UNIVERSITÄT
BERN

Graduate School for Cellular and Biomedical Sciences
University of Bern

Advanced Medical Image Analysis of the Human Facial Nerve based on Machine Learning Technologies

PhD thesis submitted by

Ping Lu

from China

for the degree of
PhD in Biomedical Engineering

Supervisor
Prof. Dr. Mauricio Reyes
Institute for Surgical Technology and Biomechanics
Medical Faculty of the University of Bern

Co-Advisors
Prof. Dr.-Ing. Stefan Weber, Dr. Nicolas Gerber
ARTORG Center for Biomedical Engineering Research
Medical Faculty of the University of Bern

Accepted by the Faculty of Medicine, the Faculty of Science and the Vetsuisse Faculty of the University of Bern at the request of the Graduate School for Cellular and Biomedical Sciences

Bern, Dean of the Faculty of Medicine

Bern, Dean of the Faculty of Science

Bern, Dean of the Vetsuisse Faculty Bern

PhD Committee

Supervisor
Prof. Dr. Mauricio Reyes
Institute for Surgical Technology and Biomechanics
Faculty of Medicine of the University of Bern

Co-advisors
Prof. Dr.-Ing. Stefan Weber, Dr. Nicolas Gerber
ARTORG Center for Biomedical Engineering Research
Medical Faculty of the University of Bern

Mentor
Prof. Dr. Martin Frenz
Institute of Applied Physics
University of Bern

External Co-referee
Prof. Dr. Rasmus Reinhold Paulsen
Department of Applied Mathematics and Computer Science
Technical University of Denmark

To My Loving Parents ...

Acknowledgements

First and foremost, I would like to express my sincere appreciation and gratitude to my supervisor, Prof. Dr. Mauricio Reyes. He provided me with the opportunity to pursue my PhD studies in the field of medical image analysis in Switzerland and advance my future career as a scientist. Without his excellent supervision, I would not be able to complete this compelling PhD project which started in August, 2013. His academic supervision and guidance in different stages of my study have been invaluable for the development of this project. Prof. Dr. Reyes' charming personality and attitude to life and work have been a source of inspiration and motivation for me during my entire PhD studies.

I would also like to express my gratitude to my PhD committee members: my co-advisors Prof. Dr. Stefan Weber, Dr. Nicolas Gerber and my mentor Prof. Dr. Martin Frenz for their insightful comments and feedback on my work. Prof. Dr. Stefan Weber, thank you for sharing the presentation slides. It immensely helped me in presenting my work in a nice way. Dr. Nicolas Gerber, thank you for your guidance in the thesis of manual facial nerve segmentation. Prof. Dr. Martin Frenz, thank you for sharing your personal study experience with me.

I am very grateful to my former colleague Ms. Livia Barazzetti. It was an invigorating experience to cooperate with her. Thank you for sharing the knowledge on the cochlear implantation project and assisting with computer programming. Many thanks to Dr. Kate Gavaghan and Dr. Tom Williamson for sharing the experience of the HearRestore project.

I would like to thank our entire MIA group, especially Dr. Elham Taghizadeh for her expertise in Amira and excellent support during my PhD project; Dr. Waldo Valenzuela for his expertise in programming and invaluable advice; Dr. Carlos Correa-Shokiche, Raphael Meier and Dr. Vimal Chandran for their advice on how to improve this thesis; Dr. Kamal Shahim for his expertise in medical image analysis. To all my colleagues at the ISTB and at the ARTORG Center, thank you all for the conversations and parties that we enjoyed together, which helped me to keep my mind at ease in the intense hours of work. I would like to express special thanks to Dr. Jarunan Panyasantisuk for her friendship, support and inspiration during my PhD studies. I appreciate Urs Rohrer's assistance in changing batteries for my wireless mouse, Ulla Jakob's assistance in the course registration, the administration

support of ISTB Karin Fahnemann-Nolte, Denise Schär, Ulla Jakob, Anke Zürn and Esther Gnahoré. Thank you to all my English teachers for polishing my English skills, especially Marina Zinecker.

I would like to express my gratitude to the funding organization Nano_Tera for financial support of my PhD project, especially Etienne Duval who coached me on presentation skills.

Studying in Bern has been a wonderful experience for me. My study life in Bern has been enriched by the Universitätssport (Unisport), StudentInnenschaft der Universität Bern (SUB), Erasmus Student Network (ESN) and the Biomedical Engineering (BME) Club. Moreover, I have met lots of international students, learned from them and become friends with them. First, I am lucky to meet Elena Kirillova in the English class. She is very keen on any sports and strongly influenced me to do sports and be fit. Second, I am fortuitous to know Jakhan Pirhulyieva from "Welcome Apéro" for new students, and thanks for joining me in many leisure activities and guiding me in English writing and speaking. Third, I am grateful to my floormates and Chinese friends for having a great time. My sincere thanks to Zhu Tang for strong support, C++ programming guidance and excellent study suggestion.

Last but not least, I am really grateful to my parents for the love and encouragement. My parents have always taught me to be independent, given me freedom to pursue my dream, and reminded me to take care of myself.

Abstract

It is a challenge for people, especially children, to speak, improve language and communicate without hearing. In addition, if people cannot hear any sounds, they will feel isolated. At this point, cochlear implantation has become widely used as a treatment for people with a severe-to-profound sensorineural hearing loss and not benefiting from hearing aids. A cochlear implant is an electrical device, which uses the electrical signal to imitate the sound waves, as replacement of the damaged part of the inner ear. The implanted electrode stimulates the hearing nerve inside the cochlea in order to transfer the signal to the brain, which enables the person to hear.

Minimally invasive cochlear implantation is one representative case of microsurgical navigation technologies, that reduces the potential surgical trauma of the conventional cochlear implant surgery. The drill passes through the facial recess to the round window via a tiny trajectory, instead of milling a big hole. It is crucial that the drill does not hurt any critical structures, especially the facial nerve, which has the highest priority for protection. If the facial nerve is damaged, the patient will lose facial movement. Therefore, it is important to accurately segment the facial nerve in surgical planning for computing an optimal drill trajectory.

In addition, Computed Tomography (CT) or Cone-beam computed tomography (CBCT) used for surgical planning, features a low image resolution in relation to the small structure of interest, as well as patient motion artefacts. For these two main reasons, the CT or CBCT images are often blurred and visualizing the border of anatomical structures becomes difficult, which make it extremely challenge to do the surgical planning in a precise way. It is hypothesised in this thesis that an advanced CT based image analysis of the ear can lead to a highly accurate facial nerve segmentation for micro-IGS cochlear implantation.

This thesis presents two main works for personalized planning of cochlear implantation in robotic image guided system, which is developed at the University of Bern. The first task is related to motion detection. In order to estimate patient head movement, we developed an image-based pipeline. Experimental results demonstrate that the proposed approach to estimate head motion is feasible.

The second task is related to highly accurate facial segmentation from CT/CBCT images. A super-resolution strategy was developed and applied for image enhancement and sub-pixel classification. For image enhancement, facial nerve image enhancement was developed using supervised learning based on multi-output ExtraTree regressor. The enhanced image is passed to a surgical planning software, OtoPlan, and facial nerve is segmented from OtoPlan. Segmentation from the enhanced CBCT and the original CBCT showed the enhanced CBCT with the proposed approach improves the segmentation. For sub-pixel classification, an automatic random forest based super resolution classification (SRC) framework is proposed for facial nerve segmentation refinement. In order to reduce computational time, a band-based region of interest selection (ROI) segmentation results from initial segmentation is selected. Then, SRC works on these ROI segmentation results for facial nerve segmentation. Preliminary results on 3D CBCT and CT ex-vivo datasets achieved a segmentation accuracy with a Dice coefficient of 0.818 ± 0.052, surface-to-surface distance of $0.121 \pm 0.030mm$ and Hausdorff distance of $0.715 \pm 0.169mm$. Compared with two other semi-automated segmentation software tools, ITK-SNAP and GeoS, the proposed method shows accurate segmentations at sub-voxel accuracy.

Keywords: Cochlear implantation, Image guidance, Surgical planning, Head movement, Motion detection, Facial nerve segmentation, Refinement, Super resolution classification, Band-based region, Sub-voxel.

Table of contents

List of figures **11**

List of tables **17**

1 Introduction **1**

1.1 Clinical Background 1

1.1.1 Anatomy of the Ear 1

1.1.2 How hearing works 3

1.1.3 Cochlear Implant 3

1.1.4 Facial Nerve Preservation in Cochlear Implant Surgery 5

1.2 Background on Medical Imaging 7

1.2.1 Cone-beam Computed Tomography 7

1.2.2 Micro-computed Tomography 8

1.2.3 Imaging the Facial Nerve 9

1.3 Minimally Invasive Cochlear Implant Surgery 10

1.3.1 Conventional Cochlear Implant Surgery 10

1.3.2 Image-Guided Procedures in Otological Surgery 12

1.3.3 Robotic Cochlear Implant Surgery 12

1.4 Challenges 17

1.4.1 Imaging Data 17

1.4.2 Manual Segmentation 17

1.4.3 Patient Movement 17

1.5 Thesis Hypothesis, Objective and Contributions 18

1.5.1 Hypothesis 18

1.5.2 Objective 18

1.5.3 Contributions 18

1.6 Outline of the Thesis 19

2 Technical Background **21**
2.1 Introduction . . . 21
2.2 Machine Learning . . . 21
2.2.1 Machine Learning in Medical Imaging . . . 21
2.2.2 Types of Machine Learning . . . 23
2.3 Random Forest . . . 24
2.3.1 Basic Definitions . . . 25
2.3.2 Decision Tree . . . 26
2.3.3 Forest Ensemble . . . 29
2.4 Extremely Randomized Trees . . . 31
2.4.1 Algorithm . . . 31
2.4.2 Split function in Extra-Tree . . . 31
2.4.3 Model Parameters . . . 32
2.4.4 Multiple Output Trees . . . 32
2.5 Registration . . . 33

3 Motion Detection **35**
3.1 Abstract . . . 35
3.2 Introduction . . . 35
3.3 Materials and Methods . . . 38
3.3.1 Phantom data sets . . . 40
3.3.2 Preprocessing . . . 41
3.3.3 Alignments of Center Line . . . 44
3.3.4 Hausdorff Distance Calculation . . . 44
3.4 Experimental Design . . . 45
3.4.1 Evaluating detected motion with a geometrically generated Ground Truth . . . 45
3.4.2 Evaluation . . . 45
3.5 Discussions . . . 52
3.6 Conclusions . . . 52

4 Facial Nerve Image Enhancement **55**
4.1 Abstract . . . 55
4.2 Introduction . . . 55
4.3 Methods . . . 56
4.3.1 Feature Extraction . . . 57
4.3.2 Multi-Output Regression Model . . . 58

4.4 Results . . . 58
4.4.1 Facial Nerve Segmentation . . . 61
4.5 Conclusions . . . 61

5 Facial Nerve Segmentation 65
5.1 Abstract . . . 65
5.2 Introduction . . . 66
5.3 Materials and Methods . . . 67
5.3.1 Image Data . . . 67
5.3.2 Preprocessing . . . 68
5.3.3 Super-Resolution Classification (SRC) . . . 69
5.4 Experimental Design . . . 71
5.4.1 Experimental detail . . . 71
5.4.2 Segmentation initialization with OtoPlan . . . 73
5.4.3 Evaluation metrics . . . 73
5.4.4 Evaluation . . . 75
5.5 Discussions . . . 80
5.6 Conclusions . . . 83

6 Conclusion and Outlook 85
6.1 Conclusion . . . 85
6.1.1 Motion Detection: a registration based approach . . . 85
6.1.2 Facial Nerve Image Enhancement: a machine learning based approach 86
6.1.3 Facial Nerve Segmentation: a Super-Resolution Classification (SRC) machine learning based approach . . . 86
6.2 Limitations of the Work . . . 87
6.2.1 Motion Detection . . . 87
6.2.2 Facial Nerve Image Enhancement . . . 87
6.2.3 Facial Nerve Segmentation . . . 88
6.3 Outlook . . . 88
6.3.1 Motion Detection . . . 88
6.3.2 Facial Nerve Image Enhancement . . . 88
6.3.3 Facial Nerve Segmentation . . . 89

7 Appendices 91
7.1 Pseudo-code of the Extra-tree algorithm [1] . . . 91

[1] Modified from [46]

7.2 Point Set to Point Set Registration Method 93
7.3 A Protocol Description of the Registration Pipeline 96
7.4 Employed parameters of the ExtraTreesClassifier 96

Bibliography **97**

List of figures

1.1 Ear anatomy. It includes the outer ear, the middle ear and the inner ear. In the outer ear, there are the pinna and the external auditory canal. In the middle ear, there are the tympanic membrane and the ossicles including the malleus, the incus and the stapes. In the inner ear, there are the cochlea, the vestibular system, the auditory nerve, and the emicircular canals. Source: [96] 2

1.2 Illustration of how hearing works within cochlea. Hair cells at the base are responsible for high frequencies. On the contrary, hair cells at the apex are responsible for the low frequencies. Modified from: [24] and [101] 4

1.3 Cochlear implant system. The implanted electrode array is used to simulate the function of the hair cell, which pass the electrical impulse to nerve. Source: [68] 4

1.4 Facial Nerve. Left: overview of the facial nerve distribution from one side of the face. Right: Zoomed facial nerve range considered in CI surgery. Source: [122] and [124] 6

