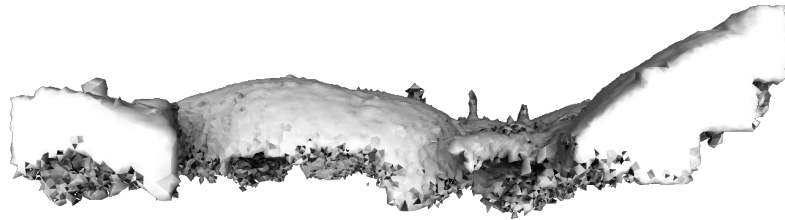


(c) The clustered binary image.

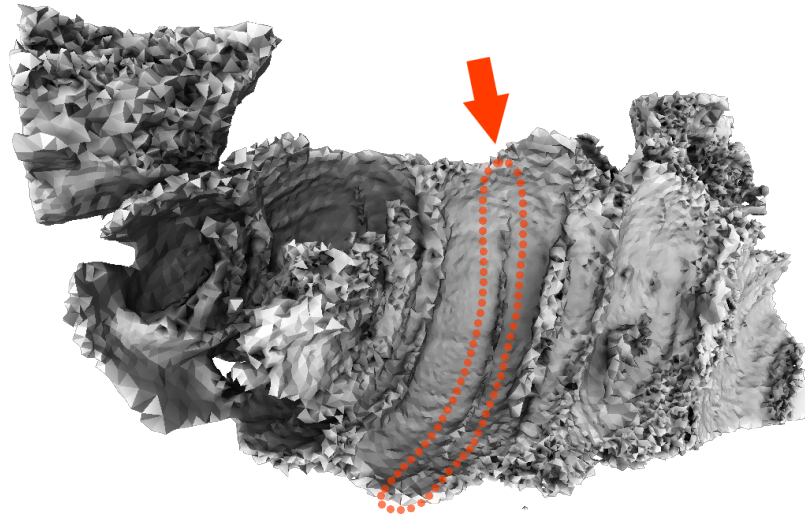
Fig. 5.4: Illustration of selected steps along the process. Red arrow indicates the position of the Reissner's membrane. Note that in (b), the superpixel intensities & structure are also affected by the neighboring slices.



(a) Mesh as seen from +Z axis.



(b) Mesh as seen from +X axis.



(c) Mesh as seen from -Z axis. Reissner's membrane is highlighted.

Fig. 5.5: A mesh created from the segmented image, visualized from different directions. Reissner's membrane is highlighted.

Chapter 6

Semi-Automatic Measurement of EH from Volumetric OCT Images

We present a novel semi-automatic method for measuring the degree of endolymphatic hydrops from volumetric optical coherence tomography images. Endolymphatic hydrops is a condition where excess endolymphatic fluid accumulates inside the inner ear, which can then cause diseases such as Menière’s disease. The exact morphological changes caused by the hydrops, and the relationship between endolymphatic hydrops and Menière’s disease are still unconfirmed, as high resolution in vivo imaging that preserves the minute structures inside the cochlea and the endolymphatic sac was not performed until recently. One possibility for exploring the relationship is through optical coherence tomography, which has been shown to have sufficient resolution and depth for obtaining partial images of cochleae from mice in vivo. We demonstrate a method for calculating the degree of endolymphatic hydrops from OCT images, and by comparing it and a previous method against ground truth values show that it produces superior results ($p = 0.029 < 0.05$).

6.1 Introduction

Menière’s disease is a disease of the inner ear that manifests itself with symptoms such as hearing loss, tinnitus, vertigo and a feeling of pressure in the ear. It has a prevalence rate of about 0.2% depending on the study cited [72, 73]. It is widely believed to be caused by endolymphatic hydrops, the accumulation of endolymphatic fluid inside the membranous labyrinth of the inner ear. The membranous labyrinth runs through the entirety of the bony labyrinth, the outer shell of the inner ear including both the vestibular system, responsible for the sense of balance, and the cochlea, responsible for the sense of hearing. In the cochlea, the endolymphatic fluid is housed inside the *scala media*, and as the fluid

accumulates, it distends the Reissner's membrane, a thin membrane that separates the fluid from the adjacent chamber, called the *scala vestibuli*. The morphological changes between a healthy and an afflicted cochlea are shown in Fig 6.1. In a healthy cochlea, the *scala media* fits between the two connected larger *scalae*, the aforementioned *scala vestibuli* and the *scala tympani* (Fig 6.1(A)). As endolymphatic hydrops aggravates, it causes swelling in the *scala media* and distends the Reissner's membrane, the thin membrane separating the *scala media* and *scala tympani* (Fig 6.1(B)). Thus, to assess the progression of endolymphatic hydrops, looking at the relative sizes of the *scalae* as well as the morphological changes related is the obvious approach.

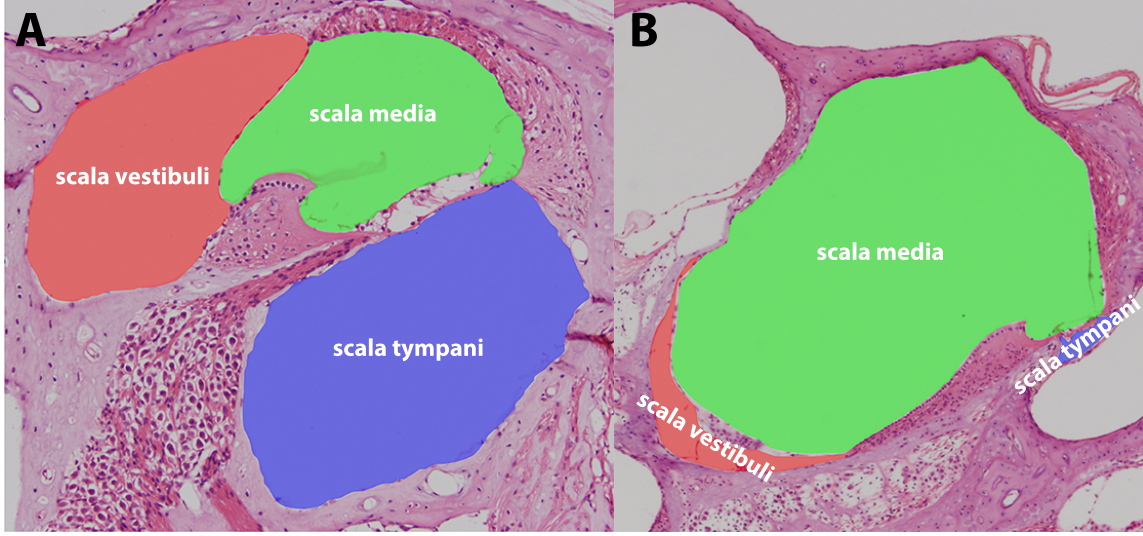


Fig. 6.1: Differences between a healthy and afflicted cochlea. Dyed histological slices from mouse cochleae imaged by a microscope. Red color corresponds to *scala vestibuli*, green to *scala media*, and blue to *scala tympani*. A: Sample from a healthy cochlea. B: Sample from an afflicted cochlea.

The exact relationship between endolymphatic hydrops and Menière's disease is unclear. For example, it is unclear to what degree does the hydrops need to progress for these symptoms to manifest. Even the morphological changes are hard to confirm due to the location and size of the inner ear. The traditional method for confirming the existence of hydrops is through histological examination. However, surgical removal of the inner ear has a chance of disrupting the fine structures inside such as the Reissner's membrane [92,93]. Thus, having the capability of quantifying endolymphatic hydrops *in vivo* could be useful when exploring the relationship between the physical manifestation and the disease, not to mention the eventual clinical value.

In this paper, we present a novel method for measuring the degree of endolymphatic hydrops from optical coherence tomography (OCT) images. We validate the method by

applying it, and a previously presented method [85] to a set of OCT volumetric images of cochleae, and by obtaining the calculated hydrops scores. By comparing these scores to those obtained from ground truth (GT) values set by an expert, we show that the proposed method outperforms the previous method by presenting the user with approximated features closer to the GT features. The method also requires less user interaction, potentially saving time spent on doing the diagnosis.

6.2 Related work

The golden standard for measuring endolymphatic hydrops is from histological sections. Various methods for quantifying hydrops from these sections include comparing the size of the *scala media* to an idealized size based on a virtual straight Reissner’s membrane [94–96]. Observing hydrops from the slices requires careful consideration of the slicing angle in order to not cause distortion in the features, so volumetric methods such as MRI or OCT are preferable to capture the three dimensional features and allow for freedom in viewpoint selection.

In vitro high resolution imaging of the cochlea has been done before by magnetic resonance microscopy [58]. The fixation required by the procedure, much like in the case of histological sectioning, can cause changes in the sample structure, and thus in vivo imaging is desirable. Previously, in vivo imaging has been achieved using magnetic resonance imaging (MRI) [59] and optical coherence tomography (OCT) [27]. Compared to MRI, OCT can do more accurate visualization of the smaller structures of the cochlea, but suffers from a high amount of noise, and is limited in the depth to which it can accurately capture the shape of the target tissue.

Multiple approaches have been proposed to grading the degree of hydrops, including a method that assigns one of three grades: significant, mild, or no hydrops [59] based on the relative sizes of the two canals, *scala vestibuli* and *scala media*, and the status of the Reissner’s membrane: if the Reissner’s membrane is undisrupted, there is no hydrops. Alternatively, if the membrane is disrupted but the *scala media* is smaller than the *scala vestibuli*, mild hydrops is present. Finally, if both membrane is disrupted and *scala media* is larger than *scala vestibuli*, significant hydrops is present. Another grading system was proposed by Gürkov et al. [97], based on a 4-point scale where the severity was decided by MRI images reviewed by medical professionals. Finally, Kakigi et al. [60] proposed an evaluation method based on OCT images, where the degree was calculated based on the ratio of the area immediate to the Reissner’s membrane to the area distal to the membrane, where the areas were separated by drawing a straight line segment between the attachment

points of the membrane and the wall.

Due to the limited depth imaging capability of OCT especially in the in vivo case with bone tissue present, it is preferable to calculate the degree using more local features, close to the surface of the cochlea. One method for doing this is to calculate the angle between the Reissner's membrane, a thin membrane separating the scala vestibuli and scala media, and the outer wall of the cochlea the membrane is connected to [85]. As the Reissner's membrane distends, the angle between the membrane and the wall changes, and thus it can be used to track the progression and/or classify the severity of hydrops. Locally speaking, this angle can be approximated by fitting two planes corresponding to the two features. A further advantage of this method is that it takes into account the aforementioned three-dimensional aspect of the features, allowing the cochlea to be imaged from angles not conforming to the main axis of the cochlea.

6.3 Methods

The proposed method can be broken to roughly three sections: preprocessing the data, segmentation of the data into background and foreground, and the calculation of the angle between the two features based on this segmentation. The background segmentation is based on preprocessing and clustering steps, and the angle calculation is based on first extracting the boundaries of the scalae and then fitting planes estimating the features to the boundaries. The individual steps taken to form the calculated planes from the raw OCT data are visualized in Fig 6.2.

OCT uses low-coherence interferometry to provide scans of the target either in one, two or three spatial dimensions. Thus it provides a strong signal in areas with a high density of scattering particles. OCT exhibits a type of noise called speckle noise which results from the superposition of backscattered light waves from different scattering points in the target. Speckle noise is characterized by high fluctuations in intensity, and thus can be hard to suppress in OCT images. On the other hand, speckle can also carry information about the imaged substance, so in some cases it can be a useful feature [34]. In our approach of pixel-intensity based segmentation, suppressing the effect of speckle is desirable.

The algorithm takes as input the OCT data, and from the user, the locations of the two *scalae* bordering the Reissner's membrane. Compared to the previous method, the proposed method requires less user input (2 points corresponding to the *scala* locations, compared to 10 points in total corresponding to the two features in the previous method). Furthermore, since the features are not marked directly, the user has less influence on the initial estimated shape of the features, leading to potentially less user-influence on the

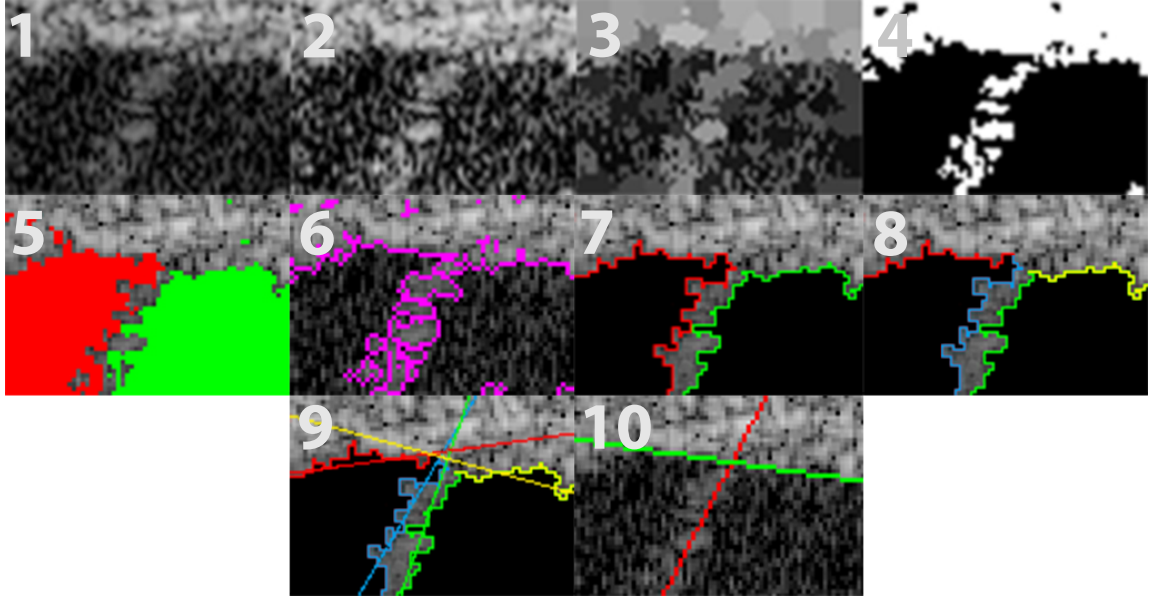


Fig. 6.2: Structure of the algorithm. Various steps are visualized from reading the OCT images to the final, calculated feature angles. Note that although the shown images are 2-dimensional slices, with the exception of median filtering and the illumination adjustment, all of the processing is done in three dimensions. 1: OCT image after median filtering. 2: Brightness constancy adjustment. 3: Superpixel segmentation. 4: K-means segmentation in diffusion space. 5: Segmenting the two scalae with s/t-cut. 6: Edge pixel labeling. 7: Largest component extraction (with scala vestibuli in red, scala media in green). 8: Wall/membrane segmentation with different colors corresponding to different segments. 9: Fitting planes to individual segments, with plane color corresponding to respective segments. 10: Average planes from step 9, with green corresponding to wall, red to segment.

interpretations of the feature locations. The data is a volumetric OCT scan S where for each pixel coordinate $(x, y, z) \in S$, an intensity value is reported. Since we are interested in the local structure, we limit our region of interest (ROI) in the Z-dimension to N slices connected to the slice the starting values are input on. On the angle calculation phase where we assign labels to individual pixels, we limit processing to the immediate vicinity around the initial values. In the evaluation process the size of this region of interest is based on the ground truth data. In future when doing measurements, it should be dependent on the feature sizes, which should for a given target (e.g. mouse, human) be roughly invariant.

The cochleae were imaged with a Thorlabs OCS1300SS OCT system, with a central wavelength of 1300 nm, a spectral bandwidth of 100 nm, an average output power of 10 mW, and an axial scan rate of 16 kHz. Five months old laboratory mice were used for cochlea harvesting. Three of the cochleae were harvested from wild type ($Slc26a4(+/-)$), and the remaining three from knockout mice ($Slc26a4(-/-)$). The study was approved by the animal experiment ethics committee at Kyoto University before conducting the experiment

(Approval number: MedKyo15155).

6.3.1 Preprocessing and segmentation

For preprocessing and segmenting the data into foreground and background, we follow a previously published method using superpixels and diffusion maps [98]. First, to reduce high frequency noise, we run the data through a median filter with window size $S_W = \{3, 3\}$ (Fig 6.2(1)). Median filter is used due to its ability to suppress the speckle noise inherent to OCT images while preserving the edges, which is needed for the segmentation step. To alleviate the signal degradation that OCT scans exhibit based on the depth of the imaging, we apply an illumination consistency algorithm. Illumination consistency aims to balance the brightness of pixels in an image despite varying amounts of light cast on different parts of the image. We use the SLRMSR [99] algorithm, which calculates the lightness L of a pixel i in an image based on

$$L(i) = \frac{1}{N} \sum_{k=1}^N \frac{I(i)}{I(x_{H_k})},$$

where N is the total amount of pixels in the image, I the original intensity of the pixels and x_{H_k} the index to the pixel with the highest intensity in a certain spray k . The sprays are randomly picked from a distribution centered on the pixel i . Effectively, the algorithm increases the intensity of pixels that are locally bright but globally dim, like in the example in Fig 6.2(2).

After preprocessing the data, to segment it into foreground and background, we use the SLIC superpixel [51] method for supervoxelization (Fig 6.2(3)). SLIC uses a k -means clustering method that groups semantically similar pixels together in local clusters. SLIC was chosen due to its performance characteristics especially for grouping volumetric data. The algorithm requires as parameters the supervoxel size, which is set as $S = \{3, 3, 3\}$, and the compactness factor, which is set as $C = 15$. Next, to split the superpixels into background and foreground, we calculate the diffusion map [88], which is a non-linear dimensionality reduction method based on the concept of diffusion. The method works on the supposition that the pixels in an image are sampled from a non-linear manifold, and attempts to uncover that non-linearity. The goal is to more accurately be able to perform the final foreground–background segmentation, as the distances in the diffusion space are strongly correlated with the distances in the original sampled object. Diffusion maps are based on a kernel, which captures the local connectivity in some meaningful way. We use

the Gaussian kernel, where the weights $w_{i,j}$ between pixels were calculated as

$$w_{i,j} = \exp\left(-\frac{d_{int}^2}{2h_{int}^2}\right),$$

where d_{int} is the L_2 -distance between the intensities of the superpixels. Although speckle noise is not usually considered normally distributed, after the abovementioned preprocessing steps we consider the noise effectively suppressed, and the normal distribution to be an appropriate model for the intensity variation. Finally, based on the calculated diffusion map, we cluster the data in the resulting diffusion space into two clusters using k -means clustering, thus obtaining the segmented data. For clustering we used the 6 eigenvectors with largest eigenvalues, excluding the top-most constant eigenvector, and weighting with the respective eigenvalues. The final segmentation is shown in Fig 6.2(4).

6.3.2 Angle calculation

Once we have segmented the cochlear shape from the background, we use the user indicated *scala vestibuli* and *scala media* locations to estimate the shape of the two features: the Reissner's membrane, and the inner wall of the cochlea. We use a max-flow/min-cut approach [100] to separate the *scala media* pixels from the *scala vestibuli* pixels. The aim of a max-flow/min-cut (or s/t cut) operation is to find a cut of a graph that separates the graph into two subgraphs S and T, where the source s belongs to S, and sink t to T. The cost of the cut is the sum of the weights assigned to the boundary edges that are cut. The objective of a min-cut algorithm is to find a cut that minimizes this cost. In our implementation, we use the user inputs as the source s and sink t, and all pixels in the segmented images belonging to the same cluster as the source/sink are considered as nodes of the graph. For each node n , there are 3 kinds of connections: n_s , n_t and n_n for node-source, node-sink and node-node, respectively. For weights between the connections, we use the spatial distance combined with a function dependent on the non-node pixels crossed by a straight line between the pixels corresponding to the two nodes. The node-node weights were set only between (3-dimensionally) adjacent pixels on the images. For all weights $w_{i,j}$ between linked nodes i, j we used the formula

$$w_{i,j} = \exp\left(-\left(\frac{d_{spat}^2(i,j)}{2h_{spat}^2} + \frac{d_{sep}^2(i,j)}{2h_{sep}^2}\right)\right),$$

where $d_{spat}(i,j)$ is the spatial L_2 -distance between i and j , and $d_{sep}(i,j)$ a distance based on the number of pixels not belonging to the node segment crossed. This effectively makes the connection weaker between nodes that are either far away from each other, or have

foreground pixels (such as the membrane) between each other. For calculating the amount of foreground pixels between two points, we use Bresenham's line drawing algorithm [101] to draw a virtual line between i and j , and check each point on the line if it was assigned to the node pixel segment. For each non-assigned pixel, $d_{sep}(i, j)$ is increased by 1. h_{spat} and h_{sep} are set to $0.95 * \max(h_{spat})$ and $0.05 * \max(h_{sep})$, respectively. The values are obtained through empirical testing. Although the graph is directional, we set $w_{i,j} = w_{j,i}$ enforcing a symmetry constraint. After labelling all of the pixels (Fig 6.2(5)), we extract the edge pixels neighbouring either a non-assigned pixel or pixels belonging to the other graph subsection (Fig 6.2(6)). This gives us two main groups of pixels corresponding to the edges of the *scala vestibuli* and *scala media*, along with some noise that we remove by extracting the largest component using connected component labelling (Fig 6.2(7)).

Next, to split the extracted edge pixels to the two features, we use an implementation of the RANSAC algorithm [102] with a plane function to fit two planes to each pixel group. RANSAC takes a minimal amount of data points to create a model for a particular mathematical shape. It then checks how well the rest of the data conforms to the generated model, and by iterating the algorithm with different data points for input, picks the one with the least outliers. For both groups, when performing many iterations, the resulting two planes should correspond to the cochlear wall and the membrane. Thus, we fit a plane equation twice to both of the groups, resulting in 4 groupings of pixels: two for either side of the membrane, and two corresponding to the *scala vestibuli* and *scala media* side cochlear wall segments (Fig 6.2(8), all four parts highlighted in different color). To calculate the final parameters for the two planes corresponding to the two features, we do least square fitting of planes on the membrane segments and the wall segments separately (Fig 6.2(9)), and calculate averages for both, resulting in the final feature approximations (Fig 6.2(10)). The angle can then be calculated by comparing the normals of the planes.

6.4 Validation

We had a medical professional who was an expert in analyzing cochlear OCT data mark the locations of the membrane and wall on 15 consecutive slices for each dataset, starting from the slice that they believed they could most accurately recognize them. This was the ground truth data used for assessing the accuracy of the algorithms. Next we generated the predicted membrane locations on the same slices by using the previous, direct feature marking (DFM) and the proposed scala-location based (SLB) methods, and compared the angles obtained to those calculated from the ground truth values. Since both methods require user input as initial values, we used the ground truth locations on the first slice to

generate the input.

To evaluate the methods, we took the output from both DFM and SLB methods, and the ground truth feature locations, and fitted planes to them by using least squares estimation for the plane parameters. We then measured the distance between the normals of the planes produced by the two methods to those obtained from the GT values. The similarity metric used was cosine similarity. The reason for comparing the individual features instead of the angles was using the angles hides some of the dynamics between the features — if both differ from the ground truth by the same amount to the same direction, the error would not reflect on the angle. We used 6 datasets of cochlear OCT data, and ran the evaluation on each dataset five times. 3 of the datasets were from mice afflicted by endolymphatic hydrops (datasets 1, 3 and 5 in Fig 6.3), and 3 from healthy mice (datasets 2, 4 and 6 in Fig 6.3). For each dataset and method, the evaluations were ran 5 times, and in the results, the average of these is presented. Fig 6.3 shows the first slices' regions of interest for each dataset, the ground truth values, and the eventual features estimated.

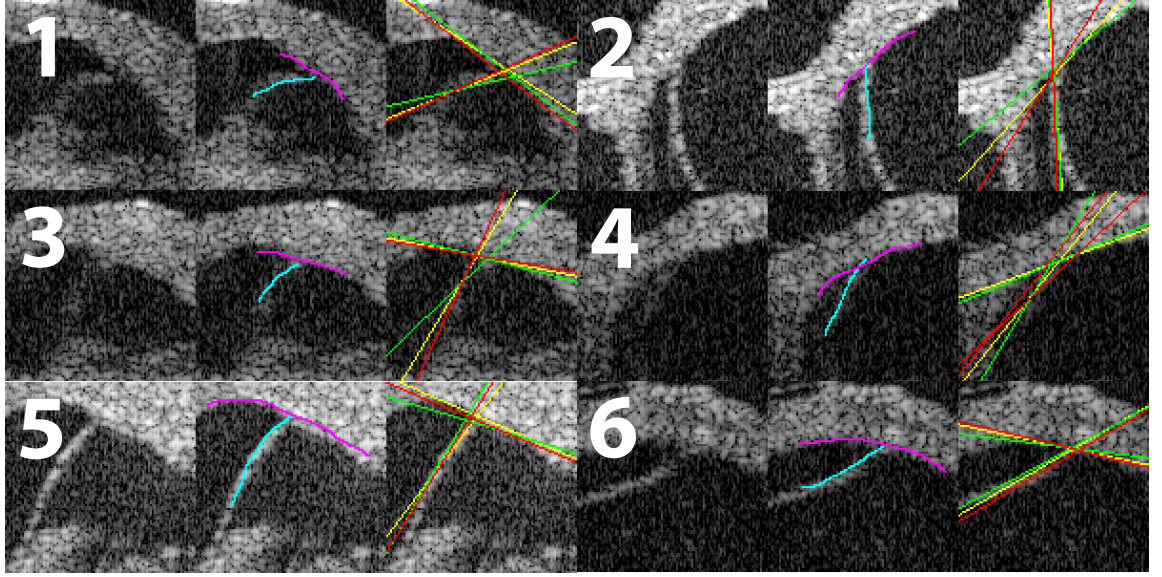


Fig. 6.3: Features, ground truth values, and estimated features for each dataset. For ground truth values, magenta corresponds to the cochlear wall and teal to the membrane. For estimated features, yellow marks the ground truth, green the DFM method, and red the SLB method.

Finally, since it can be assumed that for the OF algorithm, given a set of initial values reasonably close to the actual features most of the error will be caused by the algorithm and be distributed along the Z-axis, it is of particular interest when it comes to the performance of the algorithms.

As the two methods use different kinds of input for the initialization, a fair method for

generating the input based on the ground truth features is needed.

6.4.1 Initial value generation for DFM method

The performance of the previous method is largely dependant on the accuracy of the initial value input from the user. As the expected user input is 5 points for each feature marked in even intervals along the appropriate length of the feature (such that the local shape stays as close to linear as possible), we split the ground truth values along the features into 5 equally sized segments and sample randomly from each segment using a uniform distribution. First, we sample the inputs directly from the ground truth to get the "ideal" output from the method. Then, to simulate user input error, we use input values sampled from a radius r circle perimeter, where the center of the circle corresponds to a randomly sampled GT value.

6.4.2 Initial value generation for SLB method

For the proposed method, since just the general location of the scala vestibuli and scala media is enough, we form a pool of initial values from the triangles formed by the endpoints of the two features (see Fig 6.2). Then we sample from each pool a starting point for the algorithm by using a normal distribution with the parameters calculated from the coordinates of the points.

6.5 Results

The similarities between estimated features and GT values are shown in Table 6.1. We evaluated the DFM method with $r = [0, \dots, 6]$. As can be seen, the proposed method estimated the features generally better in all cases when compared with the direct sampling (with $r = 0$), except for the wall in dataset 6, and the membrane in dataset 4. Due to the non-normally distributed data, we did the Wilcoxon signed pair test to test for differences in the mean ranks of the distributions between the similarity scores for both methods ($N = 60$), and rejected the null hypothesis with $p = 1.449 \times 10^{-06} < 0.05$. The distributions are similar, so we can further claim that the median values' difference of 0.019 in favor of the SLB method is statistically significant, indicating that the proposed method performs better than the previous one. Both average similarities and variances for the datasets are visualized in Fig 6.4. Fig 6.5 shows the similarity score and variance averages as calculated over all datasets. For differences in performance between the healthy and afflicted datasets, a significant difference was found for the SLB method ($p = 0.0005 < 0.05$). For the

DFM method, no significant difference was found ($p = 0.7747$). Looking at the specific components of the normal vectors, the average differences from the GT axes for each dataset and overall average differences are reported in Table 6.2. Both X- and Y-coordinate differences were statistically insignificant ($p = 0.0523$ and $p = 0.6221$ respectively), but for the Z-coordinate, the difference was significant ($p = 0.0024 < 0.05$).

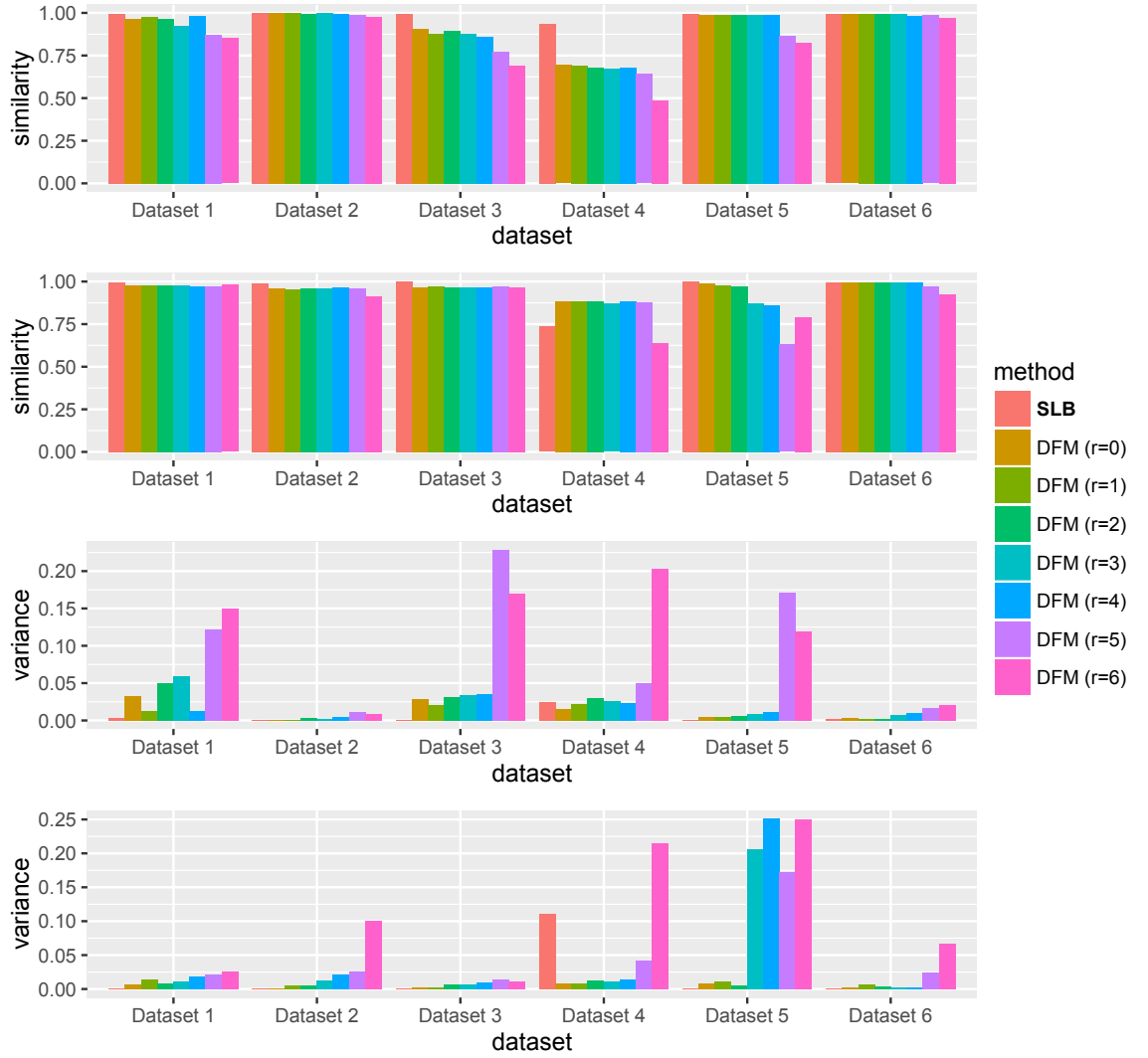


Fig. 6.4: Results for the evaluation. Top: calculated similarities for the wall feature over the datasets. Middle-top: calculated similarities for the membrane feature over the datasets. Middle-bottom: calculated variances for the wall feature over the datasets. Bottom: calculated variances for the membrane feature over the datasets.

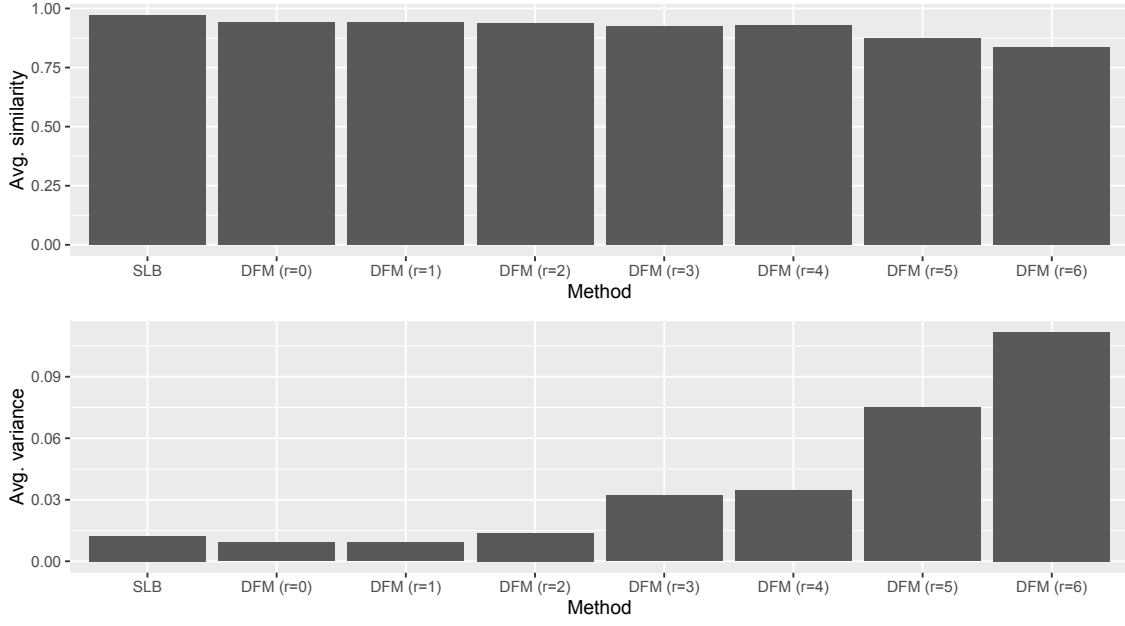


Fig. 6.5: Results averaged over datasets. Average similarities and variances for each method, averaged over the datasets.

Table 6.1: Calculated cosine similarity of the two methods to the GT plane normal vectors. DFM represents the previous method, and SLB the novel one. r equals the distance to the GT value at which the inputs were sampled from. Bolded numbers represent the closest average to the ground truth, and the smallest standard deviation for that particular dataset.