1.5 Cochlea image obtained from standard CT (left), high resolution CBCT (middle) and non-clinical micro-CT (right). The image quality increases from left to right. The border of the cochlea can be easily recognized on the highest resolution, micro-CT image. Source: [160] 9

1.6 One example of the three segments of the facial nerve canal in a CT image. These segments are used when planning a cochlear implantation. (a) The labyrinthine segment (S1), (b) The tympanic segment (S2), and (C) The mastoid segment (S3). 10

1.7 One example of manually segmented facial nerve from micro-CT. It shows the course of the facial nerve considered in the surgical planning. In fact, facial nerve segmentation demonstrate the facial nerve canal segmentation, which wraps the facial nerve. 11

1.8 A minimally invasive robotic image guided system for cochlear implant surgery. The surgical planning guides the surgeon to do cochlear implant surgery. The robot drills a small hole in the mastoid. During the drill process, the camera tracks the drill position. Source: [10] 13

1.9 The personalize surgical planning software OtoPlan for minimal invasive cochlear implant surgery developed from Bern University, Switzerland. Left: User interface of OtoPlan software. Right: Semi-automatic segmentation for the facial nerve and chorda tympani. Source: [45] 14

1.10 Segmented anatomical structures and the drill trajectory. The drill trajectory passes through the facial recess, which is covered by the facial nerve, the chorda tympani and the external auditory canal, and reaches the round window. At least $0.3mm$ is required between the drill trajectory and the facial nerve. Source: [45] 14

1.11 The surgical planning software iPlan 2.6 for cochlear implant surgery. Left: the position of the registration fuduial marker in the image space. Right: segmented anatomical structures and the drill trajectory. Source: [91] . . . 16

1.12 Surgical planning of drilling trajectory and anatomical structure of the left ear. Source: [110] 16

2.1 One example of facial nerve segmentation with a supervised learning classification approach. From left to right: an original CBCT image, a segmented facial nerve highlighted in green, 3D view of the segmented facial nerve. Details in Chapter 5. 23

2.2 Classification forest testing. The same test input data $\mathbf{v}$ passes through each single tree in the forest, until it reaches the leaf node, which stores the posterior $p_t(c|\mathbf{v})$. The forest class prediction is obtained as the average of all tree posteriors. Source: Criminisi et al.[29] 30

2.3 The aim of image registration is to find a transform that maps points from the fixed image to points in the moving image. 34

2.4 The Registration framework. The basic components of the registration are two input images (fixed image and moving image), a transform, a metric, an interpolator and an optimizer. Modified from: [72] 34

3.1 Example of head motion artifacts in CT image. It appears as shadowed and streaked. This type of motion is impossible to avoid in most cases, because the patient can not keep still during the CT scanning process. Source: [6] . 36

3.2 Motion comparison in CBCT for surgical planning of cochlear implantation. Compared with the no motion CBCT, the simulated motion CBCT image is blurred. Furthermore, the anatomical structures could not be recognized easily. It might lead to high risk for surgical trauma. Modified from: [136] and [94] 36

3.3 Sketch of the correlation between the quantity of motion and the fiducial localization error (FLE). The aim is to find the threshold that can evaluate if the blurred CBCT imaging is suitable for the surgical planning. 38

3.4 The experimental pipeline of CBCT movement detection. Center lines of rod phantoms are extracted, which are regarded as the input data for registration. The center line of the rod-phantom in the motion image is defined as the fixed image, and the center line of the rod-phantom in the no-motion image is considered as moving image. Next, the Hausdorff distance is employed to compute the maximum surface distance between the surface of rod phantoms with and without motion. The simulated head motion is quantified via Hausdorff distance calculation. 39

3.5 Simulated motion scanning system. From left to right: Planmeca 3D CBCT max imaging system and a phantom with the drilling robot, zoomed area for the scanned block, three tiny rod-phantoms attached on the scanned block. . 40

3.6 One example of motion types at 1.25$degree$. Left: sudden motion every 5 minutes to the left or the right, and return to the original position, then towards the opposite direction. Right: continuous motion in the same back-and-forth routine as the sudden motion. 41

3.7 The motion patterns we have scanned. The no motion pattern image is the least blurred image. The outline of the rod-phantom can be easily recognized. When the rotation increases, the length of the rod-phantom appears longer. . 42

3.8 One example of the scanned phantom block. From left to right: the scanned block within two rod-phantoms (A and B), zoomed area showing the phantom block with motion artifacts, and zoomed area showing the phantom block without motion. 42

3.9 Sketch of fitting the center line of the rod-phantom for registration. During the registration, the moving point set of the center line of the rod-phantom from the "no-motion" image aligns to the fixed point set from the "motion" image. 44

3.10 The phantom block used to simulate controlled motion patterns. The amount of motion can be calculated analytically as the dimensions of the rods and the magnitude of the motion are known. 46

3.11 One example of sudden motion of the Phantom A. "Illusory" rod-phantoms appear on the CBCT in the "0.75*degree*", "1.25*degree*", "2.50*degree*" sudden motion images, respectively. The distances among illusory rod-phantoms in the "0.75*degree*" sudden motion images are larger than the "2.50*degree*" one. The smallest distances are in the "1.25*degree*" sudden motion image. . 47

3.12 One example of sudden motion of the Phantom B. "no-motion" pattern of Phantom B is easily distinguished. Different degrees sudden motion make phantom B blurred in various levels. 48

3.13 One example of continuous motion of the Phantom A. "no-motion" pattern of Phantom A is the brightest among the rest of continuous motion pattern images. The continuous motion results in the uneven distributed intensities of Phantom A. 49

3.14 One example of continuous motion of the Phantom B. "no-motion" of Phantom B has the sharpest boundary among the continous motion patterns of the Phantom B. The larger motion rotation leads to the longer length of the Phantom B in the CBCT image. 50

3.15 Correlation between ground truth (computed analytically) and estimated motion using the proposed image-based pipeline. First experiments on two phantom experiments show a good correlation, indicating the feasibility of the proposed approach to estimate head motion. 51

4.1 The complete pipeline of the proposed approach for enhancing CBCT image. During training, the original CBCT image is aligned to its related micro-CT image. Features are extracted from the CBCT by image patches, and intensities from related image patches from the micro-CT are provide. The mapping from image features to intensities is learned during the training phase. 59

4.2 Results of supervised-learning based CBCT image enhancement. Image features are extracted from (a) original CBCT image, and used to produce an enhanced version (d), that presents sharper and more clear structures, as compared to the original high-resolution micro-CT image (b). For demonstration purposes, we report results obtained using features extracted from a short range (i.e. small window size) (c), indicating the ability of the model to learn and utilize local structural information for the prediction process. . 60

4.3 Panoramic view for semi-automatic segmentation of the facial nerve. The user selects a set of landmarks that approximately correspond to the middle line of the facial nerve. A threshold based scheme is then used to cast sampling perpendicular in order to find the facial nerve wall (above and below the middle line). Due to the low contrast quality of the CBCT image, manual correction of each point is commonly required. In this example, we illustrate the enhanced CBCT image. 62

4.4 Surface-to-surface distances from the ground-truth segmentation to OtoPlan segmentations generated using the original and enhanced CBCT image. Colormap encodes distances and best viewed in electronic version. 63

5.1 Proposed super-resolution classification (SRC) approach, described for the training (a), and testing phase (b). During training, the original CBCT/CT image is aligned to its corresponding micro-CT image. OtoPlan [45] is used to create an initial segmentation, from where a region-of-interest (ROI) band is created. From the original and upsampled CBCT/CT images, features are extracted from the ROI-band to build a classification model, which is used during testing to produce a final super-resolution segmented image. The zoomed square on the segmented super-resolution image shows on one voxel of the CBCT/CT image, the more accurate segmentation yielded by SRC. . 72

5.2 Evaluation on CBCT cases between the proposed super-resolution segmentation method and GeoS (version 2.3.6) and ITK-SNAP (version 3.4.0). Average values for DSC (a), Hausdorff (b), Positive Predictive Value (d), and Sensitivity (e). Case number 7 could not be segmented via GeoS. Note: best seen in colors. 77

5.3 Example results for the proposed super-resolution segmentation approach. From left to right: Original CBCT image with highlighted (in blue) facial nerve, resulting segmentation and ground truth delineation (orange contour), and zoomed area describing SRC results on four corresponding CBCT voxels. 78

5.4 The facial nerve segmentation comparison on the original CBCT image between the proposed SRC method and other segmentation software — ITK-SNAP and GeoS. The ROI selection via band 16 from OtoPlan initial segmentation. 78

5.5 Evaluation on CT cases between the proposed super-resolution segmentation method and GeoS (version 2.3.6) and ITK-SNAP (version 3.4.0). Average values for DSC (a), Hausdorff (b), Positive Predictive Value (d), and Sensitivity (e). Case number 8 could not be segmented via GeoS. Note: best seen in colors. 79

List of tables

4.1 List of texture - based features computed at each grid node. 57

5.1 Quantitative comparison on CBCT cases between our method and GeoS and ITK-SNAP, for band sizes 16 (left) and 24 (right). Dice and surface distance errors (in *mm*). The measurements are given as mean ± standard deviation (median). The best performance is indicated in boldface. The '$*$' indicates that SRC results are statistically significantly greater ($p < 0.05$) than GeoS and ITK-SNAP using a two-tailed t-test and Wilcoxon signed ranks tests. . 76

5.2 Quantitative comparison on CT cases between our method and GeoS and ITK-SNAP, for band sizes 16 (left) and 24 (right). Dice and surface distance errors (in *mm*). The measurements are given as mean ± standard deviation (median). The best performance is indicated in boldface. The '$*$' indicates that SRC results are statistically significantly greater ($p < 0.05$) than GeoS and ITK-SNAP using a two-tailed t-test and Wilcoxon signed ranks tests. . 80

7.1 Employed parameters of the ExtraTreesClassifier in sklearn. 96

Chapter 1

Introduction

1.1 Clinical Background

To fully comprehend cochlear implantation, a brief description of the anatomy of the ear and how hearing works is provided below. This section is derived from [97][58][59][60][61][57][158][157].

1.1.1 Anatomy of the Ear

The human ear is the organ of hearing and balance. It contains three main parts: the outer ear, the middle ear and the inner ear (see Figure 1.1). The outer ear captures the sound waves and directs them into the middle ear. The middle ear modifies sound waves into vibrations that pass to the inner ear. The inner ear converts vibrations into nerve impulses, which the brain then interprets.

Outer Ear

The outer ear consists of the pinna and an external auditory canal.

- **Pinna** is also named **auricle**. It is the outside part of the ear, which collects sound and funnels it into the ear canal [97][60][158].

- **External auditory canal** is also called **ear canal**. It has a tube-like shape that joins the outer ear to the middle ear and delivers sound to the middle ear [97][60][158].

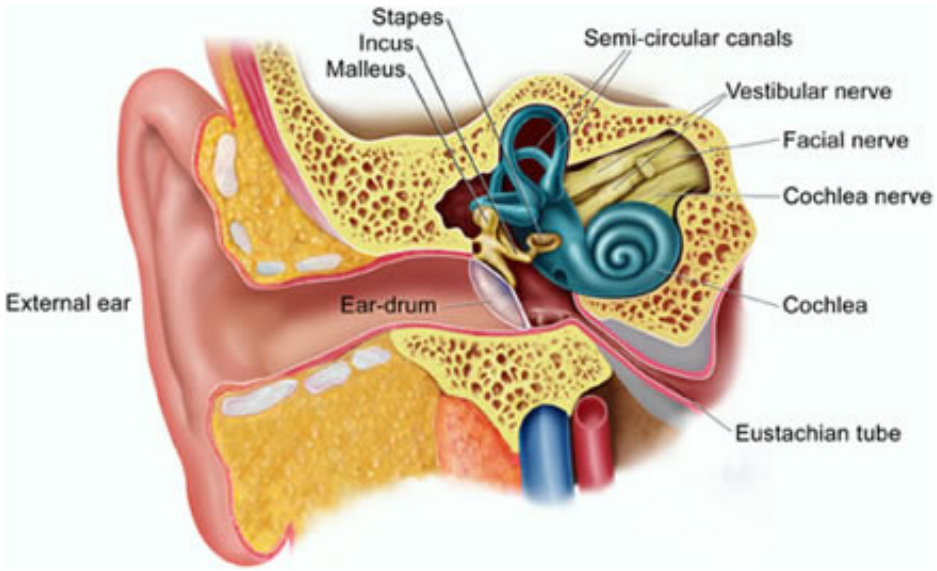


Figure 1.1: Ear anatomy. It includes the outer ear, the middle ear and the inner ear. In the outer ear, there are the pinna and the external auditory canal. In the middle ear, there are the tympanic membrane and the ossicles including the malleus, the incus and the stapes. In the inner ear, there are the cochlea, the vestibular system, the auditory nerve, and the emicircular canals. Source: [96]

Middle Ear

The middle ear is located between the outer ear and the inner ear. It comprises the tympanic membrane and ossicles.

- **Tympanic membrane** is also called **eardrum** and it converts sound into vibrations [97][61][158].
- **Ossicles** comprise a chain of three small bones called **malleus**, **incus** and **stapes** that transmit the vibrations to the inner ear [97][61][158].

Inner Ear

The inner ear includes the cochlea, vestibular system, auditory nerve and the semicircular canals.

- **Cochlea**: a spiral-shaped chamber bone that looks like a snail shell. It contains the hearing nerves, fluid and hair cells. Once the sound vibrations stimulate the hair cells, these hair-like structures move [97][59][158][157].
- **Vestibular system**: It contains cells that are used to control balance [97][59][158].

- **Auditory nerve**: transmits impulses about hearing from the inner ear to the brain [97][59][158].

- **Semicircular canals**: perform a balance function and contain three semicircular tubes which are **the lateral semicircular canal**, **the anterior semicircular canal** and **the inferior semicircular canal** [57][158].

1.1.2 How hearing works

Ears catch sound waves and convert them into signals that the brain can comprehend, such as a song or a conversation. How hearing works are described [98] [42] [99] as follows:

The sound enters the ear canal from the pinna, and makes the ear drum vibrate. The vibrations arrive at the ossicular chain, which contains malleus, incus and stapes, and from there to the inner ear. Next, the vibrations result in the movement of the liquid inside the cochlea, which in turn pushes hair cells inside the cochlea. Then, these hair cells produce electrical impulses. The auditory nerves receive these impulses and send signals to the brain. At the end, the auditory cortex of the brain interprets this information as sounds, for instance, a song or a conversation.

Hair cells, which lie in the whole length of the cochlear, play an important role in the conversion process from the sound vibration to the electrical impulse. These hair cells have various degrees of sensitivity, associated to different frequencies. Therefore, the ear has the ability of recognizing a large range of sounds. Along the whole length of the cochlea, the hair cells are arranged in order, resembling like the keys in a piano. As shown in Figure 1.2, hair cells located at the base, or lower region of the cochlear, are responsible for high frequencies. On the contrary, hair cells at the apex are responsible for low frequencies.

1.1.3 Cochlear Implant

The function of hair cells is to transmit the electrical impulses to the auditory cortex. If the hair cells are impaired (e.g. trauma, congenital defect, etc.), then they cannot produce the necessary impulses to stimulate the auditory nerves.

Hence, in order to recover hearing loss caused by disfunctional hair cells, a cochlear implant (CI) is chosen for the patient who has hearing loss [99]. A cochlear implant is an electrical device, which uses the electrical signal to pulse the auditory nerve, in order to transfer the sound signal to the brain.

The decibel (dB) is an unit for measuring the loudness or the volume of sound [100]. The amount of hearing loss can be classified into 5 different scales from the least to the

Figure 1.2: Illustration of how hearing works within cochlea. Hair cells at the base are responsible for high frequencies. On the contrary, hair cells at the apex are responsible for the low frequencies. Modified from: [24] and [101]

most important loss: minimal, mild, moderate, severe or profound [95][139] [62]. Cochlear implants support people who have a severe-to-profound sensorineural hearing loss (i.e. 70 to 120dB) [62].

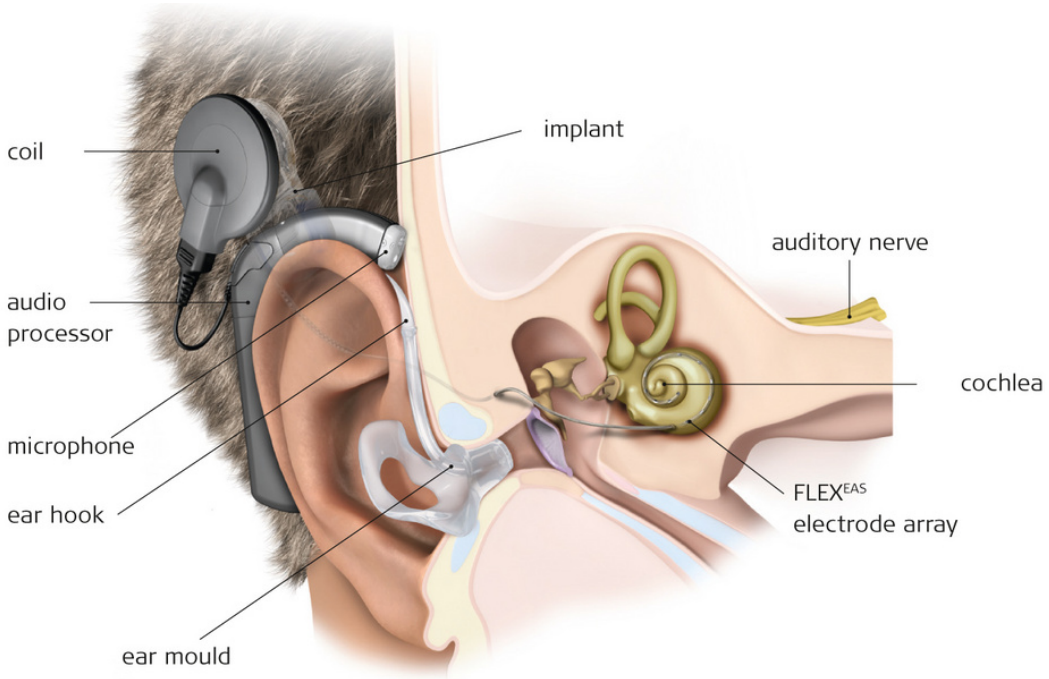


Figure 1.3: Cochlear implant system. The implanted electrode array is used to simulate the function of the hair cell, which pass the electrical impulse to nerve. Source: [68]

1.1.4 Facial Nerve Preservation in Cochlear Implant Surgery

In cochlear implant surgery (see Section 1.3), the facial nerve is given the highest rating in preservation. In order to rescue the facial nerve, which lies in the narrow facial recess, during the CI surgery, the chorda tympani nerve is allowed to be sacrificed. Bhatia et. al [11] summarized 300 consecutive pediatric cochlear implantations and informed that the chorda tympani was destroyed in 20% of these cases (59 children). On the contrary, the aim of our CI surgery is to save the facial nerve and the chorda tympani.

The facial nerve is one out of twelve cranial nerves [159][56][17][102]. It is often abbreviated as cranial nerve (CN) VII [159][56][17]. There are two facial nerves on each side of the face[102]. As shown in Figure 1.4, the facial nerve distributes throughout the face. The facial nerve controls the muscles on the face and manages facial expressions [56]. If the facial nerve is damaged, it will cause temporal or permanent facial paralysis [44]. The facial nerve also delivers the taste-feeling information from the anterior two-thirds of the tongue and the mouth [17] [159][56].

The facial nerve is a thin tubular structure inside the facial nerve canal (see section 1.1.4), which is located in the temporal bone and features a winding path (see Fig. 1.4). The three segments of the facial nerve within the temporal bone, include the labyrinthine, tympanic and mastoid, and are considered in cochlear implantation [52]. The diameter of the facial nerve is in the range of $0.8 - 1.7mm$ [126]. In the second-genu region, the facial nerve makes a turn and goes through between the lateral semicircular canal and the stapes [126]. Its size and shape are different from person to person, especially between adults and children [126]. The facial nerve has a sharper bend in children than in adults [126].

There are three different types of facial nerve fibers [71]. **Motor fibers** control the superficial muscles of the face, neck and scalp, and manage the facial expression of these muscles [71]. **Sensory fibers** pass impulses from taste sensors which are located in the anterior two-thirds of the tongue **Sensory fibers** [71]. They also carry general sensory impulses from tissues close to the tongue [71]. **Parasympathetic fibers** rule the lacrimal glands and certain salivary glands [71].

In order to fully understand the facial nerve, other related medical terminologies are briefly describes hereafter.

Facial Nerve Canal

In the temporal bone, the facial nerve canal (synonym: fallopian canal [71]) stop growing after the baby was born [52]. It is a bony canal wrapped the facial nerve in the temporal bone

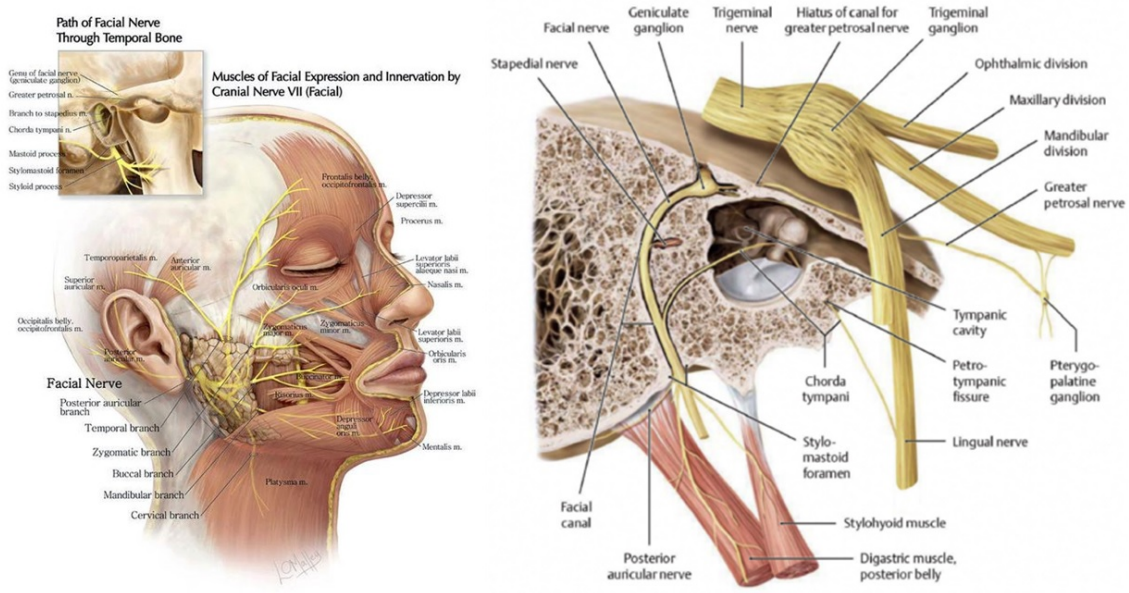


Figure 1.4: Facial Nerve. Left: overview of the facial nerve distribution from one side of the face. Right: Zoomed facial nerve range considered in CI surgery. Source: [122] and [124]

region. And it is divided into three segments: the labyrinthine segment (S1), the tympanic segment (S2), and the mastoid segment (S3) [70] [52][71].

Facial Nerve Dehiscence

Facial nerve dehiscence (FND) is discontinuity in the bony structure wall of the facial nerve canal around the facial nerve [7]. Moreover, it is a common anatomic variant that normally occurs around the oval window in the tympanic segment of the facial nerve [26]. FND is attributed to abnormalities of the ossification of the facial nerve canal, such as cholesteatoma, trauma, and the pressure effect of tumorous lesions [26] [7].

Facial Nerve Disorders

Infection, injury and other conditions can cause facial nerve disorders, which result in paralysis, twitch or weakness on one side or both sides of the face [65]. Those harmful effects may lead to facial expression loss, as well as make eating, drinking and speaking difficult [65]. In addition, closing and blinking the eyes will be effected but will also damage the cornea [65].

1.2 Background on Medical Imaging

Medical images are used for disease diagnosis via computer-assisted approaches nowadays. In this section imaging techniques, used in this thesis work, is discussed. Firstly, CT imaging is introduced, then Cone-beam Computed Tomography and Micro-computed Tomography are described. Finally, the facial nerve images, which are used in the experiments, are described.

1.2.1 Cone-beam Computed Tomography

X-rays were discovered by Whilhelm Conrad Röentgen in 1895. He detected an unknown ray which easily goes through paper and many other materials except a lead plate [149]. X-ray imaging is a transmission-based projection technique, which visualizes the inside of human body without any intervention [1]. X-ray beams from an X-rays source travel through a patient and are collected by the detector (an ionization chamber or film) [1]. When X-ray beams pass through the patient, they are attenuated by the tissues of the body which absorb X-ray beams. Because different tissues absorb various amounts of X-ray beams, which radiographic images show different intensity levels, corresponding different tissues [1]. For instance, calcium contained in bones absorbs most X-rays, soft tissue and fat absorb less, air absorbs the least. Consequently, bone appears white, while soft tissue and fat tissue appears black on the image.

Computed Tomography (CT) uses X-rays to produce images. Tomography comes from the Geek "tomos", which means cut or slice [149]. Tomography tries to produce an image of one or more slices through the body [149]. CT was invented by Godfrey N. Housefield in 1970, which stresses that images are computed from projection measurements [149]. The source beams and detectors rotate together around the patient, producing a large number of projections at different angles of view [1]. CT images provide a fairly sensible contrast between soft tissues and a high spatial resolution [1].

The image quality of any medical imaging device is based on four factors: image contrast, spatial resolution, image noise and artifacts [49]. The higher spatial resolution is, the stronger the ability of distinguishing very small adjoining structures.

Here I enumerate four types of artifact that might happen when the image is created [149][2].

- **Noise**. It is effected from scatter at the detector and in the patient's body. The bright and dark streaks will appear along the direction of greatest attenuation.
- **Partial volume effect (PVE)**. Each voxel in a CT image is described as the attenuation of a specific material volume. If the volume contains more than one different tissue,

the CT value will be an average value of these tissues. Moreover, due to the resolution limitation of CT, the tissue boundary might be blurred.

- **Metal artifacts**. They are the result of the extreme different attenuation between the metal and tissue. Metal streak artifacts appear in the CT image [149].

- **Motion artifacts**. They are caused by conscious or unconscious patient movement such as breathing or the heartbeats during the scanning process.

The serious downside of X-ray based images is the ionizing radiation effect of the technique [1]. Ionizing radiation can cause physical harm to tissue, damage DNA and speed up cell death [1]. There is a radiation dose limitation per year for a patient. Specially, young patients are more sensitive to radiation, as tissues are more easily damaged via radiation exposure [20][75][66].