Method	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5	Dataset 6
Wall	Similarity to ground truth (average $\pm$ st. dev.)					
SLB	0.9944$\pm$0.0028	0.9985$\pm$0.0009	0.9960$\pm$0.0007	0.9378$\pm$0.0252	0.9936$\pm$0.0010	0.9913 $\pm$ 0.0021
DFM (r=0)	0.9661 $\pm$ 0.0327	0.9984 $\pm$ 0.0003	0.9046 $\pm$ 0.0283	0.6929 $\pm$ 0.0149	0.9871 $\pm$ 0.0049	0.9917 $\pm$ 0.0034
DFM (r=1)	0.9771 $\pm$ 0.0124	0.9983 $\pm$ 0.0007	0.8759 $\pm$ 0.0213	0.6898 $\pm$ 0.0226	0.9889 $\pm$ 0.0042	0.9937$\pm$0.0025
DFM (r=2)	0.9629 $\pm$ 0.0503	0.9955 $\pm$ 0.0031	0.8941 $\pm$ 0.0318	0.6808 $\pm$ 0.0298	0.9895 $\pm$ 0.0065	0.9937 $\pm$ 0.0023
DFM (r=3)	0.9252 $\pm$ 0.0595	0.9977 $\pm$ 0.0015	0.8770 $\pm$ 0.0335	0.6736 $\pm$ 0.0257	0.9874 $\pm$ 0.0081	0.9919 $\pm$ 0.0075
DFM (r=4)	0.9825 $\pm$ 0.0122	0.9935 $\pm$ 0.0047	0.8579 $\pm$ 0.0359	0.6810 $\pm$ 0.0238	0.9895 $\pm$ 0.0107	0.9844 $\pm$ 0.0104
DFM (r=5)	0.8709 $\pm$ 0.1220	0.9878 $\pm$ 0.0117	0.7713 $\pm$ 0.2279	0.6408 $\pm$ 0.0505	0.8672 $\pm$ 0.1714	0.9858 $\pm$ 0.0163
DFM (r=6)	0.8515 $\pm$ 0.1500	0.9755 $\pm$ 0.0093	0.6908 $\pm$ 0.1701	0.4850 $\pm$ 0.2026	0.8267 $\pm$ 0.1194	0.9692 $\pm$ 0.0203
Membrane	Similarity to ground truth (average $\pm$ st. dev.)					
SLB	0.9958$\pm$0.0012	0.9902$\pm$0.0010	0.9989$\pm$0.0004	0.7352 $\pm$ 0.1113	0.9989$\pm$0.0003	0.9970$\pm$0.0005
DFM (r=0)	0.9794 $\pm$ 0.0060	0.9585 $\pm$ 0.0013	0.9660 $\pm$ 0.0023	0.8833$\pm$0.0077	0.9882 $\pm$ 0.0074	0.9953 $\pm$ 0.0018
DFM (r=1)	0.9780 $\pm$ 0.0143	0.9544 $\pm$ 0.0050	0.9701 $\pm$ 0.0022	0.8832 $\pm$ 0.0075	0.9795 $\pm$ 0.0113	0.9945 $\pm$ 0.0061
DFM (r=2)	0.9754 $\pm$ 0.0076	0.9573 $\pm$ 0.0049	0.9682 $\pm$ 0.0071	0.8821 $\pm$ 0.0129	0.9712 $\pm$ 0.0048	0.9948 $\pm$ 0.0038
DFM (r=3)	0.9787 $\pm$ 0.0115	0.9612 $\pm$ 0.0122	0.9675 $\pm$ 0.0073	0.8695 $\pm$ 0.0106	0.8707 $\pm$ 0.2057	0.9955 $\pm$ 0.0016
DFM (r=4)	0.9741 $\pm$ 0.0182	0.9676 $\pm$ 0.0216	0.9649 $\pm$ 0.0094	0.8867 $\pm$ 0.0133	0.8591 $\pm$ 0.2509	0.9957 $\pm$ 0.0029
DFM (r=5)	0.9693 $\pm$ 0.0213	0.9608 $\pm$ 0.0260	0.9715 $\pm$ 0.0139	0.8801 $\pm$ 0.0421	0.6301 $\pm$ 0.1719	0.9722 $\pm$ 0.0238
DFM (r=6)	0.9805 $\pm$ 0.0252	0.9138 $\pm$ 0.1009	0.9644 $\pm$ 0.0116	0.6372 $\pm$ 0.2153	0.7912 $\pm$ 0.2494	0.9238 $\pm$ 0.0665

6.6 Discussion

Overall the proposed method performed better. The dataset with the worst performance was the fourth one, mostly due to the poor visibility of the membrane (see Fig 6.3(4)). Although a significant difference in performance between the healthy and afflicted data

Table 6.2: Normal vector component differences averaged over all datasets and both features, when compared to the GT normal vectors. The score of the method that achieved the smaller difference has been bolded.

Coordinate	SLB	DFM
Wall		
X	0.0443	0.1028
Y	0.0785	0.0792
Z	0.0890	0.2563
Membrane		
X	0.1264	0.0738
Y	0.0977	0.0580
Z	0.0713	0.2163
Combined		
X	0.0853	0.0883
Y	0.0881	0.0686
Z	0.0801	0.2363

was found for the proposed method, this is mostly because of dataset 4, as comparing the results with the other two datasets with endolymphatic hydrops present, the difference in performance is quite obvious.

Further, the supposition was that if the optical flow method used in the DFM method was as susceptible to noise as thought, most of the error should be concentrated on the Z-axis component of the normal vector. As can be seen in Table 6.2, this is indeed true. For the variance scores, although in most cases the SLB method provided the better score, due to the poor performance on dataset 4, overall the DFM (with $r = 0$) performed better.

Increasing the radius doesn't have such a drastic effect on the average quality of the results until towards the larger end of r . Since we are adding essentially zero-mean noise to the user input sampling, and because the DFM-method tends to "stick" to the original inputs, on average the only case where the performance suffers is when sampling directly from the GT values would accurately track over the Z-dimension, but sampling from distance r misses the membrane and thus cannot track it. Due to this, and the error in the DFM-method mostly being introduced in the Z-dimension to begin with, increasing r doesn't contribute much to the accumulated error. Meanwhile, the variances do increase compared to SLB, due to the deviance in sampling locations growing larger with increased r .

Related to the above, something that the measurements do not show is the inter-user variability in the results. As the DFM method is very dependent on the user's idea of where the membrane is located, the SLB method could possibly result in lower variability. To investigate this, in the future a user test should be carried out. Further validation

could be done by having test participants do measurements using both algorithms, and see how the actual results compare in that case. Furthermore, based on the amount of input required from the user (10 coordinates in the previous method, 2 in the proposed), it can be assumed that the proposed method requires less interaction. Nevertheless, actual measurements should be done to showcase this.

For the SLB method, we avoid erroneous input by only considering the pixels that belong to the *scala* (the "background" area with lower average intensity in the segmented data). Thus we do not have to deal with situations where the user has clicked on a foreground pixel. This can happen especially when the membrane has bulged out due to extreme hydrops, and the area of the *scala vestibuli* has shrunk. This can be dealt with using additional information of the surrounding pixels, but nevertheless can introduce uncertainty in the solution.

6.7 Conclusion

We have presented a novel method for estimating the degree of endolymphatic hydrops from segmented OCT images that achieves higher accuracy than a previously presented state of the art method. The results can be used for modeling the progression of endolymphatic hydrops, and together with symptomatic information, to investigate the deeper relationship between endolymphatic hydrops and Menière's disease. Future work includes working together with medical professionals to bring the method to actual use.

Chapter 7

Evaluation of a novel multi-articulated endoscope: proof of concept through a virtual simulation

In endoscopic surgery such as video-assisted thoracoscopic surgery and laparoscopic surgery, providing the surgeon a good view of the target is important. Rigid endoscope has for years been the go-to tool for this purpose, but it has certain limitations like the inability to work around obstacles. To improve on current tools, a novel multi-articulated endoscope (MAE) is currently under development. To investigate its feasibility and possible value, we performed a user test using virtual prototype of the MAE with the intent to show that it outperforms the conventional endoscope while bringing minimal additional burden to the operator. To evaluate the prototype, we built a virtual model of the MAE and a rigid oblique-viewing endoscope. Through a comparative user study we evaluate the ability of each device to visualize certain targets placed inside the virtual chest cavity by the angle between the visual axis of the scope and the normal of the plane of the target, while accounting for the usability of each endoscope by recording the time taken for each task. In addition, we collected a questionnaire from each participant to obtain feedback. The angles obtained using the MAE were smaller on average ($p = 0.01$, $MD = 7.25$), indicating that better visualization can be achieved through the proposed method. A non-significant difference in mean time taken for each task in favor of the rigid endoscope was also found ($p = 0.42$, $MD = -2.09$). We have demonstrated that better visualization for endoscopic surgery can be achieved through our novel MAE. The scope may bring about a paradigm

shift in the field of minimally invasive surgery by providing more freedom in viewpoint selection, enabling surgeons to perform more elaborate procedures in minimally invasive settings.

7.1 Introduction

Abdominal and chest surgery have traditionally relied on big surgical operations such as thoracotomy or laparotomy. However, through the use of devices such as the thoracoscope and the laparoscope, the surgeon has become able to access the target organ inside the chest or abdominal cavity only by creating a small incision. This approach results in less invasive treatment for patients who require surgical interventions, and is commonly called minimally invasive surgery [103]. Largely due to only requiring a small incision it has been shown to be the superior approach in comparison to conventional invasive methods such as the thoracotomy [104], with studies showing it leads to comparatively shorter hospital stays, better cosmetic outcomes, less pain, less complications and better prognosis overall [105] [106] [107].

Especially because of the small incision, in endoscopic surgery the method of visualization through the endoscope is important. The traditional tool for visualization is the rigid endoscope, a 300-400 mm long, 3-10 mm diameter tubular device that provides a view inside the human body. In addition to the rigid endoscope, semiflexible thorascopes [108] [109] containing one joint for twisting the distal portion are also available.

The problem with minimally invasive surgery, and endoscopy in general, is that only a two-dimensional view can be obtained, necessitating training for the surgeon to successfully be able to perform the elaborate procedures needed [110]. Recently, various stereoscopic endoscopes have been developed that aim to reconstruct the three-dimensional features [111] [112]. Even still the problem of presenting this information to the user remains. Furthermore, there is the problem of the difference between the viewing axis and the working axis: the surgeon often has to perform operations where the angle between the optical axis of the endoscope and the instrument's plane is large. This has been shown to be of detrimental effect to the performance of the surgeon [113] [114].

We introduce a virtual prototype of a novel endoscope composed of multiple joints that allow the endoscope body to change its shape based on user guidance. The principle is that as the user controls the leading segment, the other joints conform to the movement in a way that allows the rest of the endoscope to follow along on the same route. Previously, similar ideas have been explored using cable actuators [8] [7], as well as for laparoscopy using servo motors [115] and shape-memory alloys [116]. In contrast to building a complete surgical

system with a high degree of freedom (DOF) such as the i-Snake [117] or the Flex robotic system [118], we focused on creating and evaluating the usability of a simple endoscope that comprises of joints with one DOF that would allow for superior visualization of targets compared to traditional rigid endoscope, while staying low cost and easy to manipulate. The goal is to produce a view for the surgeon that reduces the disorienting effect caused by the lack of depth-information and non-optimal placement of the endoscopic view in relation to the other tools in use during surgery. We validate the feasibility of this approach and compare our novel MAE with a traditional rigid endoscope through a user study.

7.2 Methods

The concept of the MAE is to manipulate the orientation of the leading segment, and as the endoscope is inserted/retracted from the body, have the rest of the endoscope body follow along the same route the leading segment takes. Thus we can avoid certain obstacles, and most importantly decide on the direction of approach when visualizing a target (see Fig. 7.1). To evaluate the endoscope we created a virtual chest cavity to simulate video-assisted thoracoscopic surgery (VATS). The project was realized using Unity 5.4.1. Basic Unity assets and self-created textures were used for the endoscope and the targets. The chest cavity was built from a CT scan using LiveVolume & Amira softwares. By using a game controller, the user can manipulate the endoscope in an environment closely resembling the VATS setting. The reason for using a game controller as opposed to simulating manual operation of the endoscope was twofold: Firstly, using a game controller shouldn't bias the test participants either way. By this, we mean that since we're using the same control method for both the novel and the conventional endoscopes, it should not give an inherent advantage to either. Secondly, multiple robotic endoscope designs already use control methods similar to a game controller, thus lending credence to it [119] [120]. After deciding on the prototype specifications, we then conducted a user study where we evaluated the performance of the MAE and a conventional rigid endoscope through tasks where we asked the user to visualize different targets inside the thoracic cavity. We scored the performance of the user based on their ability to visualize the targets appropriately as well as the time taken for each task. The expectation was that using the prototype the targets could be visualized better, but because of the added functionality and thus more complicated controls, the elapsed time per each task would be longer. At the end of the experiment, we collected a questionnaire about the experience, to further evaluate and aid in development.

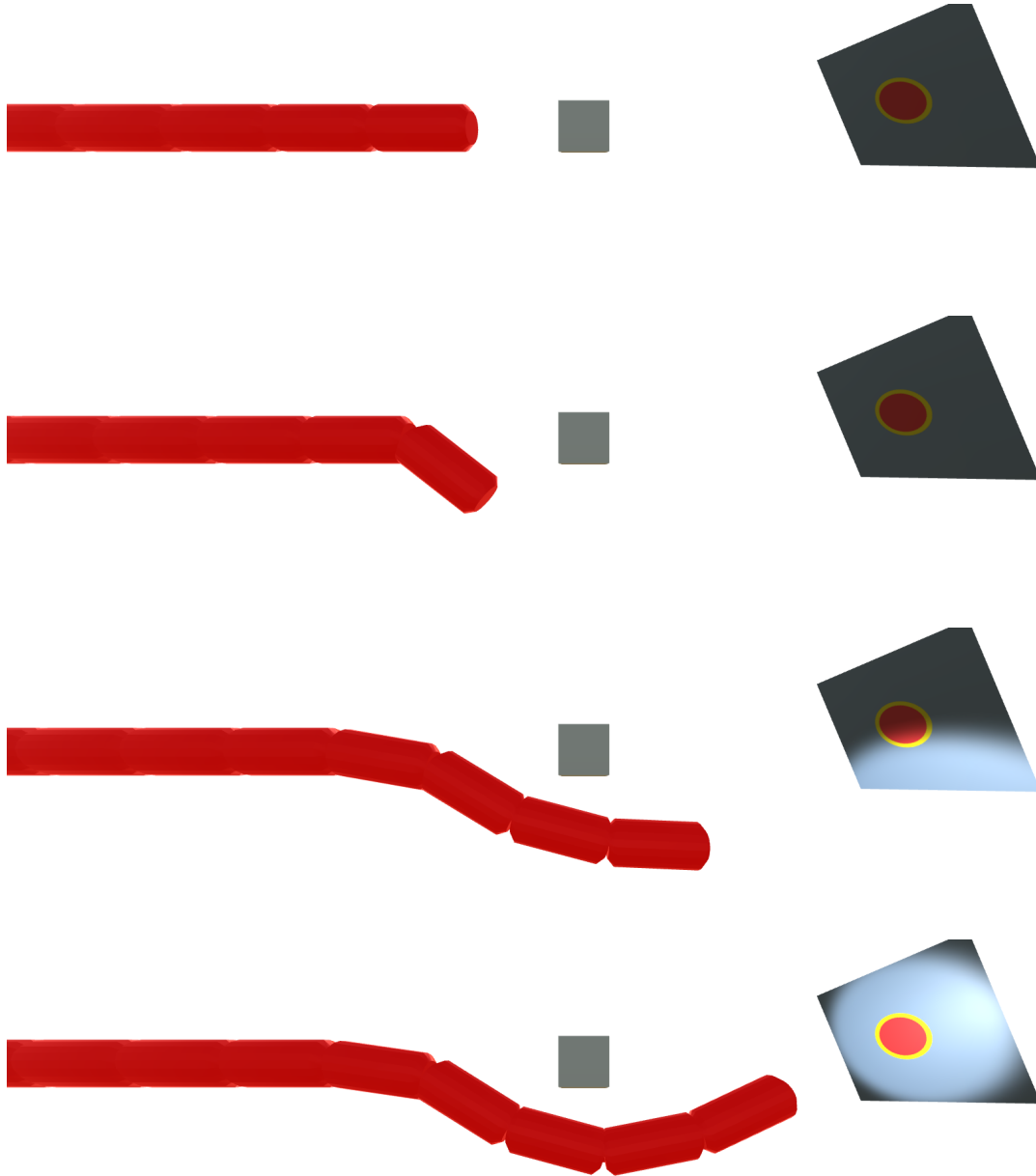


Fig. 7.1: Concept behind the movement of the endoscope. As the head of the endoscope is moved through the push/pull and angle adjustments, the rest of the body conforms to the route. This allows the endoscope to both avoid obstacles, and show the target from head on even if it is at an angle.

7.2.1 Endoscope structure

Before the user study, we considered multiple prototypes of the endoscope with varying features. The overall limiting factor was the predicted cost for the configuration. The features that were varied were:

- Length of each joint.
- DOF for each joint.
- Rotation of the camera.

To compare the different versions, we prototyped them in a virtual environment and used them for visualization of the chest, eventually arriving at a single setting we considered optimal. We prototyped joint lengths of 2, 3, 4, 5, and 6 centimeters for operating inside the cavity. 2 cm was the lower limit due to estimated technical limits for the eventual hardware implementation. It quickly became obvious that lengths longer than 2 cm were unfeasible due to the limited operating space within the cavity. Meanwhile, 2 cm was considered short enough to allow for the endoscope to bend freely while being still able to house all of the components. To keep the structure simple, each joint can only move in one dimension, and all of the joints are aligned so that the endoscope on the whole can only change its shape in two dimensions if only the joints are manipulated. Adding a joint at the leading end of the endoscope that could be orientated orthogonally from the rest of the joints was considered to allow turning the camera more freely, but ultimately it was decided that by carefully choosing the endoscope's orientation it would not be necessary. Further, mechanically rotating the endoscope camera is not needed, as we can perform the rotation in software.