Cone-beam computed tomography (CBCT), also called digital volume tomography (DVT) [119], provides a three dimensional (3D) image by a cone shaped X-ray beam [4]. When comparing with the conventional CT, there are several advantages of CBCT. CBCT imaging has a lower X-radiation dose, costless, less scanning time and fewer metal artifacts [150][51][28][144][64].

The first CBCT system was designed for the dental and maxillo-facial imaging, especially for implant planning in 1998 [104]. Nowadays, CBCT is widely used in clinical practice, especially in surgical planning, dental applications, skeletal fracture evaluation [32] and post-operative assessment of cochlear implantation [127]. It is highly regarded as an alternative to CT in the head region [37]. For instance, CBCT images are often chosen to plan a cochlear implantation procedure [168][34][25] [123][147][125].

However, there still remains some disadvantages of the CBCT images. Firstly, the CBCT is sensitive to artifacts, noise, and soft tissue contrast [129]. Secondly, the lower X-radiation dose causes the image quality of CBCT to be of lower quality than CT [162]. Hence, CBCT images are typically blurred at the border of anatomical structures.

1.2.2 Micro-computed Tomography

Micro-computed tomography (micro-CT or μCT) generates high resolution 3D images. Micro-CT was developed in early 1990*s* [18] and shows the internal structure images of various materials including bone, tissue, and medical implants at a high resolution at the cost of higher radiation fields than CT and CBCT image [14][106][19][161][78].

Figure 1.5 shows an example of a cochlea image, showing different image qualities from standard CT, high resolution CBCT and non-clinical micro-CT, where the border of the cochlea can be easily recognized.

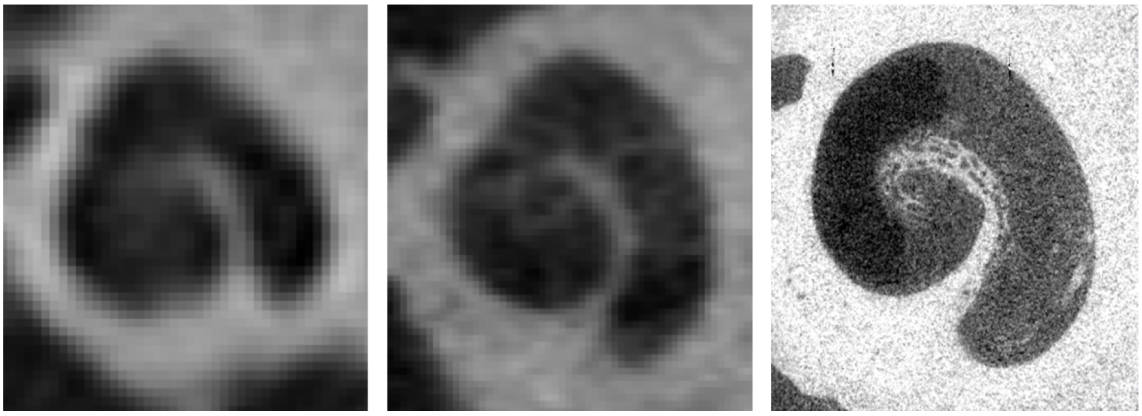


Figure 1.5: Cochlea image obtained from standard CT (left), high resolution CBCT (middle) and non-clinical micro-CT (right). The image quality increases from left to right. The border of the cochlea can be easily recognized on the highest resolution, micro-CT image. Source: [160]

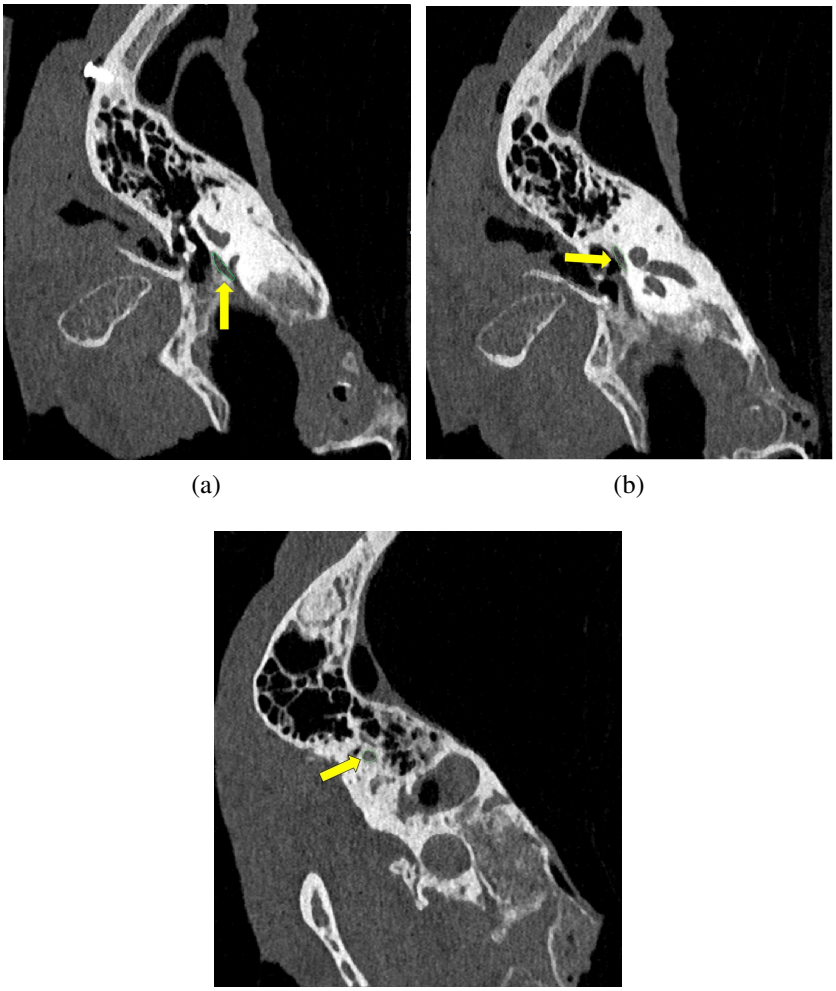
1.2.3 Imaging the Facial Nerve

Imaging plays an important role in providing information about the facial nerve in surgical planning (see Section 1.3.3). The facial nerve has a complicated anatomical shape [52]. Moreover, the facial nerve might have an abnormal pattern because of a congenital defect, infection, inflammation, trauma and cancer [52]. Computed tomography (CT) is used to image the facial nerve that is located inside the bony facial nerve canal of the temporal bone region [52]. In contrast, magnetic resonance imaging (MRI) is used to check the soft tissue and evaluate facial nerve abnormalities in these areas outside the bony facial nerve canal (fallopian canal) (see Section 1.1.4), which could not be seen in the CT image[52].

Although all the above presented techniques can be used to identify and evaluate the facial nerve, the choice of the imaging modality should be based on the goal of the surgical planning and patient's symptoms [52]. In our experiments, pairs of CBCT and micro-CT image, and pairs of CT and micro-CT image are used for the analysis of facial nerve.

Figure 1.6 shows one example of the three segments of the facial nerve canal in a CT image, which are required in the surgical planning for cochlear implant surgery. Figure 1.7 shows one example of a manually segmented facial nerve from micro-CT images, which shows the course of the facial nerve considered in the surgical planning. In fact, a facial

nerve segmentation demonstrates the facial nerve canal segmentation, which wraps the facial nerve.



(a) (b)

(c)

Figure 1.6: One example of the three segments of the facial nerve canal in a CT image. These segments are used when planning a cochlear implantation. (a) The labyrinthine segment (S1), (b) The tympanic segment (S2), and (C) The mastoid segment (S3).

1.3 Minimally Invasive Cochlear Implant Surgery

1.3.1 Conventional Cochlear Implant Surgery

Cochlear implants (CI), for deaf people who obtain artificial hearing, have come a long way since 1960*s* [133] [131]. Nowadays, more than $300,000$ people in the world have received CI surgery [69].

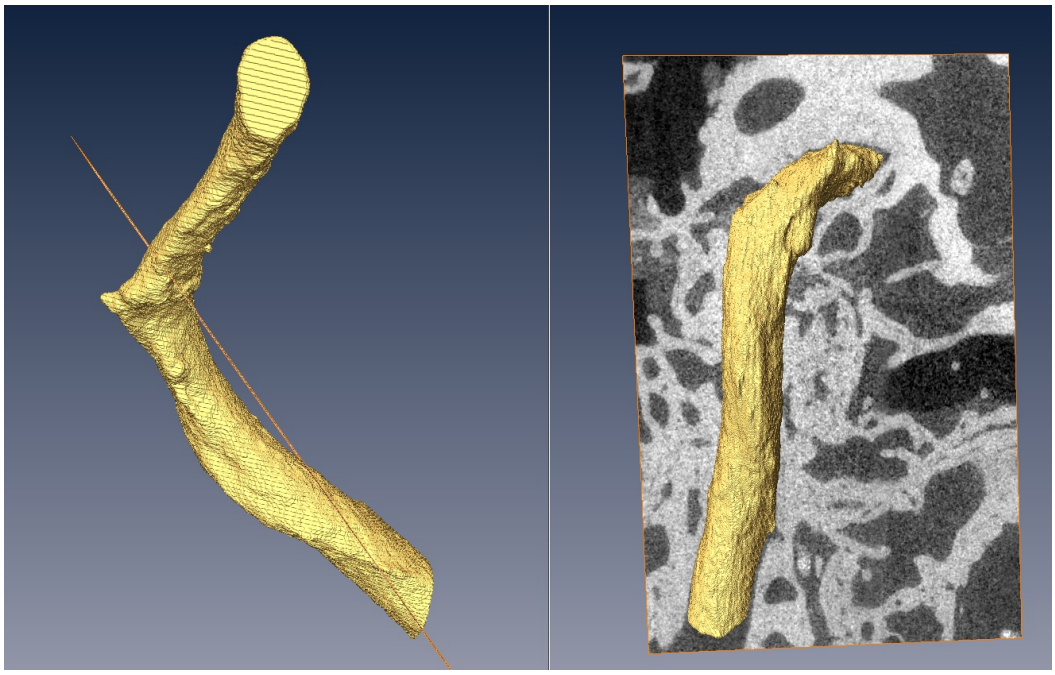


Figure 1.7: One example of manually segmented facial nerve from micro-CT. It shows the course of the facial nerve considered in the surgical planning. In fact, facial nerve segmentation demonstrate the facial nerve canal segmentation, which wraps the facial nerve.

A mastoidectomy represents the current common technology for inner ear access, which removes the mastoid region behind the ear in the skull [154]. During the mastoidectomy, the surgeon must drill a large hole in the mastoid region of the temporal bone between the skull surface and the inner ear, and expose the facial recess, which is an anatomic area covered by the facial nerve, chorda tympani nerve and the incus buttress [82] [154] [153]. The mastoidectomy allows the surgeon to see and protect the vital anatomical structures during the drill process. The main challenge of the surgery is to access the cochlea, which locates at approximately 35mm inside the temporal bone [83]. During the drilling process, the facial nerve, the chorda tympani, the sigmoid sinus, the carotid artery and the labyrinth must be avoided [91]. As described in section 1.1.4, the highest priority is given to the facial nerve [91]. The space between the facial nerve and the chorda tympani is around 2mm [83] [91]. In addition, the carotid artery brings needed blood to the brain and face [91]. It results in patient's death if this part of the carotid artery is damaged [91]. The sigmoid sinus are venous sinuses that receives blood from the posterior aspect of the temporal bone [91]. Puncture sigmoid sinus will lead to a considerable amount of blood loss and haemodynamic disturbance [91]. The labyrinth is responsible for hearing and balance. In order to prevent hearing changes and dizziness or vertigo, the labyrinth should not be damaged as well.

To avoid damaging the facial nerve or any other vital anatomical structure, the surgeon must be totally confident about the drilling position [154]. A mastoidectomy is highly dependent on the surgeon's experience and skills, and is the most important and time-consuming part of the cochlear implant surgery.

1.3.2 Image-Guided Procedures in Otological Surgery

To minimize the invasiveness of the procedure, image guided surgical techniques are increasingly applied in various surgical specialties such as neurosurgery, otolaryngology, spine surgery, and orthopedic surgery. Compared with the traditional open surgery, image-guide surgery has more advantages including reduced surgical risk and surgical time, more visual insight into the human body and reduced invasiveness [121].

Image-Guided Procedures consist of 6 key components: image acquisition, data visualization, segmentation, image-to-physical-space registration, three-dimensional tracking systems and human computer interaction (HCI) [43][163]. In a timeline view, the image-guided procedure contains three phases: pre-operative planning, intra-operative plan execution, and post-operative assessment [163]. The aim of pre-operative planning is to make a surgical plan based on pre-operative medical images and anatomical geometry information [163]. In this phase, the image acquisition requires the following techniques: medical images and image processing and data visualization.

Although image-guided techniques are widely applied in different surgical specialties, a few surgical cases in otology utilize image guidance [80]. This is because the otological surgery requires much higher accuracy than other surgical specialties, such as neurosurgery[44]. The image-guided surgical system makes the successful CI surgery less dependent on the surgeon's surgical experience, which compensates for the loss of visual control during the drill process [131][91]. Moreover, the image-guided surgical system for CI surgery is expected to assist the surgeon to find the optimal drill trajectory and insert the electrode array into the human cochlea. Schipper made an assessment of the image guided surgical procedure on CI surgery on a cadaver head in 2004 [131]. Later, the direct cochlear access (DCA), a minimal invasive method, was developed. More detailed information about DCA is given in Section 1.3.3.