As the endoscope is based on the idea that the only part actively moved by the user is the head of the endoscope, and the rest of it follows in a "follow-the-leader" manner, we have to calculate the angles of the individual joints continuously. In the prototype we have 7 joints $k \in 0, \dots, 6$ whose angles β_k are calculated from a list of key-value pairs $[t, \alpha]$ where t is the distance the endoscope has travelled from the port, and α the absolute angle taken at that point. For an endoscope that has travelled a distance T from the port, the angle ϕ_t at any point t between $[0, T[$ is defined as

$$\phi_t = \sum_{i=0}^I w(t_i, t) * (\alpha_i - \alpha_{i-1}), \quad (7.2.1)$$

where

$$w(t_i, t) = 1 - \max(0, \min(t_i - t, 1)).$$

As α represents the global angle, and we need the relative angle, we calculate $\alpha_i - \alpha_{i-1}$ and define $\alpha_i|_{i<0} = 0$. By using equation 7.2.1, we can define the angle for all of the joints k for any given t .

7.2.2 Controls

To control the virtual endoscopes, we mapped the movement to an Xbox 360 game controller. For controlling the rigid endoscope, we gave the user the following possibilities:

- Rotation of the endoscope by using the port as the fulcrum.
- Twisting of the endoscope along its main axis.
- Insertion/retraction of the endoscope.

For the prototype, we additionally gave the user the ability to manipulate the orientation of the endoscope head along one dimension.

7.2.3 Evaluation

To evaluate the MAE, we performed a user study. The general flow of the test is shown in Fig. 7.4. The participants comprised trainees and faculty members of the Department of Thoracic Surgery, Kyoto University Hospital. Total number of participants was 7, and the average age was 34.8. Informed consent was obtained from all individual participants included in the study. Fig. 7.2 shows the setup for the user evaluation. Every participant had previous experience in VATS and using an endoscope during it. We asked the participants to visualize six different targets inside the thoracic cavity, corresponding to actual surgical targets. The targets were 3 cm by 3 cm squares that had a bullseye with radius $r = 0.5$ cm in the middle. The placement, shape and the order of the targets can be seen in Fig. 7.3. Each participant evaluated both of the endoscopes. To account for bias caused by the order the tests were taken, we randomly decided which of the endoscopes would be evaluated first for each participant. Before the actual test, the participants were allowed to practice the controls of the endoscope for a period of time they deemed necessary. During the practice 4 targets separate from the actual targets but placed at similar positions were displayed to the participant one at a time, to provide context for the training. Average time for participants to complete the test was 15 minutes.

The scoring was based on the proximity of the target to the center of the screen, as well as the angle between the camera forward vector and the target's normal vector, such that if the camera forward vector as drawn from the middle of the camera didn't pass through the bullseye, it was not considered a valid placement. The validity of the placement was

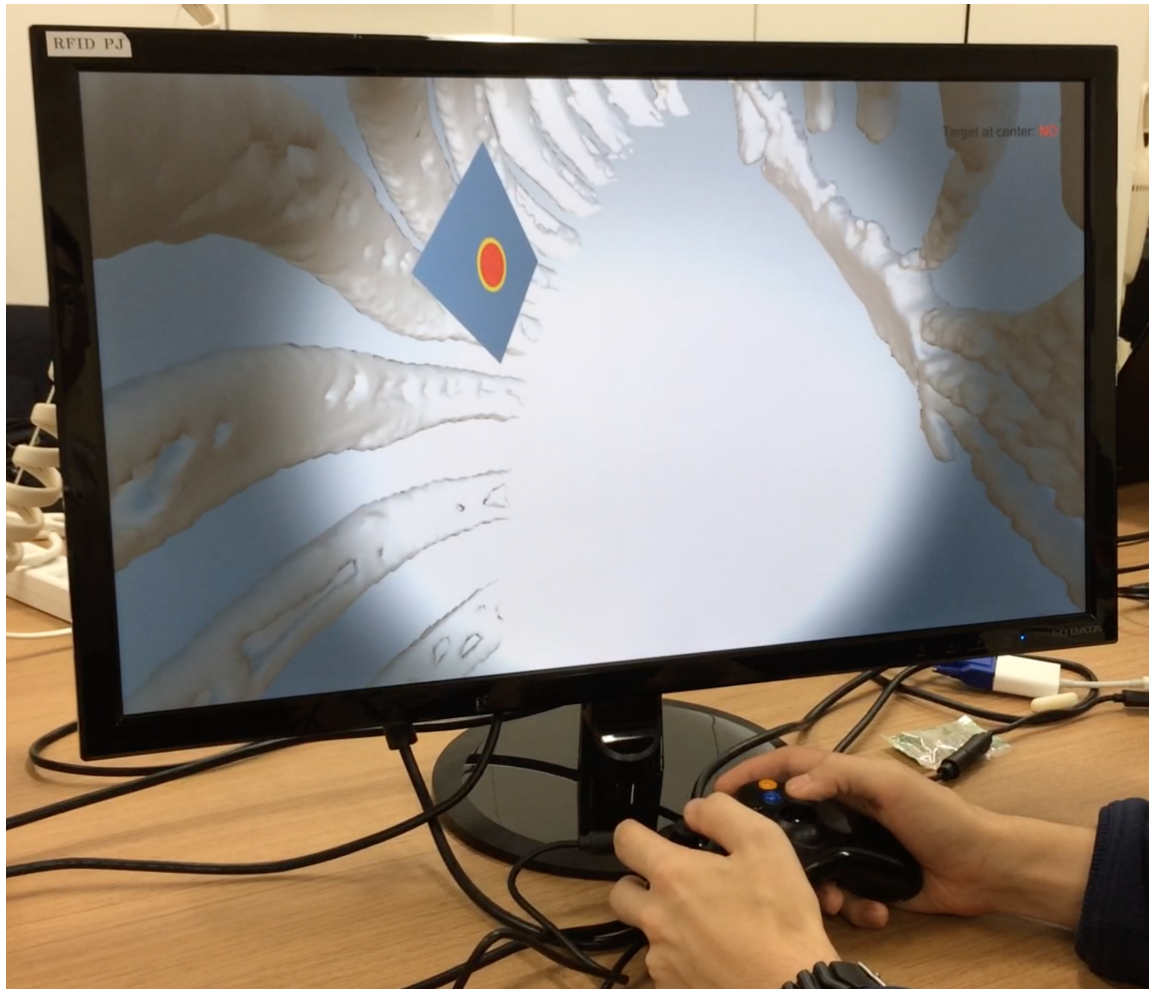


Fig. 7.2: The test setup. The controller, screen as well as the software are shown.

shown to the participant during the test. If the validity requirement is passed, and the user presses a button to record the target placement, the software calculates the angle between the abovementioned vectors, and stores it in memory. The user could not record the placement and progress until the validity requirement was passed. For evaluation purposes, the closer the endoscope optical axis is to the perpendicular view of the target, the better the score. It has been shown that the best results for surgical operations can be obtained when the working axis and the optical axis are close to same [121]. In this sense, the head-on visualization of the target can be considered to be equivalent to having the endoscope minimizing the angle to the working axis. The participants were briefed to the scoring system, and indicated to pursue the best visualization angle they deemed reasonably achievable, according to their experience in endoscopic surgery.

After the test, we collected a questionnaire to get feedback on the experience. In the questionnaire, the following questions were asked:

- Participant information
 - Age
 - Sex
 - Rigid endoscope evaluation (Likert, 5-point scale)
 - Explanation for the controls was satisfactory.
 - I could understand the controls.
 - If I became more accustomed to the controls, I feel I could better visualize the target.
 - MAE evaluation (Likert, 5-point scale)
 - Explanation for the controls was satisfactory.
 - I could understand the controls.
 - If I become more accustomed to the controls, I feel I could better visualize the target.
 - Method comparisons (Conventional/Proposed/Neither)
 - With which method was it easier to capture the target?
 - Which method felt easier to use?
 - Which method did you feel more secure in using?
 - Free-form feedback
-

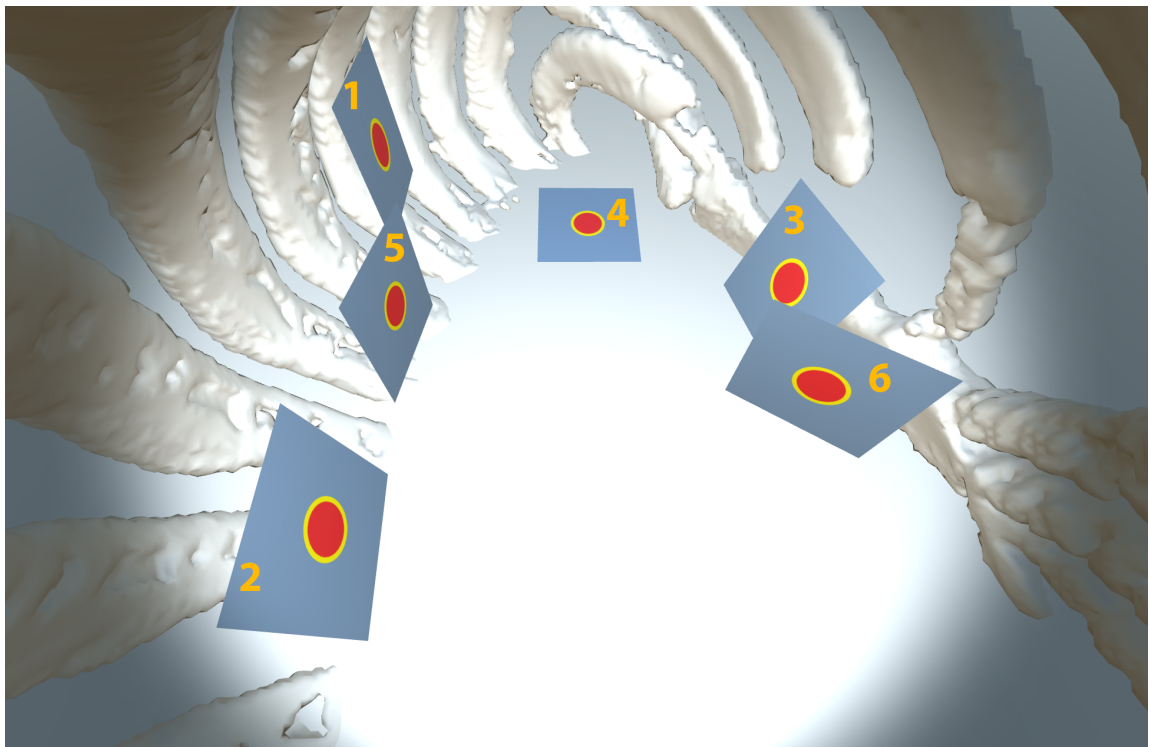


Fig. 7.3: A screen capture of the test, showing the targets the participants were asked to navigate to. Numbers indicate the order in which the targets were shown to each participant.

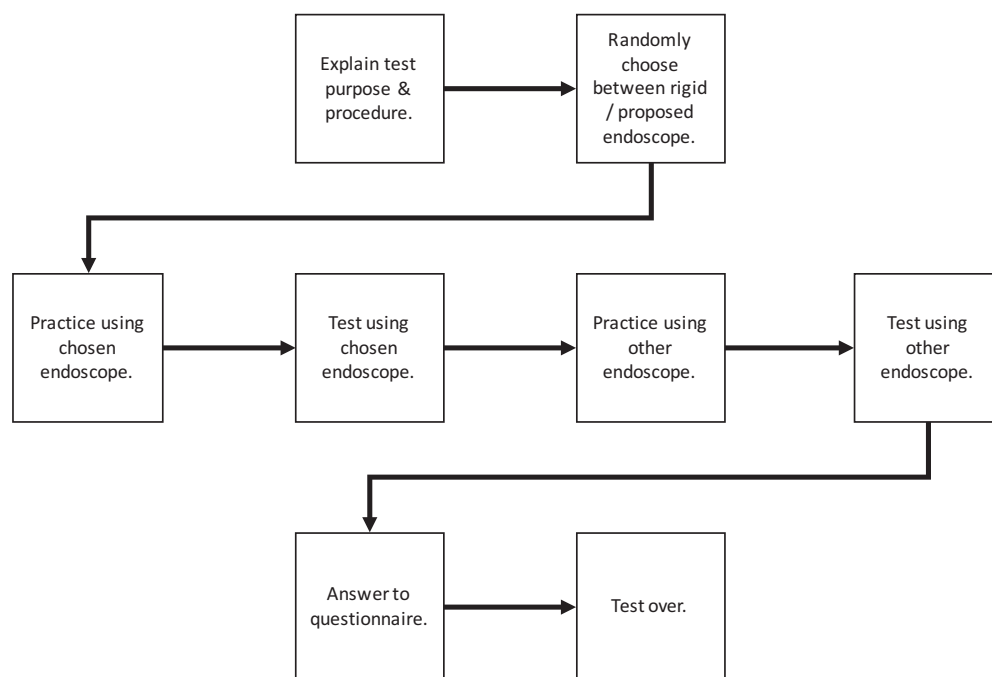


Fig. 7.4: The course of events for the test for each participant. After testing both of the endoscopes, the participants answered a short questionnaire to evaluate their experience.

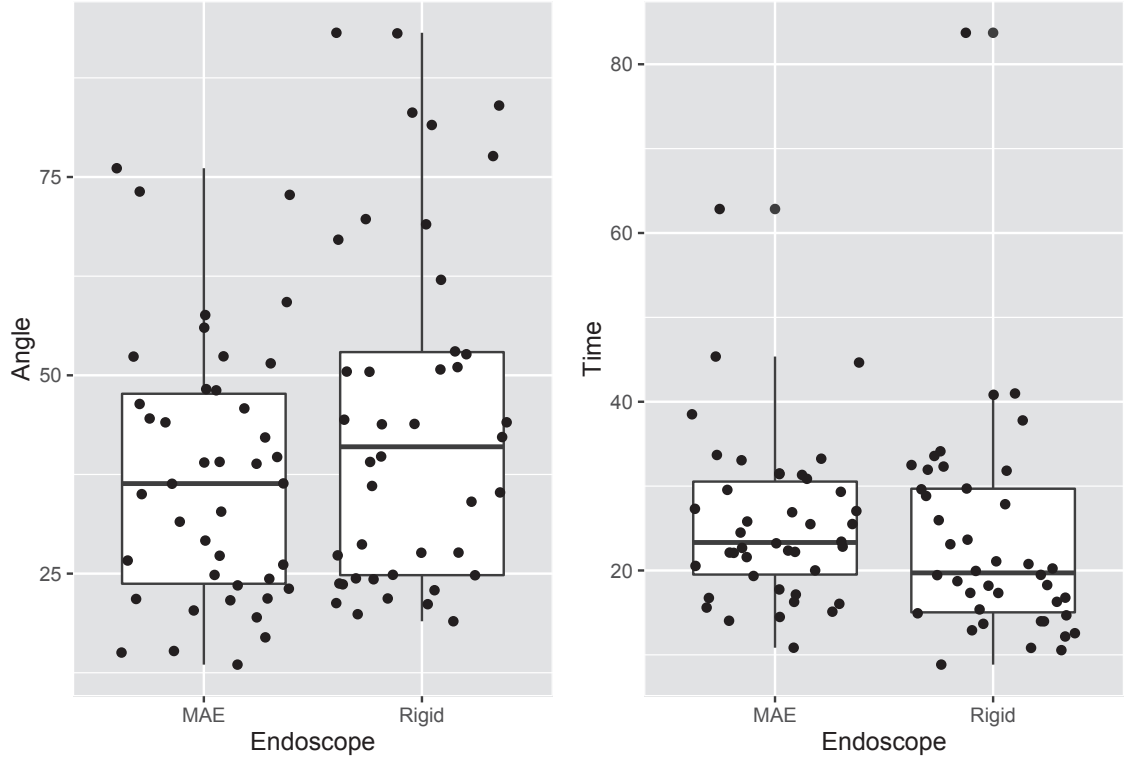


Fig. 7.5: A combined scatterplot-boxplot of the angles achieved (left) and times taken (right) per task for each endoscope.

7.3 Results

Results of the study are shown in Table 7.1. For both time and angle, a lower value equals a better score. Accordingly, the winner of the category is highlighted by bolding. As can be seen, on average the conventional endoscope achieved faster operation, while the MAE achieved better angles.

As the compared angles are not independent due to being measured by the same user, we perform the paired t-test giving $p = 0.01$ with $t = 2.74$ and a mean difference of 7.25, indicating that with a statistically significant difference, the MAE provides a lower angle between the optical axis and the target normal axis. For the times, similarly performing the paired one-tailed t-test gives $p = 0.42$ and $t = -0.81$ with mean difference of -2.09 , indicating that the mean time taken for the normal endoscope is not significantly less than the time taken for the proposed method.

Combined scatter-box plots for the angles and times taken for both types of endoscope are shown in Fig. 7.5. A Bland-Altman plot showing the differences vs. means for both metrics is shown in Fig. 7.6.

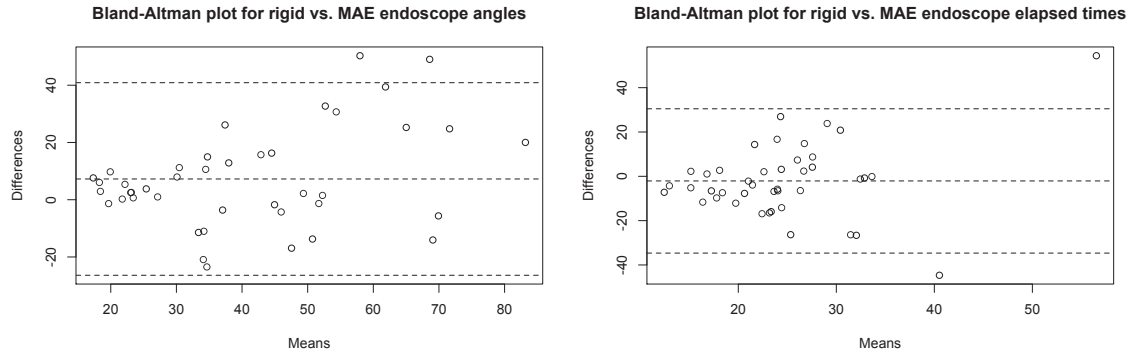


Fig. 7.6: Bland-Altman plots showing the relation between differences and means for the two endoscopes for both angles and times elapsed for the tests. For calculating the differences, a positive value indicates that the score was larger for the rigid endoscope.

For the questionnaire answers, the data is shown in Table 7.2.

7.4 Discussion

The results were as expected when concerning the achieved angle: In a statistically significant way, the proposed method performed on average better than the traditional endoscope. Similarly, regarding the average time spent on operation, the mean score was in agreement with the expectations, but the data was not sufficient to confirm significance. Both the inter-task and inter-user standard deviations were quite high, which can be explained by the difficulties in navigating using an unfamiliar game controller, as well as the differences between experience levels of the users. Indeed, as pure personal conjecture, it seemed that there was a portion of test participants who could quickly grasp the concept of the proposed method, and those who struggled for quite a while. Most difficulties seemed to come from using the “twist” function of the endoscope. Lowest scores for angles were obtained in task 4 (Fig. 7.3) (combined average 23.3°), which was the easiest to navigate to. On the opposite, the most difficult task for the users was task 1 (combined average 59.0°), which had been anticipated as one of the more difficult ones. On top of that, this task was presented to the users as the first target. Care should be taken when interpreting these figures, especially considering the limited number of participants.

Although on average the users performed tasks faster using the rigid endoscope, the differences weren’t significant. This might be caused by two factors: the users took quite a long time after arriving close to the target to find the optimal angle for the rigid endoscope, whereas in many cases the MAE allows the user to just turn the camera towards the target. Also, the Xbox controller already provided a baseline difficulty level for many of the users,

Table 7.1: Comparison of the average times and angles for the traditional endoscope and the proposed MAE. Total average times are shown on the top, and task-specific and user-specific times follow. For each row, the better scores are bolded.

	Rigid endoscope		MAE	
	Time (s)	Angle (deg.)	Time (s)	Angle (deg.)
Total avg. & st. dev.	23.49 $\pm$ 12.88	44.63 $\pm$ 21.98	25.58 $\pm$ 9.77	37.38 $\pm$ 16.34
Task avgs.				
T1	33.13 $\pm$ 23.23	61.04 $\pm$ 15.60	27.61 $\pm$ 9.72	56.93 $\pm$ 18.57
T2	23.18 $\pm$ 10.36	55.25 $\pm$ 23.98	21.34 $\pm$ 6.81	39.03 $\pm$ 12.13
T3	21.54 $\pm$ 8.93	37.85 $\pm$ 17.05	29.98 $\pm$ 16.05	32.73 $\pm$ 9.96
T4	15.89 $\pm$ 7.77	22.36 $\pm$ 2.23	29.61 $\pm$ 10.27	24.25 $\pm$ 9.96
T5	24.24 $\pm$ 10.01	56.50 $\pm$ 22.07	24.58 $\pm$ 3.67	44.80 $\pm$ 12.03
T6	22.99 $\pm$ 8.03	34.78 $\pm$ 18.60	20.37 $\pm$ 5.86	26.56 $\pm$ 10.63
User avgs.				
U1	34.81 $\pm$ 24.92	47.40 $\pm$ 22.74	29.39 $\pm$ 10.07	40.82 $\pm$ 20.30
U2	23.13 $\pm$ 8.32	61.92 $\pm$ 27.22	25.59 $\pm$ 6.78	45.35 $\pm$ 19.56
U3	13.26 $\pm$ 3.22	67.49 $\pm$ 24.07	17.51 $\pm$ 4.25	45.23 $\pm$ 14.37
U4	24.54 $\pm$ 6.61	35.35 $\pm$ 11.88	28.11 $\pm$ 3.87	44.71 $\pm$ 14.20
U5	16.09 $\pm$ 3.37	32.67 $\pm$ 13.00	33.44 $\pm$ 17.58	21.89 $\pm$ 6.34
U6	31.21 $\pm$ 10.95	34.84 $\pm$ 10.76	18.00 $\pm$ 4.76	35.05 $\pm$ 12.46
U7	21.41 $\pm$ 7.81	32.74 $\pm$ 12.01	27.03 $\pm$ 4.65	28.64 $\pm$ 13.27

Table 7.2: Questionnaire results data. For Likert-scale questions, the options for answers were: 1: Strongly disagree, 2: Disagree, 3: Neither agree nor disagree, 4: Agree, 5: Strongly agree. For the comparisons, the participant could choose between 1: “rigid endoscope”, 2: “multi-articulated endoscope”, or 3: “neither”.

Rigid endoscope					
	1	2	3	4	5
Explanation for the controls was satisfactory	0	0	1	2	4
I could understand the controls	0	0	1	2	4
If I became more accustomed to the controls, I feel I could better visualize the target.	0	0	1	3	3
Multij. endoscope					
	1	2	3	4	5
Explanation for the controls was satisfactory	0	0	0	3	4
I could understand the controls	0	0	0	5	2
If I became more accustomed to the controls, I feel I could better visualize the target.	0	0	0	2	5
Comparisons					
	1	2	3		
With which method was it easier to capture the target?	0%	100%	0%		
Which method felt easier to use?	42.9%	57.1%	0%		
Which method did you feel more secure in using?	14.3%	71.4%	14.3%		

which may have a diminishing impact on the relative difference.

Grasping the relationships between all the different input possibilities seemed to pose difficulty for most participants in the case of the MAE. For example, twisting a regular endoscope can be thought of as applying a rotation to the point of view along the endoscope axis, whereas in the case that the leading segment doesn't share that axis, like often happens in the case of a MAE, a translation of the viewpoint is also required. This makes it difficult to predict how the multijointed endoscope acts if the current shape is not clear in the user's head. Some form of visualization method through a graphical user interface could be useful here, to allow the medical professional to better grasp the shape of the endoscope from the beginning, and to quickly adjust to the changes in shape.

Due to the joints' movement being restricted to one DOF (pitch), the proposed MAE cannot compensate for movement along the joints' rotation axis, as opposed to the case where the joints have 2DOF (pitch, yaw). This can be alleviated with careful route planning. However, in cases where the endoscope is engaged in a complex shape, partial or full retraction may be needed to navigate to a new target.

For the questionnaire, similar scores were given for both of the methods. On the whole, all of the participants agreed that it was easier to show the target with the MAE, which seems to correspond with the results as well. On the ease of use, the opinions were much more divided, which also reflects in the less clear division between the usage time between the two endoscopes. On the last question regarding the feeling of security, a majority answered that they felt more secure in using the MAE, which suggests that indeed a demand for this kind of technology exists.

Recently, a virtual reality VATS lobectomy training simulator was presented [110]. In the freeform feedback section, a lot of the participants expressed that they felt the endoscope had potential, but would require more practice to be able to use it well. In a similar vein to the simulator, the evaluation software could be improved or altered to a training simulator to further get the user accustomed to the use of the endoscope.

We are currently looking forward to applying the results to a real prototype, and are considering a torque wire-driver to implement the articulation mechanism.

7.5 Conclusion

A user study using a virtual model of a novel multi-articulated endoscope was conducted. In this report, we showed that the system could provide a better view on targets inside virtual thoracic cages with no significant difference in usability of the endoscope. Although the MAE requires skill for optimal controlling, we believe that this can be solved by further

improvements, such as a more sophisticated user interface.

Chapter 8

Augmented virtuality platform for usability evaluation of a novel endoscope concept

The endoscope is a key component of minimally invasive surgery, and several methods for improving the traditional rigid endoscope have been proposed. These generally increase the complexity of the endoscope, and thus the feasibility of operation of such devices should be ensured through user testing in an environment that closely resembles the real use scenario. We present an augmented virtuality endoscopy simulator that can be used for evaluating virtual endoscope concepts. We achieve this by having the operator of the endoscope hold a physical object resembling the prototype, and through operating the physical object, have a virtual version displayed through a head-mounted display behave accordingly. We evaluate the system through implementing a previously proposed novel articulated endoscope. The system allows for usability evaluation of endoscopes in a setting closely resembling that of the real surgical room without the need for constructing a working physical prototype, potentially saving costs and allowing for quicker prototyping of novel endoscope concepts.

8.1 Introduction

Surgical interventions in the abdominal and chest areas have traditionally relied on operations such as thoracotomy or laparotomy to gain access to the surgical target. Endoscopic methods with devices such as the thoracoscope and the laparoscope have enabled surgeons to access the target organs through small incisions. This kind of approach is commonly called minimally invasive surgery [103]. Owing to the fact that it only requires a small in-

cision, it has been shown to be the superior approach when compared to some conventional invasive methods such as thoracotomy [104].

The endoscope is thus an indispensable tool when considering minimally invasive surgery. Traditionally, the rigid endoscope, a 300-400 mm in length and 3-10 mm in diameter tubular device has been used for visualization. Various improvements to it have been proposed, ranging from adding a single joint to the distal end of the rigid endoscope [108] to multi-articulated approaches that allow the user to modify the shape of the endoscope with varying degrees of freedom [7, 8, 115, 116]. These kind of multi-articulated endoscopes (MAE) generally have the advantage of allowing for finer control of the point of view to the surgical target, as well as potentially enabling manoeuvrement around obstacles within the range of motion of the endoscope.

Previously, we have looked into a novel endoscope concept based on joints with 1 degree of freedom (DOF). Through a comparative user study, we have showed that without a significant difference in elapsed time, the operators could achieve a significantly better angle of visualization for the surgical target, indicating that usage of such a device could result in less complications caused by the surgery due to increased operational efficiency. The evaluation for this endoscope was done by having the user operate the endoscope with a game controller, and having the endoscopic view presented through a computer screen. While this may be sufficient for comparing the maneuverability of the proposed endoscope, it does not necessarily reflect on the actual usability, as the endoscope is likely to be manually operated.

We propose a system to allow the evaluation of these kind of novel endoscopes in an augmented virtual environment. Using a commercial VR head-mounted display (HMD) and motion tracking software, the surgeon can manipulate a physical model of the endoscope, and through wearing the HMD, see the virtual prototype model with multiple joints behave accordingly. Similar systems have been previously applied to delivery simulation for medical training [122], and virtual reality has been used widely for endoscopic training [15, 123]. We show that the system is a feasible platform for evaluating the usability and user experience of the novel MAE in a situation closely resembling the actual use scenario, by allowing the operator the tactile feedback from holding a physical device while virtually modeling the behavior of the MAE. Such a system can be used for measuring the usability and user experience of novel endoscope concepts in a way that closely mimics the real hospital environment. This closes the gap between the purely virtual environment that lacks physical feedback, and the purely real environment, in which costly prototypes would need to be developed for accurate evaluation.

8.2 Methods

We display the test environment to the user through the HTC Vive HMD. For tracking the endoscope movement we attach markers to the physical model and use the OptiTrack motion tracking system to follow the markers' movements, and update the virtual endoscope position and orientation accordingly. The virtual environment was created with the Unity game engine. By streaming the marker locations from the motion tracking system to Unity, we match the physical endoscope model's movements to the virtually modeled MAE, and by having the user wear the HMD, overlay the virtual model on top of the physical model to give the illusion of operating a more complex endoscope while allowing for realistic manipulation and tactile feedback. Figure 8.1 shows the main components of the system. V_1 and V_2 represent the Vive lighthouses, while O_n represent the OptiTrack cameras. T marks the physical operating target, i.e. the endoscope, while HMD is the head-mounted display worn by the user.

The endoscope itself is based on a previously proposed novel MAE, and is composed of seven aligned joints with each having 1DOF. By manipulating the angle of the joint at the distal end of the endoscope, and through insertion and withdrawal of the endoscope from the target, the user can proceed to the desired direction while the rest of the joints conform to the path taken by the leading portion. The distal joint angle manipulation was achieved by using two buttons for angle increase/decrease on one of the controllers provided by the HTC Vive system.

Several requirements were determined to have to be fulfilled for the environment to be usable. First of all, the movements in the physical world have to closely match to the movements in the virtual reality. This applies to both the user and to the endoscope: if position tracking is not accurate, the user will have problems focusing on the task, and the accuracy of the evaluation suffers. Related to this, the system should not interfere with what is being tested. Physical equipment used should be as unobtrusive as possible, to allow for free interaction with the endoscope. Further, the system should allow the user enough space for movement to perform the tasks without becoming a hindrance. The evaluated device should be modeled as close to the specifications as possible, to ensure relevancy of the results. Likewise, the dimensions of the physical component corresponding to the device should match the dimensions of the virtual counterpart on the relevant parts. In general, we propose three main requirements for the system to be fit for its purpose:

1. The mapping between the real world positioning and the virtual reality coordinates should be accurate.
2. The system should be both unobtrusive and accommodating for relevant actions.

3. The physical and virtual counterparts of the prototyped device should match the concept to a high degree for the relevant parts.

Given that these three requirements are fulfilled, the system should be suitable for modeling the endoscope and evaluating its usability in a way that the results translate well to the surgical room.

8.3 Evaluation

To evaluate the system, we look at each requirement and see how well the system fulfills it. By considering the performance in regards to each requirement, we can say if the proposed system is suitable for usability evaluation of the particular kind of endoscopic system.

8.3.1 Mapping between real/virtual world

Figure 8.2 shows the test scenario in both real and virtual environments. To achieve a high correlation between them, we use the OptiTrack system with Prime 13 infrared cameras for tracking markers attached to the physical endoscope. The cameras have a latency of 4.2 ms and can track the markers with sub-millimeter accuracy¹. For tracking the user, we use the HTC Vive system, which uses various sensors in unison with two base stations that project structured light and can track the HMD with roughly 2 mm accuracy¹. To combine the two sets of data from the devices, we use the Unity game engine which has plug-ins to stream data from both. Matching the coordinate systems is a question of providing proper positional and rotational offsets, as both systems use similar scaling for coordinates.

The moment the user rotates the joints of the virtual model, it ceases to be fully in agreement with the physical model. We avoid any problems caused by this by limiting the extent of withdrawal of the endoscope, so that for the parts where it undergoes deformation, neither the physical nor the virtual model are directly accessible by the user.

Since both the Vive and the OptiTrack use infrared light for tracking, there is a possibility for interference. In practice however, this was not found to be an issue, and in the case that interference is exhibited, the systems can be synchronized to eliminate the problem.

8.3.2 Accommodativeness

The general layout of the area can be seen in Figure 8.1. The Vive allows an area with diagonal length of 5 m to be set up for tracking. For the OptiTrack system, multiple

¹The accuracy numbers for the Vive are from <http://doc-ok.org/?p=1478>, and for the Prime 13 camera from <http://optitrack.com/products/prime-13/>. Both pages were accessed on 2017-02-27.

cameras are needed for accurate tracking but the maximum range for each camera is 12.2 m. For our experiments we set up a 2 m x 2 m area within which the tests are carried out.

The HTC Vive has a cord that connects the HMD to the computer, which can get in the way, but careful wiring can eliminate this problem, and wireless solutions exist for transmitting data. Likewise, the markers that are required for the OptiTrack system can interfere, but with proper placement should get in the way only minimally.

8.3.3 Models' suitability for concept

For the physical counterpart of the endoscope we repurposed a laparoscopy training simulator's endoscope handle for the system (Figure 8.2). This provided a realistic feel for the endoscope operation, but 3D printed parts could also be considered for replacement. For the virtual counterpart, we modeled the movement of the joints in a "follow-me" kind of way as described before. The target for endoscopy was a CT-scan from the right lung area with separate segmented meshes for the muscular and skeletal structures, and the pleura.

8.4 Discussion

The proposed system allows the evaluation of the usability of the novel endoscope in a more realistic setting. By having the user operate the physical endoscope model, the disconnect caused by having to use e.g. a gamepad for controls is removed, and the user can get a feel for how the endoscope would actually handle.

Limitations for the proposed system include the difficulty of implementing realistic haptic feedback for the parts where the virtual model doesn't agree with the physical one. A vibration mechanism could be implemented to improve the experience, but limitations remain when compared to using a physical prototype and e.g. an animal carcass. This further emphasizes that the system is best employed in the early stage of the development process.

To control the angle of the distal joint, we used a controller provided by the HTC Vive system, held in the operator's free hand. In practice dedicating a whole hand for this operation is not feasible, so alternatives such as integrating the buttons to the endoscope, or using foot pedals should be considered.

Using the system we can compare endoscopes with varying complexities to model the tradeoffs between usability and capability. Since the user is completely immersed in the virtual environment, it also allows for the simulation of alternative visual feedback methods such as HMD's within the virtual environment. Methods for reducing user disorientation caused by either the environment or the endoscope's dynamic shape (such as displaying

the endoscope like in Figure 8.3) can be also be explored, and user evaluations conducted to see what kind of information can be displayed for positive effect. In a similar vein, the system could be applied for endoscopy training.

The proposed system can be expanded to be a platform for exploring related concepts such as visualization and endoscope manipulation techniques, stereoscopic endoscopy, and training in an environment closely resembling the surgical room. For future work, continuing development in this direction should be considered.

8.5 Conclusion

We presented a platform for evaluating the usability of a novel endoscope system. By showing that it adheres to the requirements, it can be considered to be a viable platform for not only evaluating the usability of the previously proposed endoscope, but also for future evaluation of various concepts related to endoscopy, such as methods for visualization and operation of the endoscope.