1.3.3 Robotic Cochlear Implant Surgery

Microsurgical procedures require surgeons to perform operations on body structure, smaller than $1mm$ [39]. Minimally invasive cochlear implantation is one delegate example of microsurgical navigation technologies [39].

Direct cochlear access (DCA) is a minimally invasive method for drilling a trajectory to access the inner ear, which differs from the traditional mastoidectomy, as it uses a small-diameter trajectory [154]. The trajectory is within the mastoid and passes through the facial recess and reaches the round window [154]. Considering the diameter of a cochlear implant electrode, which is less than $1.5mm$, the size of the drilling hole is close to the size of the electrode, which is $1.5-2mm$ in tunnel diameter [154]. Hence, it is challenging to reach an acceptable accuracy in DCA.

A surgical robotic image guided system (see Figure1.8) was designed for DCA surgery by the Univerisity of Bern [8][10][154]. In this system, a force torque sensor at the robot wrist can do haptic control, registration and evaluate drilling process [154]. This system needs a highly precise surgical planning to ensure a save minimal invasive cochlear implantation. Fiducial screws markers and critical anatomical structures are segmentation with the associated planning software tool, OtoPlan [44][45]. After all the required structures are segmented, the surgeon can plan a drill trajectory. As shown in Figure 1.10, the minimal distance of $0.3mm$ is required between the drill path and the critical structure facial nerve.

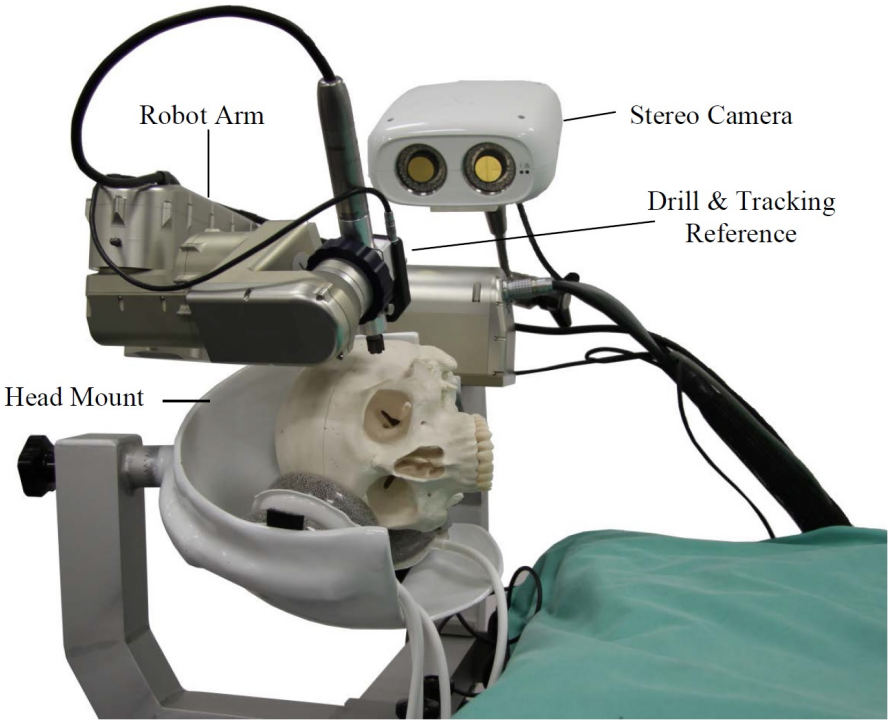


Figure 1.8: A minimally invasive robotic image guided system for cochlear implant surgery. The surgical planning guides the surgeon to do cochlear implant surgery. The robot drills a small hole in the mastoid. During the drill process, the camera tracks the drill position. Source: [10]

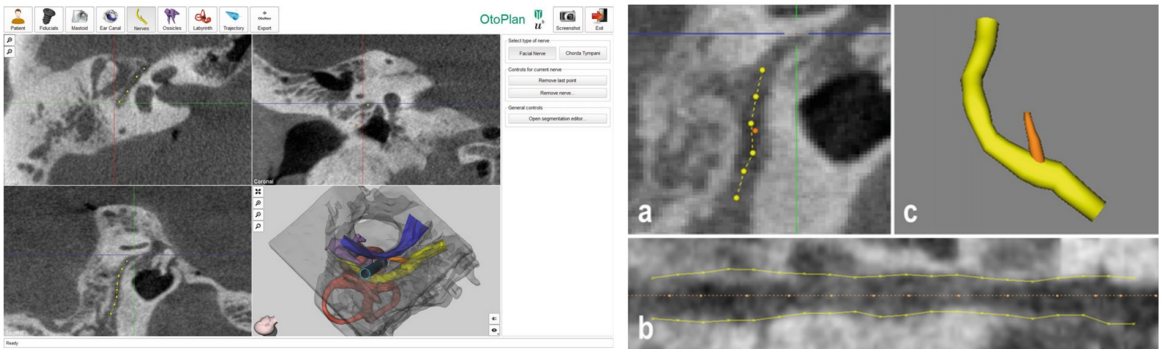


Figure 1.9: The personalize surgical planning software OtoPlan for minimal invasive cochlear implant surgery developed from Bern University, Switzerland. Left: User interface of OtoPlan software. Right: Semi-automatic segmentation for the facial nerve and chorda tympani. Source: [45]

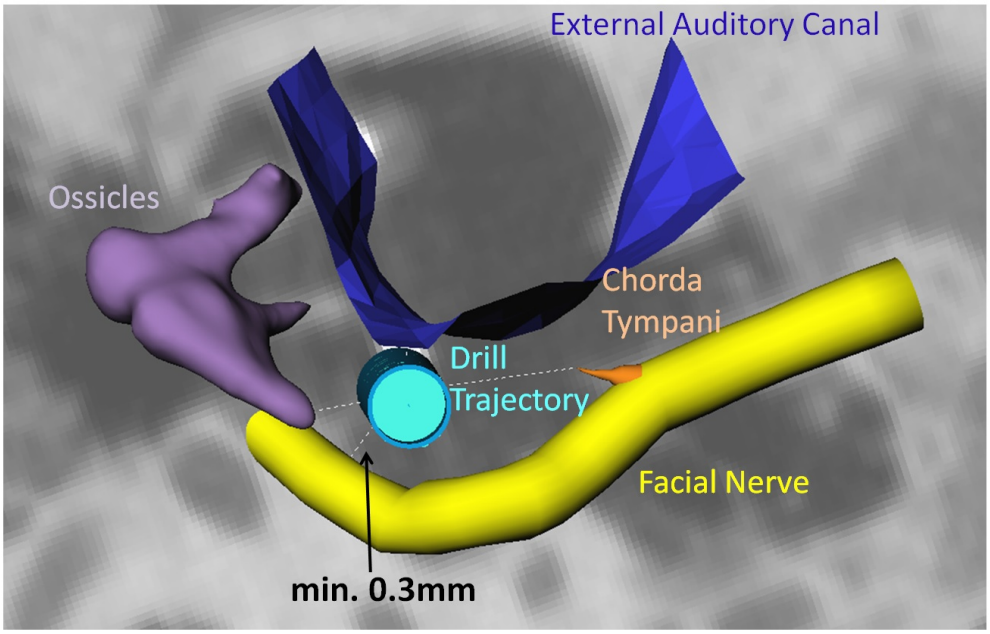


Figure 1.10: Segmented anatomical structures and the drill trajectory. The drill trajectory passes through the facial recess, which is covered by the facial nerve, the chorda tympani and the external auditory canal, and reaches the round window. At least 0.3*mm* is required between the drill trajectory and the facial nerve. Source: [45]

Surgical Planning

Surgical planning uses patient specific medical imaging and computer programs to visualize the anatomical structures of interest to yield a surgical plan [44]. In addition, in robotic image guided system, the surgical planning is used for controlling the robotic equipment, which is loaded in the robotic system [44].

Surgical planning involves the segmentation of critical anatomical structures as well as the identification of vulnerable structures [44]. Various image modalities are employed in surgical planning, such as a pre-operative CT or CBCT images providing the geometric information among the surrounding tissues and structures.

In surgical planning, segmentation plays a critical important. Segmentation is defined as the boundary delineation of anatomical structures within the medical image data. There are different segmentation methods, such as manual segmentation, semi-automatic segmentation and automatic segmentation. Manual segmentation requires the experts, such as surgeon, radiologist, to mark the borders of the anatomical structure on the images slice by slice. This method is a time-consuming and error-prone process. For this reason, it is necessary to develop semi-automatic or automatic segmentation methods for surgical planning.

In [91], the surgical planning software iPlan 2.6 BrainLAB AG, Feldkirchen, Germany) was proposed, which defines registration fiducial markers and segments essential anatomical structures. The facial nerve, chorda tympani and sigmoid sinus are manually segmented, while the cochlea, semicircular canals and the middle-ear ossicles are segmented semi-automatically. According to the 3D segmented anatomical models, the planning of the drill trajectory is automatically decided via a software.

The automatic segmentation of anatomical structures in a temporal bone CT scan is demonstrated in [109] for cochlear implant surgery. The labyrinth, ossicles, and the external auditory canal are automatically segmented based on atlas-based registration techniques. The facial nerve and the chorda tympani are segmented via a navigated optimal medial axis and a deformable-model algorithm. Furthermore, the study in [110] shows a method of designing a drill path for percutaneous cochlear implant surgery, which calculates an optimal safety drill patch automatically (see Figure 1.12). Figure 1.12 shows the spatial orientation between the drill path and ear structures.

Moreover, atlas-based approaches combined with level-set segmentation have been proposed to segment the facial nerve and chorda tympani in CT images in adults [108] and pediatric patients [126] for percutaneous cochlear implant surgery. In [126], the study stressed that the anatomical structures of the children and adults are different, and hence a statistical model created from adult data could not handle pediatric cases.

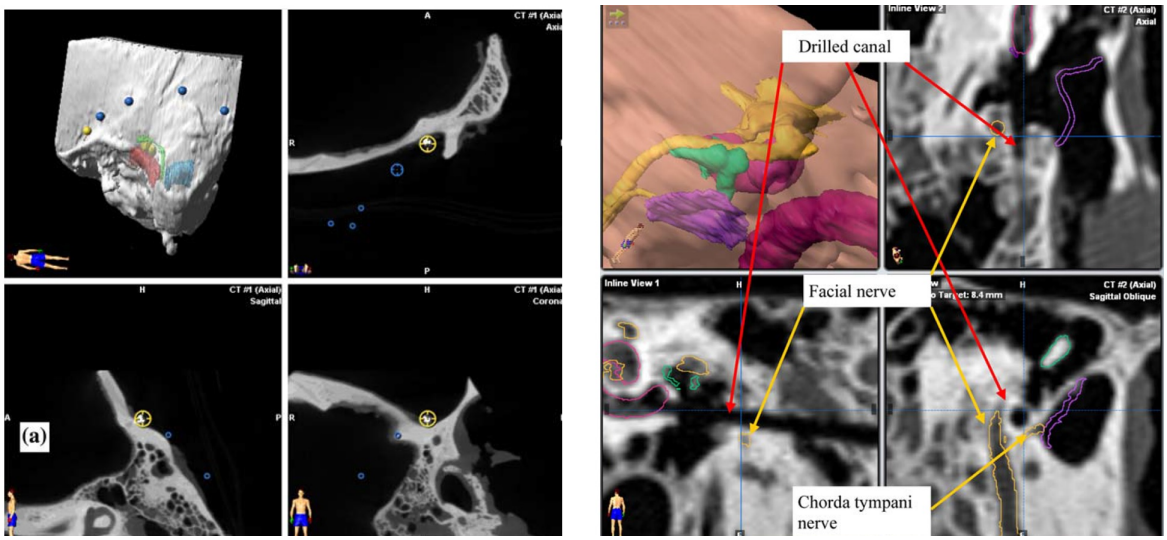


Figure 1.11: The surgical planning software iPlan 2.6 for cochlear implant surgery. Left: the position of the registration fuduial marker in the image space. Right: segmented anatomical structures and the drill trajectory. Source: [91]

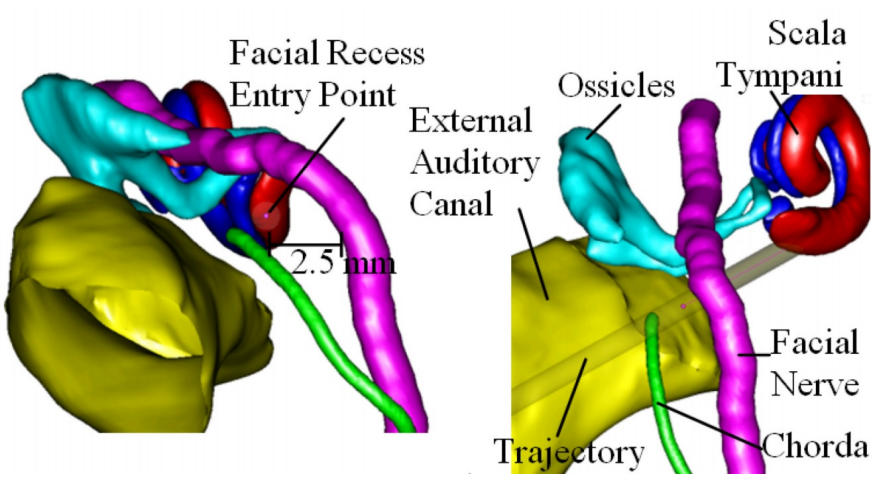


Figure 1.12: Surgical planning of drilling trajectory and anatomical structure of the left ear. Source: [110]

There is great potential in the surgical planning, which is an important part of a successful treatment. Until now, many surgical planning software were able to solve several specific clinic tasks. To date, there is no commercial image guided system incorporate surgical planning for a complete cochlear implant surgery, due to the limitation of the required accuracy in planning and navigation system [44]. Cochlear implant surgery requires submillimeter precision because very tiny anatomical structures, such as facial nerve and chorda tympani, should not be hurt.