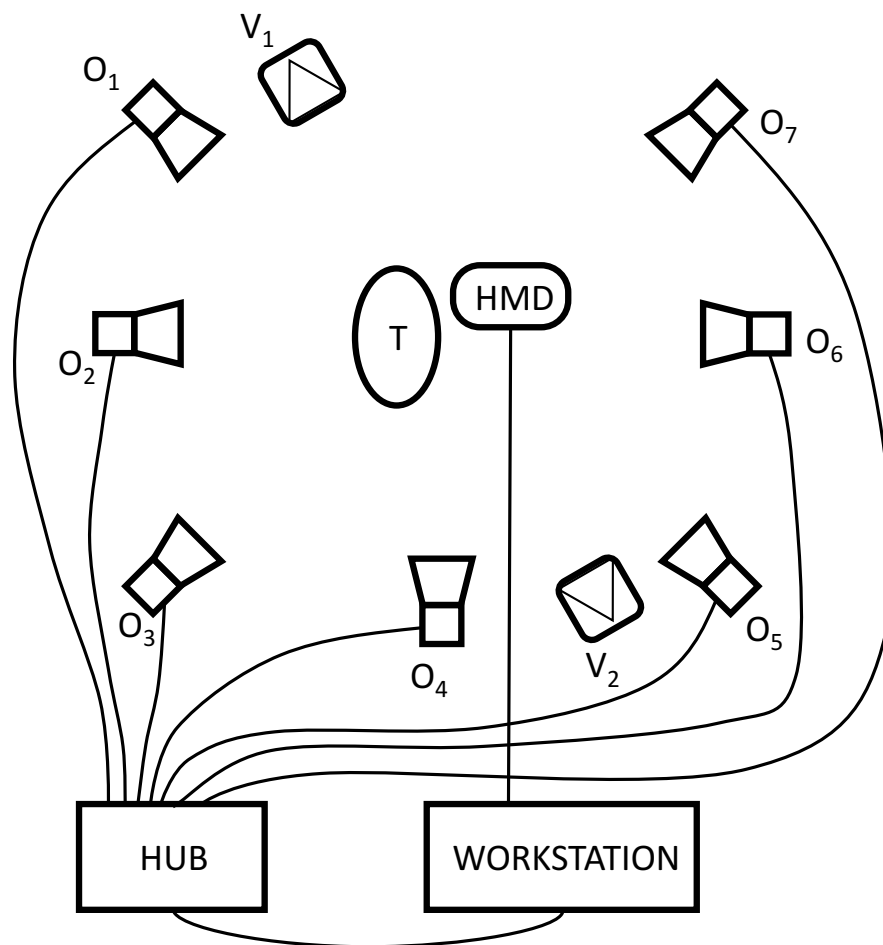


Fig. 8.1: Overview of the components of the system.

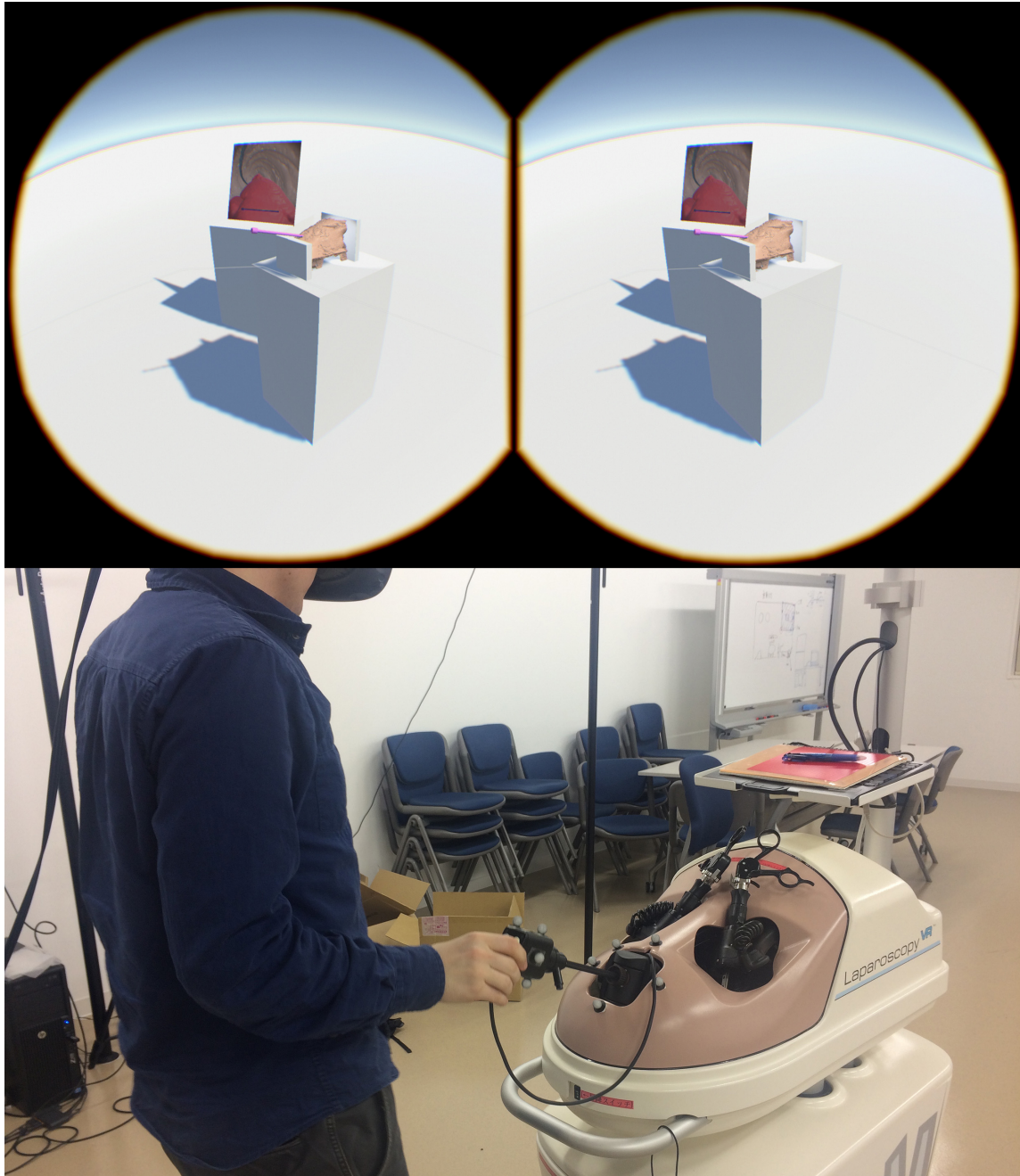


Fig. 8.2: Top: the virtual stereoscopic view towards the endoscopic target provided by the HMD. Bottom: view in the physical world from the same viewpoint.

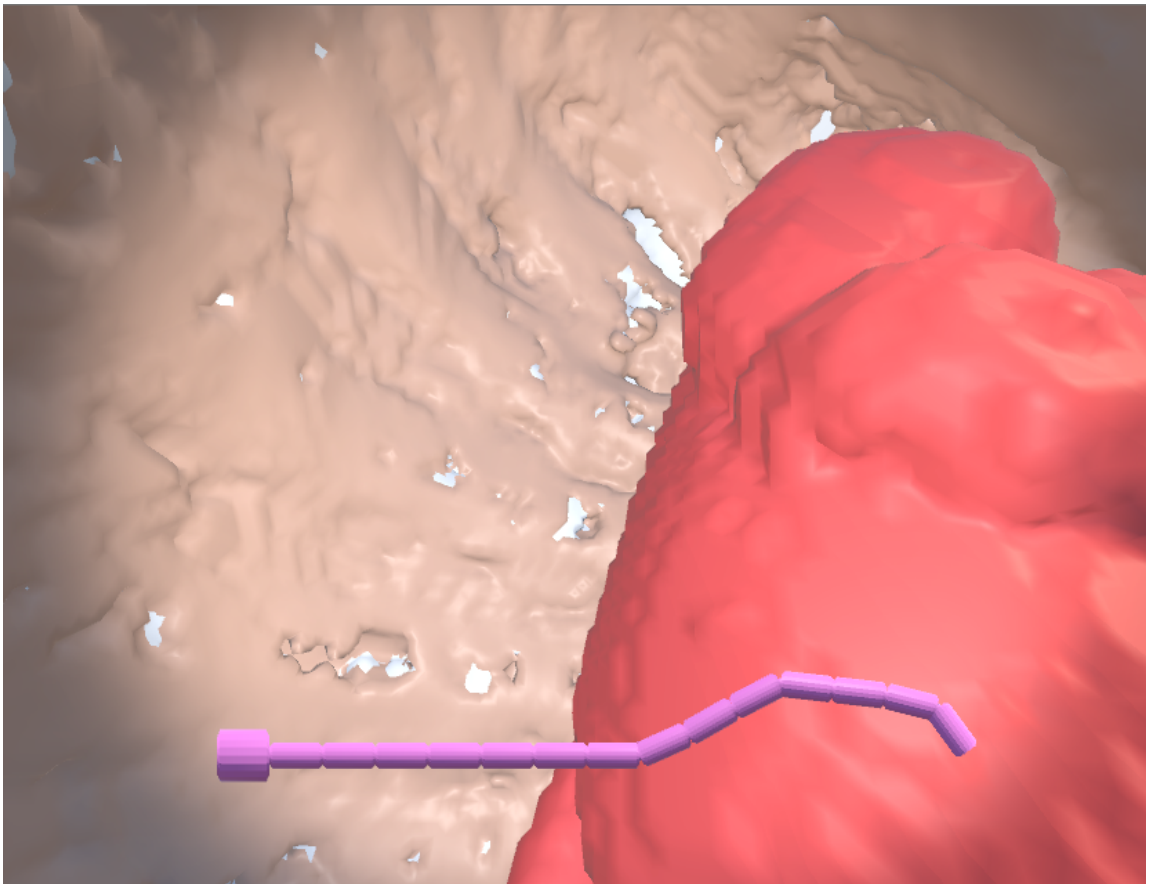


Fig. 8.3: The endoscopic view inside the target, with current endoscope shape overlayed.

Chapter 9

Discussion

In this thesis, we explored different ways to support the development and use of novel endoscopes through means of visual computing. To this end, in Chapters 3, 4, 5, and 6, we explored tools and clinical applications for a novel endoscope combining conventional visible light imaging and OCT imaging capabilities. We applied visuocomputational endoscopy for both operational support of the novel dual-mode endoscope through image processing and reconstruction of the OCT images, as well as for clinical decision support in developing methods for measuring the degree of endolymphatic hydrops in cochlea. In Chapters 7 and 8, we explored visuocomputational support for the development and usability evaluation of a novel multi-articulated endoscope (MAE).

Within the visuocomputational endoscopy pipeline, as shown in Fig. 2.3, the works covered all aspects labeled, with the first half covering the middle three (multimodal endoscopy, image processing, and image analysis), and the second half covering the structural modifications and computer graphics aspect. Creating a guideline for endoscope research and development based on and combining the researches done for both projects would be an interesting endeavor, and might be considered in the future.

A combination of the two approaches, a dual-mode MAE, and the usability challenges it implies might be an interesting project to tackle through visuocomputational endoscopy means. By modeling the endoscope virtually and generating mock data resembling the OCT/digital endoscope data, usability issues related to the endoscope shape and how to present information to the user could be explored. Especially methods for controlling and visualization of the endoscope orientation could be explored through a setup similar to that presented in Chapter 8.

In a broader perspective, the researches introduced in this dissertation showcase the different aspects of visuocomputational endoscopy as described in Fig. 9.1. The figure shows the examples of visuocomputational support introduced in Chapter 2, as well as

the research results presented in the previous chapters. For the latter, the parts dealing with visual computing applied to the novel OCT/digital endoscope mostly focused on the operational support (image segmentation, dual-mode image synthesis, and surface reconstruction methods) to decision support (measurement of endolymphatic hydrops). The research related to the MAE mainly applied visual computing in the development phase, but the system could be feasibly extended to provide training support for the novel endoscope, also. The framework in Fig. 9.1 could further be split into applications prior to use of endoscopy (top row) and applications during use of endoscopy (bottom row). This classification seems to make sense in that the tools developed for support in one of the four aspects can be readily applied for the horizontally neighboring aspect too.

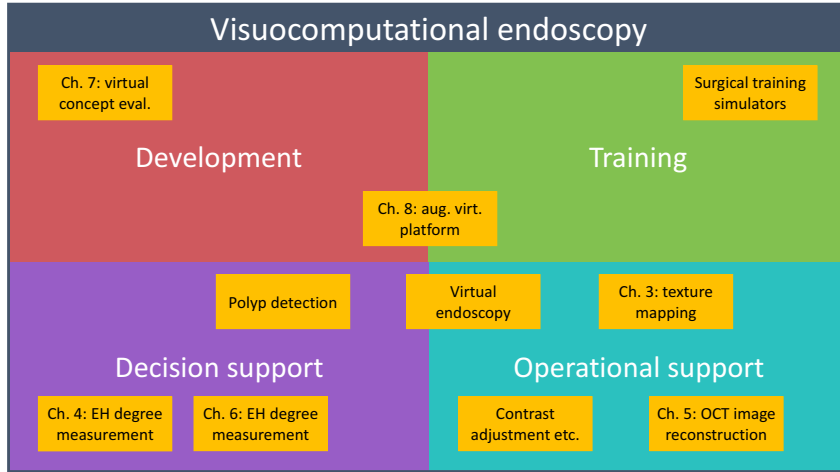


Fig. 9.1: The four aspects of visuocomputational endoscopy: development of novel endoscopic methods, training for endoscopic procedures, operational support, and decision support.

In the next few paragraphs, we will go through discussion relevant to the two projects, starting with the OCT/digital endoscope project, and moving to the MAE project.

9.1 OCT project

In general, more robust methods for handling the noise need to be developed. Due to the nature of the speckle noise, a strongly reflecting particle in the imaged sample can cause whole column clusters to become corrupted with noise of high intensity. This kind of noise is hard to filter with traditional noise filtering methods, and can get detected as a feature by even the most sophisticated segmentation and feature detection algorithms. In the future, algorithms for specifically dealing with this type of noise could be developed to

reduce their effect on segmentation and such.

For the dual-mode endoscope, considering that its ultimate application is in real-time endoscopic imaging, algorithms that take advantage of the continuous nature of imaging such as motion estimation, perhaps in the form of optical flow, should be taken advantage of to reduce the amount of unnecessary re-segmentation from frame to frame. This could help push the proposed method more towards true real-time performance.

The first algorithm focusing on Ménière's disease, based on optical flow, turned out to be quite vulnerable to the speckle noise present in OCT images. Performance could be improved either by better management of the noise, which likely would increase processing time, or by using another method for motion estimation that doesn't rely so much on the brightness consistency principle like optical flow does.

Evaluating the tools for the measurement of endolymphatic hydrops proved to be quite difficult. Although it is possible to manufacture hydrops in mice via knocking out the relevant gene, it still only enables the procurement of only cochlea with no hydrops, or those with severe hydrops. Harvesting cochleae that exhibit a mild form of hydrops would require going through a lot of mice. However, this kind of large scale validation should be considered, not only due to the validation aspect on the tools created, but because it would also enable the forming of a precise model of how the Reissner's membrane behaves in general when the hydrops progresses. Having data to "fill the gap" between the none/severe hydrops groups would allow us to come up with similar classification for the angle, leading to a truly quantitative assessment method for the severity of hydrops.

Although evaluations were made on mice cochleae and the method should translate to work well in human cochleae, some modifications are bound to be needed. The larger size of the cochlea in humans means that less of the features are probably visible in an OCT volumetric image. This means that more skill is needed for the medical professional to get the image so that the needed features are visible, but due to the increased feature sizes, the algorithm should perform better. Due to this it would be interesting and enlightening to be able to work with some human cochlea data in the future.

Throughout the publications, we used a simplification of the Reissner's membrane structure and the cochlear wall shape by fitting planes onto them. In reality, neither of the structures, except the Reissner's membrane in the case of no hydrops, are perfectly straight, being more closely estimated by a parabolic surface if anything. Thus, instead of the angle, the curvature of the membrane could be used as an indicator of the severity of endolymphatic hydrops.

Similar to this, the diffusion maps we used for segmentation could also be used for noise reduction. Combined with a proper kernel that takes into account the characteristics

of speckle noise that OCT images exhibit, it could prove to be a powerful tool for noise reduction when combined with some form of local noise filtering method. Exploring possibilities for this in the future could lead to some nice results when it comes to de-noising OCT images.

It would be interesting to try to perform some sort of automatic detection of the OCT features to solve the initial value problem that in the presented methods were solved through user input. Although full cochlear modelling based on the spiral structure was briefly considered, it was discarded as being too complicated, and furthermore due to only part of the cochlea usually being able to be captured by the OCT sensor. To this end, modelling local features like the three scalae, or even simpler the curved shape of the inner wall of the tunnels could be utilised to automatically detect the location of the Reissner's membrane, for example.

Automatic segmentation of different tissue types could also be considered, for example to separate the outer layer of the cochlear wall from the inner layer. Methods for achieving this could include taking advantage of the speckle noise, and the different types of texture it creates depending on the material. For this, texture analysis methods could be considered and might provide good results. This could prove valuable for delineating cancer cells from benign soft tissue, for example.

9.2 MAE project

In Chapter 7 we built a virtual model of a novel multi-articulated endoscope (MAE), and through a user test, confirmed its feasibility as a clinical tool. Building upon the experiences, in Chapter 8 we explored an augmented virtuality platform for modeling and evaluating the endoscope and similar kinds of clinical equipment in a highly realistic fashion.

For the MAE, although previous researches have investigated the applicability of multi-articulated endoscopes with ≥ 2 DOF per joint to various clinical targets, usability and UX research has been lacking. In this sense, it would be valuable to compare the usability of MAE to models of these endoscopes, too.

The platform for evaluating the novel MAE has some limitations. The markers required for the motion tracking system to work require careful placement to not get in the way during operation. Unfortunately, a good solution for this does not exist yet. A marker-less Kinect-type system could be used for tracking the movements, but the accuracy of such a solution is questionable.

Chapter 10

Conclusion

10.1 Contributions

In this dissertation, we investigated multiple methods for improving upon conventional endoscopy through visual computing. We looked at some previous approaches in Chapter 2, and provided a structured view to endoscopic research and development in Fig. 2.3. Based on this structure, and our contributions from Chapters 3–8, we proposed a model of visuocomputational endoscopy in Fig. 9.1.

Texture mapping for dual-mode en face-OCT endoscope system

In Chapter 3 we explored image processing methods for a novel endoscopic system combining a traditional digital endoscope with an OCT imaging sensor. The focus was on providing a real-time system for mapping the camera image to surface data obtained from the OCT image. This kind of functionality could be valuable for medical professionals when diagnosing diseases, and especially in surgery to see not only the surface image but the structural changes in tissue more accurately.

Estimating the degree of EH using optical flow

For Chapter 4, we showed that it is feasible to use the angle between the Reissner’s membrane and the cochlear wall it is connected to, to measure the degree of endolymphatic hydrops. We used motion estimation to calculate the slice-by-slice movement of the features marked by the user, and fitted planes using least squares method to obtain the angle. We then performed some tests on cochleae obtained from mice for further validation. The results obtained show that there is a correlation between the area calculated and the angle between these features, indicating that the tool could be used for clinical purposes when

calculating the degree of endolymphatic hydrops.

Segmentation of OCT images using diffusion maps

In this chapter, we used diffusion maps and superpixels to show that we could segment the cochlear structure from OCT images that include the delicate features such as the Reissner's membrane, suggesting that the segmented data could be used in further analysis of the degree of endolymphatic hydrops.

Semi-automatic measurement of endolymphatic hydrops from volumetric optical coherence tomography images

In Chapter 6, we used the method from the previous chapter, along with a novel method for delineation of the two previously mentioned features for calculating the degree of endolymphatic hydrops. Although the results had a higher standard deviation than those presented in Chapter 4, the method works as a proof of concept and further adjusting of the parameters should work to allow for more accurate results.

Evaluation of a novel multi-articulated endoscope: proof of concept through a virtual simulation

In Chapter 7, a new point of view was introduced: utilizing virtual prototyping to demonstrate the feasibility of a novel endoscope design. Through building a virtual model, it was possible to conduct a feasibility study and evaluate the usability of the design. By avoiding the need to construct physical prototypes, quick evaluation of various configurations, and facilitation of user study was made possible.

Augmented virtuality platform for usability evaluation of a novel endoscope concept

Finally, Chapter 8 introduced a platform whose design was inspired by the research conducted in Chapter 7. By having a user interact in an augmented virtuality environment, they can evaluate novel endoscopic prototypes in a highly realistic setting without the need for a working prototype.

10.2 Future direction

The work undertaken for this dissertation could continue in several ways. In general, applications of visuocomputational endoscopy involve multiple areas of visual computing,

allowing for a broad selection of possible future research avenues.

For the evaluation of endolymphatic hydrops, further improvement of the delineation of the inner cochlea features could help obtain even more accurate measurements of the angle, and thus lead to more accurate analysis. Other ways to measure the degree such as the beforehand mentioned curvature of the Reissner's membrane could be explored. Definitely, more thorough evaluation of the method is needed before any clinical use can be considered. This could be done by harvesting mouse cochleae with various levels of endolymphatic hydrops.

Another direction would be to focus on improving the handling of the endoscopic data, particularly for real-time processing. Eventually as the endoscope is finished and as the applications for it become clear, adding functionality in respect to those applications could prove to be valuable.

Furthermore, feature delineation in form of detecting cancer cells from surrounding benign tissue, or bone tissue from soft tissue, could be attempted. To this end, texture analysis methods could be considered.

For the virtual endoscope prototype, an obvious next step is the development of a physical prototype based on the specifications. Replicating the user study with the physical prototype would also give valuable information about the validity of the virtual prototype user experiment results.

Finally, regarding the augmented virtuality platform, although it was created specifically for the evaluation of the MAE, very little domain-specific knowledge pertaining to the endoscope was used, and its application for other medical device development could be a direction for future research.

Bibliography

- [1] Henry Gray. *Anatomy of the human body*. Lea & Febiger, 1918.
- [2] William G Bradley. History of medical imaging. *Proceedings of the American Philosophical Society*, 152(3):349–361, 2008.
- [3] Ronnie Beth Bush, Hanna Leonhardt, Irving M Bush, and Ralph R Landes. Dr. bozzini’s lichtleiter: A translation of his original article (1806). *Urology*, 3(1):119–123, 1974.
- [4] G Berci and KA Forde. History of endoscopy. *Surgical endoscopy*, 14(1):5–15, 2000.
- [5] BI Hirschowitz, LE Curtiss, CW Peters, and HM Pollard. Demonstration of a new gastroscope, the fiberscope. *Gastroenterology*, 35(1), 1958.
- [6] Gavriel Iddan, Gavriel Meron, Arkady Glukhovsky, and Paul Swain. Wireless capsule endoscopy. *Nature*, 405(6785):417, 2000.
- [7] Amir Degani, Howie Choset, Alon Wolf, and Marco A Zenati. Highly articulated robotic probe for minimally invasive surgery. In *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pages 4167–4172. IEEE, 2006.
- [8] Takeyoshi Ota, Amir Degani, David Schwartzman, Brett Zubiate, Jeremy McGarvey, Howie Choset, and Marco A Zenati. A novel highly articulated robotic surgical system for epicardial ablation. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pages 250–253. IEEE, 2008.
- [9] Kazuhiro Gono, Takashi Obi, Masahiro Yamaguchi, Nagaaki Ohyama, Hirohisa Machida, Yasushi Sano, Shigeaki Yoshida, Yasuo Hamamoto, and Takao Endo. Appearance of enhanced tissue features in narrow-band endoscopic imaging. *Journal of biomedical optics*, 9(3):568–577, 2004.

-
- [10] Eugene P Dimagno, Patrick T Regan, Jonathan E Clain, EM James, and James L Buxton. Human endoscopic ultrasonography. *Gastroenterology*, 83(4):824–829, 1982.
 - [11] Guillermo J Tearney, Mark E Brezinski, Brett E Bouma, Stephen A Boppart, Costas Pitris, James F Southern, and James G Fujimoto. In vivo endoscopic optical biopsy with optical coherence tomography. *Science*, 276(5321):2037–2039, 1997.
 - [12] Jun-Hyung Cho. Advanced imaging technology other than narrow band imaging. *Clinical endoscopy*, 48(6):503, 2015.
 - [13] Stéphane Nicolau, Luc Soler, Didier Mutter, and Jacques Marescaux. Augmented reality in laparoscopic surgical oncology. *Surgical oncology*, 20(3):189–201, 2011.
 - [14] Michael Liedlgruber and Andreas Uhl. Computer-aided decision support systems for endoscopy in the gastrointestinal tract: a review. *IEEE reviews in biomedical engineering*, 4:73–88, 2011.
 - [15] Uwe Kühnapfel, Hüseyin Kemâl Cakmak, and Heiko Maaß. Endoscopic surgery training using virtual reality and deformable tissue simulation. *Computers & graphics*, 24(5):671–682, 2000.
 - [16] AF Fercher, K Mengedoht, and W Werner. Eye-length measurement by interferometry with partially coherent light. *Optics letters*, 13(3):186–188, 1988.
 - [17] David Huang, Eric A Swanson, Charles P Lin, Joel S Schuman, William G Stinson, Warren Chang, Michael R Hee, Thomas Flotte, Kenton Gregory, Carmen A Puliafito, et al. Optical coherence tomography. *Science*, 254(5035):1178–1181, 1991.
 - [18] Joseph A Izatt, Michael R Hee, Eric A Swanson, Charles P Lin, David Huang, Joel S Schuman, Carmen A Puliafito, and James G Fujimoto. Micrometer-scale resolution imaging of the anterior eye in vivo with optical coherence tomography. *Archives of ophthalmology*, 112(12):1584–1589, 1994.
 - [19] Carmen A Puliafito, Michael R Hee, Charles P Lin, Elias Reichel, Joel S Schuman, Jay S Duker, Joseph A Izatt, Eric A Swanson, and James G Fujimoto. Imaging of macular diseases with optical coherence tomography. *Ophthalmology*, 102(2):217–229, 1995.
 - [20] Michael R Hee, Carmen A Puliafito, Carlton Wong, Jay S Duker, Elias Reichel, Joel S Schuman, Eric A Swanson, and James G Fujimoto. Optical coherence tomography of macular holes. *Ophthalmology*, 102(5):748–756, 1995.
-

- [21] Raheleh Kafieh, Hossein Rabbani, and Saeed Kermani. A review of algorithms for segmentation of optical coherence tomography from retina. *Journal of medical signals and sensors*, 3(1):45, 2013.
 - [22] Zahid Yaqoob, Jigang Wu, Emily J McDowell, Xin Heng, and Changhuei Yang. Methods and application areas of endoscopic optical coherence tomography. *Journal of biomedical optics*, 11(6):063001–063001, 2006.
 - [23] John Strasswimmer, Mark C Pierce, B Hyle Park, Victor Neel, and Johannes F de Boer. Polarization-sensitive optical coherence tomography of invasive basal cell carcinoma. *Journal of biomedical optics*, 9(2):292–298, 2004.
 - [24] Julia Welzel, Maike Bruhns, and Helmut H Wolff. Optical coherence tomography in contact dermatitis and psoriasis. *Archives of dermatological research*, 295(2):50–55, 2003.
 - [25] A Baumgartner, S Dichtl, CK Hitzenberger, H Sattmann, B Robl, A Moritz, AF Fercher, and W Sperr. Polarization-sensitive optical coherence tomography of dental structures. *Caries research*, 34(1):59–69, 2000.
 - [26] Hrebesh M Subhash, Viviana Davila, Hai Sun, Anh T Nguyen-Huynh, Xiaorui Shi, Alfred L Nuttall, and Ruikang K Wang. Volumetric in vivo imaging of microvascular perfusion within the intact cochlea in mice using ultra-high sensitive optical microangiography. *Medical Imaging, IEEE Transactions on*, 30(2):224–230, 2011.
 - [27] Yosuke Tona, Tatsunori Sakamoto, Takayuki Nakagawa, Tsunemichi Adachi, Mirei Taniguchi, Hiroko Torii, Kiyomi Hamaguchi, Shin-ichiro Kitajiri, and Juichi Ito. In vivo imaging of mouse cochlea by optical coherence tomography. *Otology & Neurotology*, 35:e84–e89, 2014.
 - [28] M Bashkansky, MD Duncan, M Kahn, D Lewis, and J Reintjes. Subsurface defect detection in ceramics by high-speed high-resolution optical coherent tomography. *Optics Letters*, 22(1):61–63, 1997.
 - [29] Eugene Hecht. *Optics, 5th edition*. Pearson Education, Inc., 2016.
 - [30] Michelle L Gabriele, Gadi Wollstein, Hiroshi Ishikawa, Larry Kagemann, Juan Xu, Lindsey S Folio, and Joel S Schuman. Optical coherence tomography: history, current status, and laboratory work. *Investigative ophthalmology & visual science*, 52(5):2425–2436, 2011.
-

-
- [31] Maciej Wojtkowski, Rainer Leitgeb, Andrzej Kowalczyk, Tomasz Bajraszewski, and Adolf F Fercher. In vivo human retinal imaging by fourier domain optical coherence tomography. *Journal of biomedical optics*, 7(3):457–463, 2002.
- [32] Benjamin Potsaid, Iwona Gorczynska, Vivek J Srinivasan, Yueli Chen, James Jiang, Alex Cable, and James G Fujimoto. Ultrahigh speed spectral/fourier domain ophthalmic imaging at 70,000 to 312,500 axial scans per second. *Optics express*, 16(19):15149–15169, 2008.
- [33] Vivek J Srinivasan, Desmond C Adler, Yueli Chen, Iwona Gorczynska, Robert Huber, Jay S Duker, Joel S Schuman, and James G Fujimoto. Ultrahigh-speed optical coherence tomography for three-dimensional and en face imaging of the retina and optic nerve head. *Investigative ophthalmology & visual science*, 49(11):5103–5110, 2008.
- [34] Joseph M Schmitt, SH Xiang, and Kin M Yung. Speckle in optical coherence tomography. *Journal of biomedical optics*, 4(1):95–105, 1999.
- [35] Kirk W Gossage, Tomasz S Tkaczyk, Jeffrey J Rodriguez, and Jennifer K Barton. Texture analysis of optical coherence tomography images: feasibility for tissue classification. *Journal of biomedical optics*, 8(3):570–575, 2003.
- [36] Robert J Zawadzki, Alfred R Fuller, David F Wiley, Bernd Hamann, Stacey S Choi, and John S Werner. Adaptation of a support vector machine algorithm for segmentation and visualization of retinal structures in volumetric optical coherence tomography data sets. *Journal of biomedical optics*, 12(4):041206–041206, 2007.
- [37] Ying T Pan, Zi L Wu, Zhi J Yuan, Zheng G Wang, and Cong W Du. Subcellular imaging of epithelium with time-lapse optical coherence tomography. *Journal of biomedical optics*, 12(5):050504–050504, 2007.
- [38] Joseph W Goodman. Some fundamental properties of speckle*. *JOSA*, 66(11):1145–1150, 1976.
- [39] Maciej Szkulmowski, Iwona Gorczynska, Daniel Szlag, Marcin Sylwestrzak, Andrzej Kowalczyk, and Maciej Wojtkowski. Efficient reduction of speckle noise in optical coherence tomography. *Optics express*, 20(2):1337–1359, 2012.
- [40] Michael Pircher, Erich Götzinger, Rainer Leitgeb, Adolf Fercher, and Christoph Hitzenberger. Measurement and imaging of water concentration in human
-

- cornea with differential absorption optical coherence tomography. *Optics Express*, 11(18):2190–2197, 2003.
- [41] SH Xiang, Liang Zhou, and Joseph M Schmitt. Speckle noise reduction for optical coherence tomography. In *BiOS Europe '97*, pages 79–88. International Society for Optics and Photonics, 1998.
- [42] Delia Cabrera Fernández, Harry M Salinas, and Carmen A Puliafito. Automated detection of retinal layer structures on optical coherence tomography images. *Optics Express*, 13(25):10200–10216, 2005.
- [43] Ruikang K Wang. Signal degradation by multiple scattering in optical coherence tomography of dense tissue: a monte carlo study towards optical clearing of biotissues. *Physics in medicine and biology*, 47(13):2281, 2002.
- [44] Mona Kathryn Garvin, Michael David Abramoff, Xiaodong Wu, Stephen R Russell, Trudy L Burns, and Milan Sonka. Automated 3-d intraretinal layer segmentation of macular spectral-domain optical coherence tomography images. *Medical Imaging, IEEE Transactions on*, 28(9):1436–1447, 2009.
- [45] Stephanie J Chiu, Xiao T Li, Peter Nicholas, Cynthia A Toth, Joseph A Izatt, and Sina Farsiu. Automatic segmentation of seven retinal layers in sdopt images congruent with expert manual segmentation. *Optics express*, 18(18):19413–19428, 2010.
- [46] Raheleh Kafieh, Hossein Rabbani, Michael D Abramoff, and Milan Sonka. Intraretinal layer segmentation of 3d optical coherence tomography using coarse grained diffusion map. *Medical image analysis*, 17(8):907–928, 2013.
- [47] Andrew M Rollins, Rujchai Ung-Arunyawee, Amitabh Chak, Richard CK Wong, Kenji Kobayashi, Michael V Sivak, and Joseph A Izatt. Real-time in vivo imaging of human gastrointestinal ultrastructure by use of endoscopic optical coherence tomography with a novel efficient interferometer design. *Optics letters*, 24(19):1358–1360, 1999.
- [48] Yingtian Pan, Huikai Xie, and Gary K Fedder. Endoscopic optical coherence tomography based on a microelectromechanical mirror. *Optics letters*, 26(24):1966–1968, 2001.
- [49] Shuliang Jiao, Zhixing Xie, Hao F Zhang, and Carmen A Puliafito. Simultaneous multimodal imaging with integrated photoacoustic microscopy and optical coherence tomography. *Optics letters*, 34(19):2961–2963, 2009.
-

- [50] R Andrew Wall and Jennifer K Barton. Fluorescence-based surface magnifying chro-moendoscopy and optical coherence tomography endoscope. *Journal of biomedical optics*, 17(8):0860031–0860037, 2012.
 - [51] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Susstrunk. Slic superpixels compared to state-of-the-art superpixel meth-ods. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(11):2274–2282, 2012.
 - [52] Lorraine A Everett, Inna A Belyantseva, Konrad Noben-Trauth, Raquel Cantos, Amy Chen, Sneha I Thakkar, Shelley L Hoogstraten-Miller, Bechara Kachar, Doris K Wu, and Eric D Green. Targeted disruption of mouse pds provides insight about the inner-ear defects encountered in pendred syndrome. *Human Molecular Genetics*, 10(2):153–161, 2001.
 - [53] Hamed Sajjadi and Michael M Paparella. Meniere’s disease. *The Lancet*, 372(9636):406–414, 2008.
 - [54] Tsutomu Nakashima, Shinji Naganawa, Makoto Sugiura, Masaaki Teranishi, Michi-hiko Sone, Hideo Hayashi, Seiichi Nakata, Naomi Katayama, and Ieda Maria Ishida. Visualization of endolymphatic hydrops in patients with meniere’s disease. *The Laryngoscope*, 117(3):415–420, 2007.
 - [55] Ilmari Pyykkö, Jing Zou, Dennis Poe, Tsutomu Nakashima, and Shinji Naganawa. Magnetic resonance imaging of the inner ear in meniere’s disease. *Otolaryngologic Clinics of North America*, 43(5):1059–1080, 2010.
 - [56] Charles Skinner Hallpike and H Cairns. Observations on the pathology of meniere’s syndrome. *The Journal of Laryngology & Otology*, 53(10):625–655, 1938.
 - [57] Kyoshio Yamakawa. Über die pathologische veränderung bei einem meniere-kranken. *J Otorhinolaryngol Soc Jpn*, 44:2310–2312, 1938.
 - [58] Alec N Salt, Miriam M Henson, Sally L Gewalt, Arthur W Keating, John E DeMott, and OW Henson. Detection and quantification of endolymphatic hydrops in the guinea pig cochlea by magnetic resonance microscopy. *Hearing research*, 88(1):79–86, 1995.
 - [59] Tsutomu Nakashima, Shinji Naganawa, Ilmari Pyykkö, William PR Gibson, Michi-hiko Sone, Seiichi Nakata, and Masaaki Teranishi. Grading of endolymphatic hydrops using magnetic resonance imaging. *Acta Oto-Laryngologica*, 129(sup560):5–8, 2009.
-