1.4 Challenges

1.4.1 Imaging Data

Considering the patient's safety, the facial nerve imaging is performed by low dose CBCT or CT imaging. Nevertheless, CBCT and CT imaging cannot image the facial nerve very clearly, which influences the surgical planning. There are two main disadvantages of CBCT and CT imaging. One is the partial volume effects (PVE) and the tiny facial nerve and the surrounding structures. The other one is that the structures surrounding the facial nerve have similar intensity values. Furthermore, due to their low resolution, CT/CBCT image cannot recognize temporal bone fractures, which exists in some cases. This fracture leads to the facial nerve canal and the bone air cell being visualized together.

1.4.2 Manual Segmentation

The segmentation of the facial nerve plays the most critical role in surgical planning. In order to obtain an accurate segmentation, the expert manually segments the facial nerve and verifies the segmented facial nerve. Manual segmentation is a time consuming, tiring and tedious task. Moreover, different experts have different segmentation results based on the same data (inter-observer variability). Furthermore, even the same expert might give different segmentation results if the expert segments the same data several times (intra-observer variability). On the other hand, semi- or fully automated segmentation are not precise enough.

1.4.3 Patient Movement

Patient could not keep still during the CBCT scanning process, especially children. Patient head movement is the main reason for blurred images, which might happen during scanning. The blurred CBCT image may lead to suboptimal surgical planning, which leads to the cochlear trauma and damage any critical structure.

1.5 Thesis Hypothesis, Objective and Contributions

1.5.1 Hypothesis

We hypothesize that an advanced CT based image analysis of the ear can lead to a highly accurate facial nerve segmentation for micro-IGS cochlear implantation.

1.5.2 Objective

The objective of this thesis was to develop approaches for motion detection and accurate image segmentation of the facial nerve, which can improve the performance of the surgical planning.

1.5.3 Contributions

There are three main contributions of the thesis, which are organized as follows:

- **Motion Detection**. We propose a practical method for motion detection in medical scanning image can be accepted in the surgical planning. We register the center-lines of the phantom data with motion and without motion, and then we calculate the Hausdorff distance which measured the maximum distance between phantoms' surface. Next, the Geometric distance and the Hausdorff distance are compared. The method and the preliminary evaluation are explained in Chapter 3.

- **Facial Nerve Image Enhancement**. We present a supervised-learning approach to enhance facial nerve imaging information from Cone-beam Computed Tomography (CBCT). A multi-class random forest is employed to learn the mapping between pairs of image patches from CBCT and micro-CT images of the facial nerve. The mapping is performed by learning the relationship between texture features and the intensities. The method and its performance are reported in Chapter 4.

- **Facial Nerve Segmentation**. We propose a super-resolution classification method, to refine the facial nerve segmentation to a sub-voxel classification level from CBCT images. The super-resolution classification method learns the mapping from low-resolution CBCT images to high-resolution facial nerve label images, obtained from manual segmentation on micro-CT images. The mapping is performed by learning the relationship between imaging features, such as first order statistic, percentiles, texture, and label information, such as facial nerve and background. The method and the experimental design are explained in Chapter 5.

1.6 Outline of the Thesis

In Chapter 2, the theoretical background of machine learning is introduced. In our work, the image enhancement and the image segmentation are based on one type of machine learning method called supervised learning. The detail of the supervised learning method is given in this chapter. The three main contributions are then presented, and divided into three chapters: In Chapter 3, the method of motion detection for the scanning image, which is accepted for the surgical planning, is presented. In Chapter 4, a supervised learning method is given, which deals with facial nerve image enhancement. In Chapter 5, a supervised learning method, referred to super-resolution classification method, is presented for facial nerve segmentation, which reconstructs the high resolution of facial nerve image from the low resolution image. In Chapter 6, the general results and the outcomes of the contribution work are summarized. Furthermore, the potential of future work is discussed.

Chapter 2

Technical Background

2.1 Introduction

This chapter summarizes the technical concepts used in the following chapters. The main technique used in this thesis is machine learning applied to facial nerve image enhancement (Chapter 4) and facial nerve image classification (Chapter 5). This chapter will firstly introduce machine learning in general, and then concentrate on the specific technique used in Chapter 5 – extremely randomized trees. Since chapters 3, 4 and 5 require registration as preprocesing, registration will be introduced at the end of this chapter.

2.2 Machine Learning

Machine learning is an essential branch of Artificial Intelligence (AI), which provides the algorithms that mimic the ability of humans to learn from previous experiences and deal with new observations [130]. The key factor of machine learning is its automation, which is required to design a learning algorithm that automatically completes a learning process without any human intervention [130]. Instead of solving a task directly via the computer program, a given machine learning method is designed to solve a task via producing its own program based on previously provided examples [130].

2.2.1 Machine Learning in Medical Imaging

Machine Learning (ML) has seen an explosion of interest in many various fields such as weather forecast, detection of e-mail spam, genetics, computer vision, stock market analysis, language processing and search engines as Google [156]. In the last decade, machine learning has been applied to the medical imaging field and plays an important role in medical image

analysis, computer aided diagnosis, image registration, image fusion, image-guided therapy, medical image categorization and retrieval [90][145][81]. Medical imaging applications of machine learning are very similar to the traditional works in computer vision. Examples of these include object detection, object segmentation and object tracking in computer vision, and anatomical structure detection, segmentation and tracking in medical imaging [31]. Furthermore, machine learning applied in medical imaging follows a growing trend [118]. In addition, workshops on machine learning in medical imaging appear on medical imaging conferences such as Medical Image Computing and Computer Assisted Intervention (MICCAI) [118].

Applying machine learning techniques in medical imaging is very challenging [31][118]. The reasons for this challenge include [31][118],

1. **Medical images are very large, and often they have three or more dimensions.** Medical image may have various resolutions and a certain amount of noise stemming from different imaging systems [118]. For this reason, machine learning approaches are expected to be capable of handling large data sets and noise.

2. **Machine learning algorithms need lots of training data to tune parameters of the predictive model avoiding overfitting or overgeneralization.** The number of medical images is often limited, as it is a long process to obtain medical images [31]. This curation step includes feature selection, anonymization, data annotation, etc. Hence, building sufficiently large training data sets is a difficult task.

3. **Training data is difficult to obtain.** Most machine learning tasks in medical imaging are based on supervised learning [118]. It needs experts to annotate medical data for building a reliable ground truth. However, producing ground truth data is tiring and tedious, and requires experts' experience in the anatomical structures of interest. Building ground truth data via multiple experts can lower inter-observer effects. A high intra- or inter-observer variability indicates that the cognition of anatomical data is not clear enough, or cannot easily be employed [148]. Even if the same expert works on the same anatomical data, he or she might have different judgments at different time intervals [15]. Therefore, generating high quality ground truth data, typically considers multiple judgments of each data, followed by their combination [118]. An ideal result of a medical image analysis method should be close to a result on which experts agree, and which is within the acceptable variation range among experts.

4. **It requires highly reliable learning methods.** As the target are patients, machine learning approaches in medical image analysis need to feature high accuracy and robustness.

Machine learning contains a broad range of techniques which are very helpful in medical applications, including classification, regression, clustering and dimensionality reduction [16]. For instance, automatic disease detection and anatomical structure segmentation can be used in any type of images for surgical planning.

Medical Image Segmentation

Medical image segmentation is a popular topic that can be handled with machine learning-based approaches. As explained in Figure 2.1, the facial nerve is segmented from a CBCT image with a supervised learning method. This method learns the mapping from extracted features in the CBCT image to labels in the manually segmented facial nerve image. Many other examples can be mentioned. For instance, in [48], a supervised learning method of joint classification-regression is proposed to handle multi-organ segmentation in 3D medical CT scans. In [93], extremely randomized trees is employed as a supervised method for biomedical image classification. In [152], random forest is applied to the 2D segmentation of placenta from fetal MRI, and lung segmentation from radiographic images.

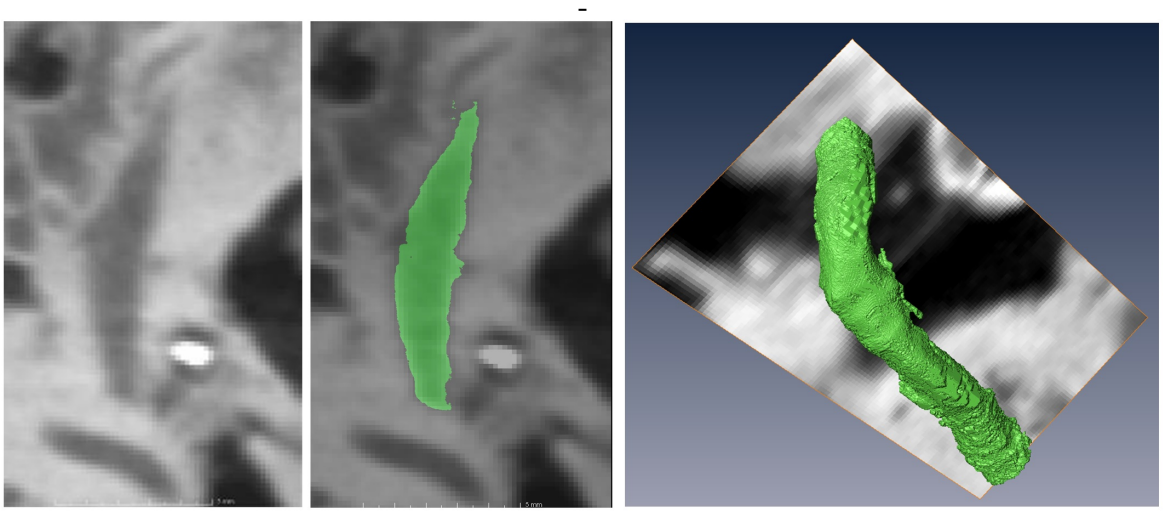


Figure 2.1: One example of facial nerve segmentation with a supervised learning classification approach. From left to right: an original CBCT image, a segmented facial nerve highlighted in green, 3D view of the segmented facial nerve. Details in Chapter 5.

2.2.2 Types of Machine Learning

Machine learning can be divided into three categories: supervised learning, unsupervised learning and semi-supervised learning [21]. The difference among these three methods lies in the training data. In supervised learning, the training data includes the input data along

with the output labeled data. In unsupervised learning, the training data includes the input data without any corresponding output data. In semi-supervised learning, the training data contains a large number of input data but only some of the input data has output labeled data [21].

In the presented work, supervised learning method was used to enhance and segment a medical image for the facial nerve. Hence, the supervised learning method will be introduced in the following paragraph.

Supervised Learning

Supervised learning is a type of learning in which input and output data, called labeled data is available, and an algorithm is used to learn the mapping from input data to output labeled data [21]. Once the mapping has been learned, the model can be used on new unseen data.

Why is it called supervised learning? The mapping process can be regarded as a process in which a teacher is in charge of the students' learning process and is responsible for making sure the students learn properly [21]. Given the correct answers (the related labeled data), the algorithm makes predictions on the training data and it is corrected via the teacher's instructions. The learning process will stop when the algorithm achieves a satisfactory performance [21].

Supervised learning can be divided into classification and regression.

- **Classification** If the output data is a category, such as "agree" or "disagree", the task is called classification.
- **Regression** When the output data is one or more continuous variables, such as "height=1.60cm" or "weight=45.33kg", the task is called regression.

Most of the practical applications are based on supervised learning. Popular supervised learning algorithms include random forest and support vector machines. In order to build a supervised learning algorithm, the following components should be considered: building a dataset, defining a model, defining a objective function, and defining an optimization procedure [40] (see Section 2.3). In the following, random forest is described.

2.3 Random Forest

Random forest is a machine learning approach, which is widely used in the medical field. The main concepts to understand this approach are decision trees, randomization and ensembling. Therefore, in this section, the basic definitions and concepts of the tree will be introduced.

Next, the split function of the decision tree will be described. At last, the way of combining decision trees into a forest ensemble will be described.

This section is based from the book of Criminisi et al.[29]. There is a large amount of literature related to random forest. As a result, some concepts have different definitions. For instance, **random forest** is also known as **decision forest**, **randomized forest**, or **randomized decision forest**. **Decision tree** is also called **randomized tree**, **randomized decision tree**. **Split node** is also named **internal node**, **decision node**, or **branch node**. **Lead node** is also called **terminal node**.

2.3.1 Basic Definitions

Tree Data Structure

Binary trees have one internal node with two outgoing edges. There are two types of nodes in a binary tree: internal nodes (also called as **split nodes**) and terminal nodes (also called **leaf nodes**). Each internal node has a split function, which tests the incoming data. Each terminal leaf leads to the final prediction.