- [60] Akinobu Kakigi, Yuya Takubo, Naoya Egami, Akinori Kashio, Munetaka Ushio, Takashi Sakamoto, Shinji Yamashita, and Tatsuya Yamasoba. Evaluation of the internal structure of normal and pathological guinea pig cochleae using optical coherence tomography. *Audiology and Neurotology*, 18(5):335–343, 2013.
 - [61] WL Curati, M Graif, DPE Kingsley, T King, CL Scholtz, and RE Steiner. Mri in acoustic neuroma: a review of 35 patients. *Neuroradiology*, 28(3):208–214, 1986.
 - [62] PJ Hulcelle, GC Doods, P Mathurin, and Guy Cornelis. Mri assessment of unsuspected dural sinus thrombosis. *Neuroradiology*, 31(3):217–221, 1989.
 - [63] Michael R Hee, Carmen A Puliafito, Carlton Wong, Jay S Duker, Elias Reichel, Bryan Rutledge, Joel S Schuman, Eric A Swanson, and James G Fujimoto. Quantitative assessment of macular edema with optical coherence tomography. *Archives of ophthalmology*, 113(8):1019–1029, 1995.
 - [64] A.M. Gilroy, B.R. MacPherson, and L.M. Ross. *Atlas of Anatomy*. Thieme Anatomy. Thieme, 2008.
 - [65] D Purves, GJ Augustine, D Fitzpatrick, LC Katz, AS LaMantia, JO McNamara, and SM Williams. *Neuroscience, 2nd Edition*. Sunderland (MA) Sinauer Associates, Inc., 2001.
 - [66] L Naftalin and M Spencer Harrison. Circulation of labyrinthine fluids. *J Laryngol Otol*, 72(2):118–136, 1958.
 - [67] National Institute of Health. Meniere’s disease. NIH Publication no. 10–3404, 2010.
 - [68] P Ménière. Sur une forme de surdit  grave d pendant d’une l sion de l’oreille interne. *Gaz M d de Paris*, 16:29, 1861.
 - [69] Franz Altmann and Edmund P Fowler. Histological findings in meniere’s symptom complex. *Ann Otol Rhinol Laryngol*, 52(1):52–80, 1943.
 - [70] Robert S Kimura. Lii experimental blockage of the endolymphatic duct and sac and its effect on the inner ear of the guinea pig a study on endolymphatic hydrops. *Annals of Otology, Rhinology & Laryngology*, 76(3):664–687, 1967.
 - [71] H Shojaku, M Nagasaki, S Yasumura, RL Zang, K Akaogi, M Ito, and Y Watanabe. Vestibular evoked myogenic potentials among dizzy patients. *Equilibrium Res*, 56:354–9, 1997.
-

- [72] Mari Havia, Erna Kentala, and Ilmari Pyykkö. Prevalence of meniere's disease in general population of southern finland. *Otolaryngology-Head and Neck Surgery*, 133(5):762–768, 2005.
 - [73] Jeffrey P Harris and Thomas H Alexander. Current-day prevalence of meniere's syndrome. *Audiology and Neurotology*, 15(5):318–322, 2010.
 - [74] Ferdinand CA Timmer, Guangwei Zhou, John J Guinan, Sharon G Kujawa, Barbara S Herrmann, and Steven D Rauch. Vestibular evoked myogenic potential (vemp) in patients with ménière's disease with drop attacks. *The Laryngoscope*, 116(5):776–779, 2006.
 - [75] Berthold K Horn and Brian G Schunck. Determining optical flow. In *1981 Technical symposium east*, pages 319–331. International Society for Optics and Photonics, 1981.
 - [76] Bruce D Lucas, Takeo Kanade, et al. An iterative image registration technique with an application to stereo vision. In *IJCAI*, volume 81, pages 674–679, 1981.
 - [77] Jean-Yves Bouguet. Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm. *Intel Corporation*, 5:1–10, 2001.
 - [78] Michael D Abràmoff, Paulo J Magalhães, and Sunanda J Ram. Image processing with imagej. *Biophotonics international*, 11(7):36–42, 2004.
 - [79] Ajit H Goenka, Chandan J Das, and Raju Sharma. Nephrogenic systemic fibrosis: a review of the new conundrum. *National Medical Journal of India*, 2009.
 - [80] Yongjian Yu and Scott T Acton. Speckle reducing anisotropic diffusion. *Image Processing, IEEE Transactions on*, 11(11):1260–1270, 2002.
 - [81] Simon Baker, Daniel Scharstein, JP Lewis, Stefan Roth, Michael J Black, and Richard Szeliski. A database and evaluation methodology for optical flow. *International Journal of Computer Vision*, 92(1):1–31, 2011.
 - [82] Takashi Shibata, Sumiko Matsumoto, Tetsuzo Agishi, and Teiko Nagano. Visualization of reissner membrane and the spiral ganglion in human fetal cochlea by micro-computed tomography. *American journal of otolaryngology*, 30(2):112–120, 2009.
 - [83] Diao Xianfen, Chen Siping, Liang Changhong, and Wang Yuanmei. 3d semi-automatic segmentation of the cochlea and inner ear. In *Engineering in Medicine*
-

- and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, pages 6285–6288. IEEE, 2006.
- [84] Jack H Noble, Robert F Labadie, Omid Majdani, and Benoit M Dawant. Automatic segmentation of intracochlear anatomy in conventional ct. *Biomedical Engineering, IEEE Transactions on*, 58(9):2625–2632, 2011.
- [85] Tuukka Karvonen, Yuki Uranishi, Tatsunori Sakamoto, Yosuke Tona, Kazuya Okamoto, Hiroshi Tamura, and Tomohiro Kuroda. Estimation of the degree of endolymphatic hydrops using optical coherence tomography. *Advanced Biomedical Engineering*, 5(0):19–25, 2016.
- [86] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):888–905, 2000.
- [87] Luc Vincent and Pierre Soille. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 13(6):583–598, 1991.
- [88] Ronald R Coifman and Stéphane Lafon. Diffusion maps. *Applied and computational harmonic analysis*, 21(1):5–30, 2006.
- [89] Zeev Farbman, Raanan Fattal, and Dani Lischinski. Diffusion maps for edge-aware image editing. *ACM Trans. Graph.*, 29(6):145:1–145:10, December 2010.
- [90] Jingen Liu, Yang Yang, Imran Saleemi, and Mubarak Shah. Learning semantic features for action recognition via diffusion maps. *Computer Vision and Image Understanding*, 116(3):361–377, 2012.
- [91] Qianqian Fang and David A Boas. Tetrahedral mesh generation from volumetric binary and grayscale images. In *Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on*, pages 1142–1145. IEEE, 2009.
- [92] Ari S Brunschwig and Alec N Salt. Fixation-induced shrinkage of reissner’s membrane and its potential influence on the assessment of endolymph volume. *Hearing research*, 114(1):62–68, 1997.
- [93] N Slepecky and M Ulfendahl. Glutaraldehyde induces cell shape changes in isolated outer hair cells from the inner ear. *Journal of submicroscopic cytology and pathology*, 20(1):37–45, 1988.
-

- [94] Taizo Takeda, Setsuko Takeda, Hiroya Kitano, Teruhiko Okada, and Akinobu Kakigi. Endolymphatic hydrops induced by chronic administration of vasopressin. *Hearing research*, 140(1):1–6, 2000.
 - [95] Alec N Salt, John E Demott, and Robert S Kimura. Comparison of endolymph cross-sectional area measured histologically with that measured in vivo with an ionic volume marker. *Annals of Otology, Rhinology & Laryngology*, 104(11):886–894, 1995.
 - [96] Masaya Takumida, Nana Akagi, and Matti Anniko. A new animal model for meniere’s disease. *Acta oto-laryngologica*, 128(3):263–271, 2008.
 - [97] Robert Gürkov, Wilhem Flatz, Julia Louza, Michael Strupp, and Eike Krause. In vivo visualization of endolymphatic hydrops in patients with meniere’s disease: correlation with audiovestibular function. *European Archives of Oto-Rhino-Laryngology*, 268(12):1743–1748, 2011.
 - [98] Tuukka Karvonen, Yuki Uranishi, Tatsunori Sakamoto, Yosuke Tona, Kazuya Okamoto, Hiroshi Tamura, and Tomohiro Kuroda. 3d reconstruction of cochlea using optical coherence tomography. In *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, pages 5905–5908. IEEE, 2016.
 - [99] Nikola Banić and Sven Lončarić. Smart light random memory sprays retinex: a fast retinex implementation for high-quality brightness adjustment and color correction. *JOSA A*, 32(11):2136–2147, 2015.
 - [100] Yuri Boykov and Vladimir Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(9):1124–1137, 2004.
 - [101] Jack E Bresenham. Algorithm for computer control of a digital plotter. *IBM Systems journal*, 4(1):25–30, 1965.
 - [102] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.
 - [103] TN Robinson and GV Stiegmann. Minimally invasive surgery. *Endoscopy*, 36(01):48–51, 2004.
-

-
- [104] Shizuka Kaseda, Teruhiro Aoki, Nanae Hangai, and Kunihiro Shimizu. Better pulmonary function and prognosis with video-assisted thoracic surgery than with thoracotomy. *The Annals of thoracic surgery*, 70(5):1644–1646, 2000.
- [105] Daniel J Boffa, Anish Dhamija, Andrzej S Kosinski, Anthony W Kim, Frank C Detterbeck, John D Mitchell, Mark W Onaitis, and Subroto Paul. Fewer complications result from a video-assisted approach to anatomic resection of clinical stage i lung cancer. *The Journal of thoracic and cardiovascular surgery*, 148(2):637–643, 2014.
- [106] Christopher Cao, Con Manganas, Su C Ang, Sheen Peeceeyen, and Tristan D Yan. Video-assisted thoracic surgery versus open thoracotomy for non-small cell lung cancer: a meta-analysis of propensity score-matched patients. *Interactive cardiovascular and thoracic surgery*, 16(3):244–249, 2012.
- [107] Davy Cheng, Robert J Downey, Kemp Kernstine, Rex Stanbridge, Hani Shennib, Randall Wolf, Toshiya Ohtsuka, Ralph Schmid, David Waller, Hiran Fernando, et al. Video-assisted thoracic surgery in lung cancer resection: a meta-analysis and systematic review of controlled trials. *Innovations: Technology and Techniques in Cardiothoracic and Vascular Surgery*, 2(6):261–292, 2007.
- [108] Peter B Licht and Lars Ladegaard. Flexible thoracoscopy may facilitate video-assisted thoracoscopic lobectomy. *World journal of surgery*, 34(7):1470–1474, 2010.
- [109] Pyng Lee and Henri G Colt. Rigid and semirigid pleuroscopy: the future is bright. *Respirology*, 10(4):418–425, 2005.
- [110] Katrine Jensen, Flemming Bjerrum, Henrik Jessen Hansen, René Horsleben Petersen, Jesper Holst Pedersen, and Lars Konge. A new possibility in thoracoscopic virtual reality simulation training: development and testing of a novel virtual reality simulator for video-assisted thoracoscopic surgery lobectomy. *Interactive cardiovascular and thoracic surgery*, 21(4):420–426, 2015.
- [111] Kenshi Yao, Toshiyuki Matsui, Hisashi Furukawa, Tsuneyoshi Yao, Toshihiro Sakurai, and Tomoko Mitsuyasu. A new stereoscopic endoscopy system: accurate 3-dimensional measurement in vitro and in vivo with distortion-correction function. *Gastrointestinal endoscopy*, 55(3):412–420, 2002.
- [112] Patrick Honeck, Gunnar Wendt-Nordahl, Jens Rassweiler, and Thomas Knoll. Three-dimensional laparoscopic imaging improves surgical performance on standardized ex-vivo laparoscopic tasks. *Journal of endourology*, 26(8):1085–1088, 2012.
-

-
- [113] GB Hanna and A Cuschieri. Influence of the optical axis-to-target view angle on endoscopic task performance. *Surgical endoscopy*, 13(4):371–375, 1999.
- [114] PV Patil, GB Hanna, and A Cuschieri. Effect of the angle between the optical axis of the endoscope and the instruments’ plane on monitor image and surgical performance. *Surgical endoscopy*, 18(1):111–114, 2004.
- [115] Koji Ikuta, Makoto Nokata, and Satoshi Aritomi. Hyper redundant active endoscope for minimally invasive surgery. *Journal of the Robotics Society of Japan*, 16(4):569–575, 1998.
- [116] Koji Ikuta, Masahiro Tsukamoto, and Shigeo Hirose. Shape memory alloy servo actuator system with electric resistance feedback and application for active endoscope. In *Robotics and Automation, 1988. Proceedings., 1988 IEEE International Conference on*, pages 427–430. Ieee, 1988.
- [117] Jianzhong Shang, David P Noonan, C Payne, James Clark, Mikael Hans Sodergren, Ara Darzi, and G-Z Yang. An articulated universal joint based flexible access robot for minimally invasive surgery. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 1147–1152. IEEE, 2011.
- [118] Carlos M Rivera-Serrano, Paul Johnson, Brett Zubiate, Richard Kuenzler, Howie Choset, Marco Zenati, Stephen Tully, and Umamaheswar Duvvuri. A transoral highly flexible robot. *The Laryngoscope*, 122(5):1067–1071, 2012.
- [119] Roland Polet and Jaques Donnez. Using a laparoscope manipulator (lapman) in laparoscopic gynecological surgery. *Surg Technol Int*, 17(187):45, 2008.
- [120] Sonja Gillen, Benedikt Pletzer, Arthur Heiligensetzer, Petra Wolf, Jörg Kleeff, Hubertus Feussner, and Alois Fürst. Solo-surgical laparoscopic cholecystectomy with a joystick-guided camera device: a case-control study. *Surgical endoscopy*, 28(1):164, 2014.
- [121] William CS Meng, Samuel PY Kwok, KL Leung, CC Chung, WY Lau, and Arthur KC Li. Optimal position of working ports in laparoscopic surgery: an in vitro study. *Surgical Laparoscopy Endoscopy & Percutaneous Techniques*, 6(4):278–281, 1996.
- [122] Tobias Sielhorst, Tobias Obst, Rainer Burgkart, Robert Riener, and Nassir Navab. An augmented reality delivery simulator for medical training. In *International Work-*
-

shop on Augmented Environments for Medical Imaging-MICCAI Satellite Workshop, volume 141, pages 11–20, 2004.

- [123] Neal E Seymour, Anthony G Gallagher, Sanziana A Roman, Michael K O'brien, Vipin K Bansal, Dana K Andersen, and Richard M Satava. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Annals of surgery*, 236(4):458–464, 2002.

List of Publications

Journals

- Tuukka Karvonen, Yuki Uranishi, Tatsunori Sakamoto, Yosuke Tona, Kazuya Okamoto, Hiroshi Tamura and Tomohiro Kuroda, “Estimation of the degree of endolymphatic hydrops using optical coherence tomography”, *Advanced Biomedical Engineering*, 5(0):19-25, 2016.
- Tuukka Karvonen, Yusuke Muranishi, Goshiro Yamamoto, Tomohiro Kuroda, Toshihiko Sato, “Evaluation of a novel multi-articulated endoscope: proof of concept through a virtual simulation”, *International Journal of Computer Assisted Radiology and Surgery*, 1-8, 2017.
- Tuukka Karvonen, Yuki Uranishi, Tatsunori Sakamoto, Yosuke Tona, Kazuya Okamoto, Hiroshi Tamura and Tomohiro Kuroda, “Semi-Automatic Measurement of Endolymphatic Hydrops from Volumetric Optical Coherence Tomography Images”, *PLOS ONE*, 2017. (Under review)

International Conferences

- Tuukka Karvonen, Yuki Uranishi, Tatsunori Sakamoto, Yosuke Tona, Kazuya Okamoto, Hiroshi Tamura and Tomohiro Kuroda, “3D Reconstruction of Cochlea Using Optical Coherence Tomography”, *Engineering in Medicine and Biology Society (EMBC), 2016 38th Annual International Conference of the IEEE*, 2016.
- Tuukka Karvonen, Yusuke Muranishi, Goshiro Yamamoto, Tomohiro Kuroda, Toshihiko Sato, “Evaluation of a novel multi-articulated endoscope: proof of concept through a virtual simulation”, *The 8th International Conference on Information Processing in Computer-Assisted Interventions*, 2017.
- Tuukka Karvonen, Goshiro Yamamoto, Yusuke Muranishi, Toshihiko Sato, Tomohiro Kuroda, “Augmented virtuality platform for usability evaluation of a novel endoscope concept”, *EMBECE & NBC 2017*, 575–578, 2017.

Domestic Conferences

- Tuukka Karvonen, Yuki Uranishi, Tatsunori Sakamoto, Yosuke Tona, Kazuya Okamoto, Hiroshi Tamura and Tomohiro Kuroda, “Segmentation of Cochlear Structure in Optical Coherence Tomography Images”, The 15th Convention of the Japanese Society for Medical Virtual Reality, 2015.

Patents

- Tatsunori Sakamoto, Juichi Ito, Tomohiro Kuroda and Tuukka Karvonen, “Determination of relative orientation between two structures using tomographic images”, *Japanese Patent Application No. 053790/2014*