Data Point and Features

Considering a **data point** as a vector $\mathbf{v} = (x_1, x_2, ..., x_d) \in \mathbb{R}^d$, each component x_i holds an the attribute of the data point, which is named as **feature**. The number of features depends on the attribute of the data point (e.g. voxel intensity, patient's age, etc.) and the application. While in theory, the dimension d of the feature space can be very large, even infinite, in practice, a smaller feature vector is preferred to comply with computational limitations such as memory and speed, and to avoid the curse of dimensionality [50] [74]. A subset of features is represented as $\phi(\mathbf{v}) = (x_{\phi_1}, x_{\phi_2}, ..., x_{\phi_{d'}}) \in \mathbb{R}^{d'}$, where $d^{'}$ stands for the dimension of the subspace, and $\phi_i \in [1, d]$ stands for the extracted dimensions. Here, the dimension $d^{'}$ of the subspace is typically much smaller than $d (d^{'} << d)$.

Split Functions

The split function is also called as **test function** or **weak learner**. Each node stores a different split function. The data point $\mathbf{v}$ arrives at the split node and then it goes to its left or right child node depending on the result of the split function. The split function at a split node j, has a function with binary outputs, denoted as

$$h(\mathbf{v}, \theta_j) : \mathbb{R}^d \times \mathscr{T} \rightarrow \{0, 1\}, \tag{2.1}$$

where 0 and 1 could be understood as "yes" and "no" respectively, θ_j describes the split parameters related with the jth node, and $\mathscr{T}$ means the space of all split parameters.

Training Points and Training Sets

A training point is a data point. Its attributes are already known and used to calculate the tree parameters. In supervised learning, a training point is a pair $(\mathbf{v}, \mathbf{y})$, where $\mathbf{v}$ is the input data point (feature vector), and $\mathbf{y}$ is a known label. [1]

Regarding training sets, they are seen as subsets of training points related with different tree branches. For example, S_j represents the subset of the training points arriving at the node j, and S_j^L , S_j^R correspond to the subset going to the left and the right child node of the node j, respectively.

2.3.2 Decision Tree

Training Phase

The split functions in the internal nodes are learned from the training data, which are very important for the correct functioning of the tree. Therefore, the training phase focuses on choosing the parameters of the split functions $h(\mathbf{v}, \theta)$ for each internal node via optimizing the **objective function**, which defines a training set. The objective function is also called **loss function** or **cost function**.

At each node j, the optimal split function splits S_j into S_j^L and S_j^R. This split function can be described as the maximization of an objective function I at the jth split node

$$\theta_j = \arg\max_{\theta \in \mathscr{T}} I(S_j, \theta). \tag{2.2}$$

The objective function can be defined as an information gain. Training starts at the root node, $j = 0$. Then, two child nodes receive a different subset of the training set. This process works on all the newly built internal nodes. Furthermore, the training phase will stop when the stopping criterion is met.

Testing Phase

During testing phase, an internal node applies its related split function $h(.,.)$ to the input data $\mathbf{v}$, and sends it to the left or right child node. This process starts at the root, and it is repeated

[1] In contrast, in an unsupervised learning method, the training point only contains the feature vector, which does not have a related label.

until the data point **v** reaches a leaf node. The leaf node has a predictor, which produces an output **y** for the input data **v**.

In the following, details on the energy model, leaf prediction model, and randomness in decision tree are provided.

Energy Model

Usually the information gain is chosen as the objective function. During training phase, the objective function is used for building decision trees. Moreover, the optimization of the objective function influences the parameters of the split functions, which decides the path of a data point and its prediction. In sum, the energy model affects on the prediction ability of a decision tree.

Information gain is related to a tree split node and describes the reduction in uncertainty, which is represented by splitting the training data into different child nodes. The information gain is defined as

$$I = H(S) - \sum_{i \in \{L,R\}} \frac{|S^i|}{|S|} H(S^i), \tag{2.3}$$

where the dataset S is split into two subsets S^L and S^R (binary trees), and $\frac{|S^i|}{|S|}$ stands for the weight of the i^{th} partition. In (2.3), H is the entropy, which measures the uncertainty related to the random variable that is expected to predict.

Entropy is a common way to tell the level of impurity in a data group. For discrete probabilities, the Shannon entropy is defined as,

$$H(S) = -\sum_{c \in C} p(c) \log(p(c)). \tag{2.4}$$

Here S is the set of training sets, c corresponds to the class label, belonging to C, and $p(c)$ corresponds to the probability of class c from the training points.

The aim of choosing the split parameters is to maximize the information gain, which produces the lowest uncertainty in the final distribution. After splitting, the children distributions become purer, their entropy decreases and their information content increases. Hence, to build a tree, the split function reduces uncertainty via maximizing the information gain and minimizing the entropy.

Along with information gain, there are some popular objective functions such as gain ratio and gini index. **Gini index** is a way to measure the impurity in a data group. It is defined as,

$$Gini(S) = 1 - \sum_{c \in C} p(c)^2. \tag{2.5}$$

The reduction in impurity is defined as

$$\Delta Gini = Gini(S) - \sum_{i \in \{L,R\}} \frac{|S^i|}{|S|} Gini(S^i). \tag{2.6}$$

Here the selected split parameter should maximize the reduction in impurity and minimum Gini index. In our work, the extra-trees classifier in Chapter 5 chooses gini as a criterion to measure the split quality for building trees.

Leaf Prediction Model

The training phase not only estimates the optimal split function in the internal nodes, but also learns good prediction models which are to be stored at the leaf nodes. In supervised learning, each leaf node contains a subset of training data. Through the split function, the input test data will reach a leaf node related to training points which are very similar to itself.

The leaf statistics can be presented as the conditional distributions $p(c|\mathbf{v})$ or $p(\mathbf{y}|\mathbf{v})$, where c corresponds to categorical labels, and $\mathbf{y}$ corresponds to continuous labels (i.e. regression), $\mathbf{v}$ is the test data point, and the conditioning corresponds to the probability distributions of each label class. Different leaf predictors can be applied. For example, a Maximum A-Posteriori (MAP) estimate is chosen as,

$$c^* = \arg\max_c p(c|\mathbf{v}) \qquad \text{or} \qquad p^* = \arg\max_p p(\mathbf{y}|\mathbf{v}) \tag{2.7}$$

for the categorical and continuous case, respectively.

Randomness

There are two main methods to add randomness into trees during the training phase, bagging and randomized node optimization [29].

- **Bagging** In bagging, each single tree in a forest is trained on a various training subset, which are sampled randomly from the same dataset. There are two advantages from this strategy. First, it helps to avoid specialization in parameters selection and improves generalization. Second, the training is faster than using the whole training data.

- **Randomized Node Optimization**

  At each node, the optimization (2.2) works on the whole parameter space τ. The size of τ might be extremely large, if the data has large dimensions. Hence, optimizing (2.2) over the whole parameter space τ is intractable. In practice, only a small random

subset $\tau_j \in \tau$ of parameter values are selected for training of the jth internal node. Therefore, training a tree results in optimizing each internal j node as,

$$\theta_j = \arg\max_{\theta \in \mathscr{T}_j} I(S_j, \theta). \tag{2.8}$$

A parameter $\rho = \mathscr{T}_j$ is defined. The parameter $\rho \in \{1, ..., |\mathscr{T}|\}$ controls the randomness degree ρ in a tree. Moreover, the value of randomness degree is fixed for all nodes. If $\rho = |\mathscr{T}|$, all the internal nodes use all the information from the data point. Hence, there is no randomness in the training phrase. If $\rho = 1$, each internal node only has a single value in the parameter θ. Hence, there is no optimization at all, and a maximum randomness is employed.

2.3.3 Forest Ensemble

Ensemble methods are learning models that combine several models' prediction results (instead of a single prediction) to get better predictive performance. Like a human, a better decision comes from a group opinion instead of a personal one. Furthermore, ensembles are parallel and independent, which can be more efficient during the training and testing phase. There are many different methods to combine basic learning models into ensembles. Commonly they are classified into two kinds of ensemble methods, averaging methods and boosting methods [135].

In **boosting methods**, estimators are built in a sequence and one estimator reduces the bias of the combined estimator. For instance, AdaBoost is a boosting method. An AdaBoost classifier starts by fitting a classifier on the original dataset. Next it fits additional copies of the classifier on the same dataset and weights of the wrong classified examples are adjusted in order to concentrate on the wrong cases [135].

In **averaging methods**, several estimators are built independently, and their prediction are averaged. After averaging, the combined estimators typically perform better than any of the single estimators, because they reduce their variance. In essence, bagging methods and ensembles of decision trees are averaging methods.

The following provides details regarding random forest and extremely randomized trees, which were used in this work.

Random Forest

A random forest is an ensemble of random decision trees. Each single tree is randomly different from other trees in the forest model. This leads to de-correlation among tree

predictions. Therefore, a random forest is expected to be more general and robust than a single tree prediction. The prediction of the forest is effected from the type of split function, energy model, leaf predictors and the type of randomness. In addition, the randomness degree parameter ρ controls the randomness degree in each tree and the correlation degree among different trees in the forest. The lower the randomness, the higher the tree correlation will be. If $\rho = \mathscr{T}$, all trees are the same as one another. If ρ declines, the trees become more de-correlated (different from each other).

In a forest with T trees, the variable $t \in \{1, ..., T\}$ is defined as the index of each single component tree. All trees are trained independently and in parallel. During the testing phase, each test data point $\mathbf{v}$ goes through each tree, and arrives at the leaves. Similarly, during testing all trees run in parallel. A single forest prediction is then obtained from a combination of all tree predictions via an averaging operation. For example, in classification, the forest prediction is obtained as,

$$p(c|\mathbf{v}) = \frac{1}{T}\sum_{t=1}^{T} p_t(c|\mathbf{v}), \tag{2.9}$$

where $p_t(c|\mathbf{v})$ represents the posterior distribution obtained by the t^{th} tree. As shown in Figure 2.2, in the classification testing phase, the same test input data $\mathbf{v}$ passes through each single tree in the forest, until it reaches the leaf node. The forest class prediction is obtained as the average of all tree posteriors.

Random forest features high accuracy on unknown data and is very efficient. The main limitation of random forest is the risk of overfit when training models via deep trees [63].

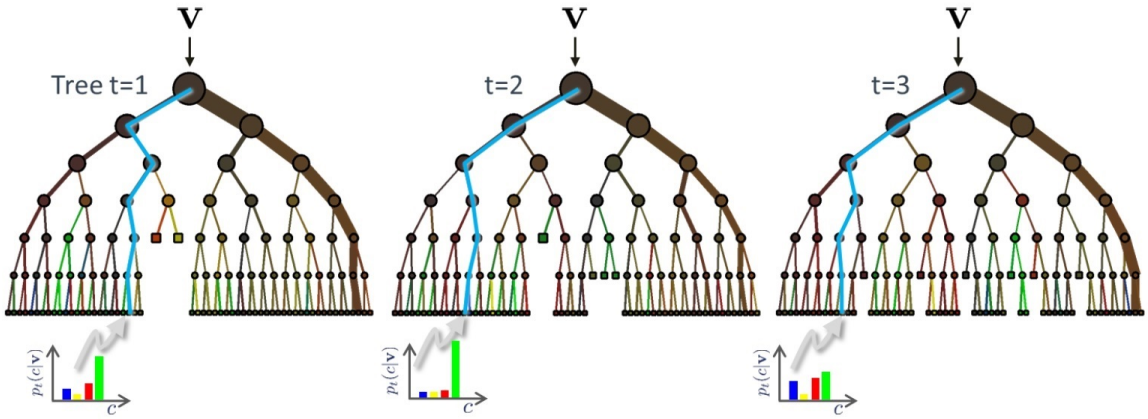


Figure 2.2: Classification forest testing. The same test input data $\mathbf{v}$ passes through each single tree in the forest, until it reaches the leaf node, which stores the posterior $p_t(c|\mathbf{v})$. The forest class prediction is obtained as the average of all tree posteriors. Source: Criminisi et al.[29]

2.4 Extremely Randomized Trees

In this section a special type of random forest, called extremely randomized trees, is presented, and which forms the basic of the work presentation in Chapter 4 and Chapter 5. This section is based from Criminisi et al.[29] and [46] [74] [155] [155] [134].

2.4.1 Algorithm

Extremely randomized trees (Extra-Trees) [46] is a type of random forest (see Appendix 7.1). The aim of Extra-Trees is to increase randomization during the splitting of internal nodes, in order to improve accuracy via reducing variance and reduce training time. There are two main differences between Extra-Trees and other tree-based ensemble methods. First, Extra-Trees use all training data to build trees without random sampling and replacement. Second, Extra-Trees randomly select features to split a node. A random split is picked at each node based on a random cut point of a feature in the $[min, max]$ range of the feature.

The stop splitting criteria is as follows: all features are constant (i.e. no change from previous iteration) in the node, or the output is constant in the node, or the number of training data points in the node is smaller than a specified minimum data n_min.

2.4.2 Split function in Extra-Tree

Any objective function can be chosen with the Extra-Tree to optimize splits at internal nodes. To select a split s, the score measurement is used. The following criteria are used in Chapter 4 and Chapter 5, respectively.

In classification, the score measurement is based on gini index (see Equation 2.5). In regression, the score measurement is defined as

$$Score(s,S) = Var(\mathbf{y}|S) - \frac{|S^L|}{|S|}Var(\mathbf{y}|S^L) - \frac{|S^R|}{|S|}Var(\mathbf{y}|S^R), \tag{2.10}$$

where S^L and S^R represent the two subsets of S according to a split which makes two outputs. $Var(\mathbf{y}|S)$ stands for the variance of the output $\mathbf{y}$ in the data sample S, and it is computed as,

$$Var(\mathbf{y}|S) = \frac{1}{|S|}\sum_{i=1}^{|S|}(\mathbf{y} - \bar{\mathbf{y}})^2. \tag{2.11}$$

Where $\bar{\mathbf{y}}$ stands for the mean of $\mathbf{y}$.

2.4.3 Model Parameters

The parameters that influence an Extra-Tree [46] model are as follows: the forest size (the number of trees T); the number of features randomly selected at each node ρ; the minimum sample size to split a node n_{min}. The parameters T, ρ and n_{min} have different effects for building trees: ρ reflects the strength of the feature selection process, n_{min} determines the strength of averaging output noise, and T stresses the strength of the variance reduction of the ensemble model aggregation. Usually, these parameters could be tuned into a manual or automatic way for specific problems.

- **Feature selection strength** ρ

  The smaller the ρ is, the stronger the randomization of the trees is. The default setting values of ρ are $\rho = \sqrt{d}$ and $\rho = d$, for classification and regression, respectively, where d is the feature dimensionality. When $\rho = 1$, only one random split is chosen at each split node. Then, the tree growing is completely independently from the output in the training data. Therefore, this kind of tree is called ***totally randomized trees***. It is fast to build such trees, but the trees are very prone to noise .

- **Smoothing strength** n_{min}

  The number n_{min} is required for splitting a node. The larger value of n_{min} makes smaller trees, with higher bias and smaller variance.

- **Averaging strength** T

  The parameter T is the number of trees in the ensemble. Higher value of T tends to yield a better accuracy.

The balance between bias and variance is different between classification and regression. Classification can accept high levels of bias of class probability without making high classification error rates. To build good randomized models, the rule of thumb is to keep a low bias on the training data and merge several randomized models for a low variance.

2.4.4 Multiple Output Trees

In supervised learning methods, classification and regression trees can be extended to predict several outputs instead of a single output. For instance, in regression, a vectorial output $\mathbf{y} = (y_1, y_2, ..., y_Y) \in \mathbb{R}^n$, or in classification, $\mathbf{y} \in C^n$, where $C = \{c_1, c_2, ..., c_C\}$ is a set of all classes. There is no correlation between the outputs, but these outputs are related to the same input data. Ultimately, more than one value is stored in the leaves. Comparising with single

output trees, there are two advantages of the multiple output trees. It reduces the training time, and it increases generalization accuracy. In our work, multiple output regression is used (see Chapter 4).

2.5 Registration

This section is based from the ITK software guide [72] and [167] [89]. Image registration is a process that aligns a moving image $I_m(x)$ to the fixed image $I_f(x)$ with the optimal spatial mapping, in which x stands for the position in the n dimensional space. The elements of the registration framework are two input images $I_m(x)$ and $I_f(x)$, a transform $T(x)$, a metric $S(I_f, I_m \circ T)$, an interpolator and an optimizer (see Figure 2.4). Here $T(x)$ stands for the spatial mapping of points x from I_f to points x in I_m, and $S(I_f, I_m \circ T)$ represents a estimation to what extent $I_f(x)$ is matched by the transformed $I_m(x)$. The interpolator measures voxel intensity values at non-grid positions. The optimizer searches the optimal parameters of $T(x)$.

There are different transformation models such as rigid transformation and affine transformation. The work presented in Chapter 3, Chapter 4 and Chapter 5 employed rigid transformation. For each point (x_1, x_2, x_3) in the moving image, a mapping is defined into the coordinate of another space (y_1, y_2, y_3), expressed as,

$$
\begin{aligned}
y_1 &= m_{11}x_1 + m_{12}x_2 + m_{13}x_3 \\
y_2 &= m_{21}x_1 + m_{22}x_2 + m_{23}x_3 \\
y_3 &= m_{31}x_1 + m_{32}x_2 + m_{33}x_3
\end{aligned}
$$

The rigid transformation matrix is defined as,

$$
T_{rigid} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & t_x \\ m_{21} & m_{22} & m_{23} & t_y \\ m_{31} & m_{32} & m_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.12}
$$

where t_x, t_x, t_x represent the translation along each axis of the coordinate system, and the $m_{i,j}$ are the rotation parameters.

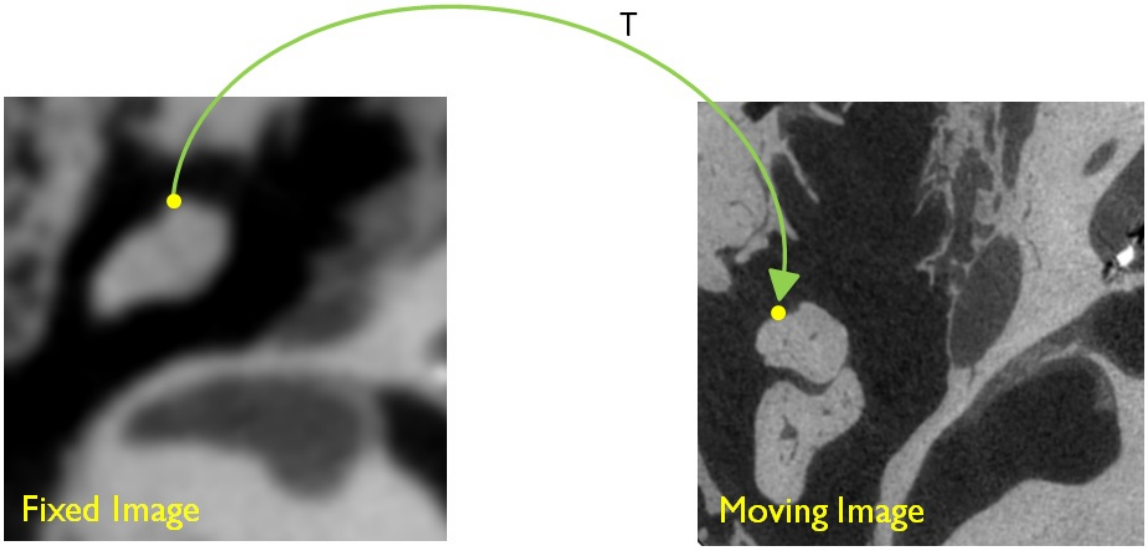


Figure 2.3: The aim of image registration is to find a transform that maps points from the fixed image to points in the moving image.

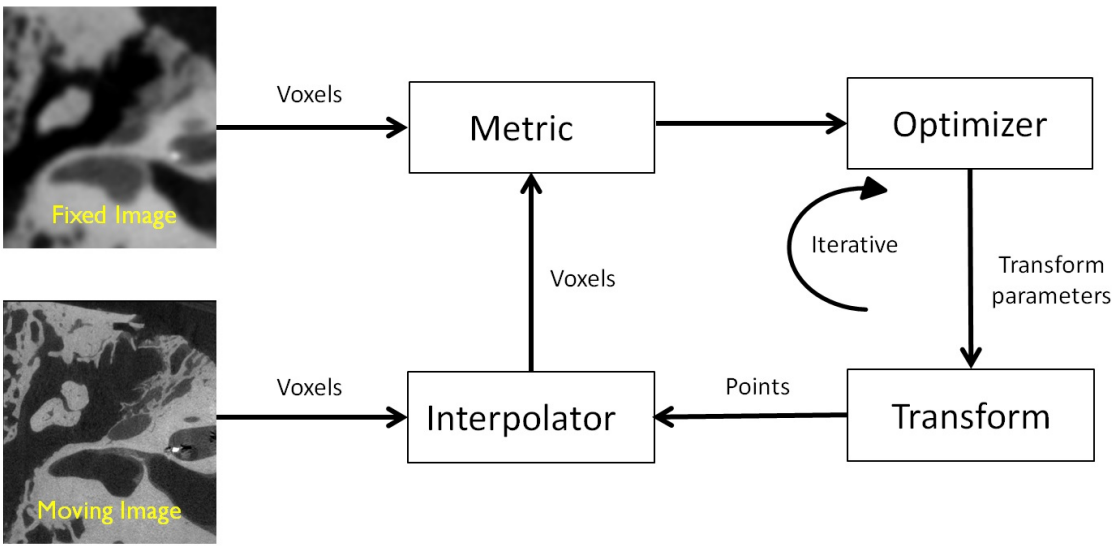


Figure 2.4: The Registration framework. The basic components of the registration are two input images (fixed image and moving image), a transform, a metric, an interpolator and an optimizer. Modified from: [72]

Chapter 3

Motion Detection

3.1 Abstract

Patient head movement is the main reason for blurred images, which might happen during the Cone-beam Computed Tomography (CBCT) scanning process. A blurred CBCT image may lead to suboptimal surgical planning, which in turn can lead to cochlear trauma and damage of critical structures. Therefore, it is important to know the maximum acceptable blurring in the CBCT image for image guided surgery (IGS). In this chapter, we propose a practical and accurate motion detection method in CBCT imaging. A simulated motion scanning system was set up, then a phantom study with different types of robot-controlled motion were obtained. Simulated motion was: rotation at various degree ($0.75 degree, 1.25 degree, 2.50 degree$) with sudden and continuous modes. The demonstrated motion detection method is based on a registration framework and Hausdorff distance. The motion detection method is evaluated by geometric distance. Our preliminary study shows that this practical solution is promising for head motion detection in CBCT imaging and it is the first step towards clinical use.

Keywords blurred CBCT, head movement, Hausdorff distance, rod-phantom, motion detection.

3.2 Introduction

In the clinical workflow, human motions cannot be avoided completely during the medical imaging process (e.g. heart beat, breath, intestinal peristalsis and unconscious movement [85]). In particular, patient head movement (see Figure 3.1) has a big impact on the quality of Cone-beam Computed Tomography (CBCT) images for surgical planning of cochlear implantations. As shown in Figure 3.2, head motion artifacts degenerate the image quality.

The low quality image hinders an optimal surgical planning, which may increase surgical risks for cochlear trauma and damaging of critical structures, such as the facial nerve.

Figure 3.1: Example of head motion artifacts in CT image. It appears as shadowed and streaked. This type of motion is impossible to avoid in most cases, because the patient can not keep still during the CT scanning process. Source: [6]

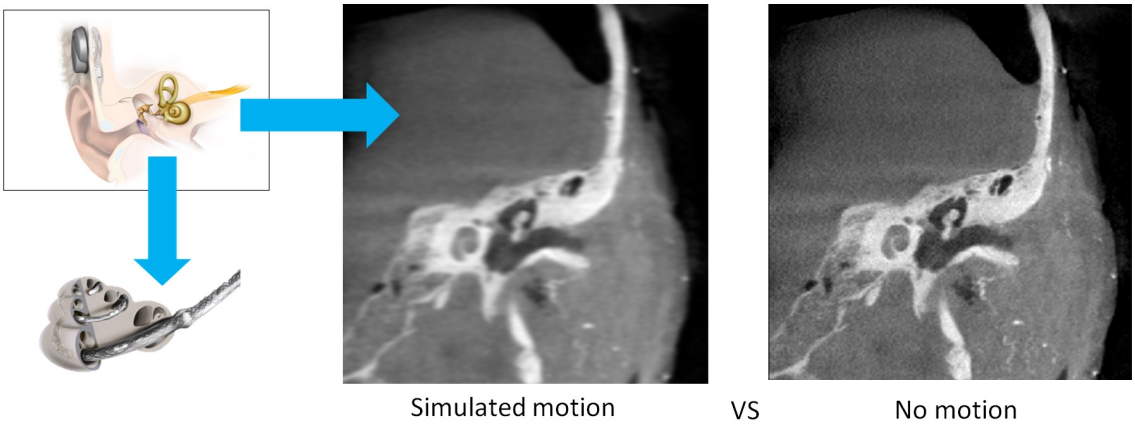


Figure 3.2: Motion comparison in CBCT for surgical planning of cochlear implantation. Compared with the no motion CBCT, the simulated motion CBCT image is blurred. Furthermore, the anatomical structures could not be recognized easily. It might lead to high risk for surgical trauma. Modified from: [136] and [94]

While most of previous techniques have focused on motion correction, few solutions have been proposed for motion detection. In [86], motion correction for respiratory motion blurring in PET/CT with internal-external motion correlation was proposed. In [166], a correction method for motion artifacts in CBCT has been suggested, which uses a patient-specific respiratory motion model. In [77], head motion correction in helical x-ray CT has been achieved by reconstruction from a modified source-detector orbit. Motion has been tracked from volunteers with four types of motion. In [76], a rigid motion correction method for helical CT was performed by motion estimations during the CT scan, and a three dimensional (3D) reconstruction algorithm called likelihood transmission reconstruction (MLTR). In [13], a motion correction method has been proposed based on 3D reconstruction and 2D/3D registration. In [12], motion artifacts in 3D brain CBCT has been detected via a marker based system. In [113] [112] [114] [115], markerless head tracking systems in 3D brain imaging were presented. These approached employ fixed markers attached on the patient's head, such as helmet, headband and goggles. Nevertheless, artifacts on positron emission tomography (PET) images may still occur. In a markerless tracking system [113] [112] [114] [115], structured light projects patterns on the face of a mannequim phantom head, and cameras capture these patterns. Then 3D face surfaces are reconstructed for motion correction. Hence, to the best of our knowledge, no direct method for head motion detection in CBCT image has been proposed before and evaluated a first prototype on phantom CBCT data and robot-controlled motion patterns.

As we known, registration plays a substantial role in robotic image guided system for cochlear implant surgery (see Section 1.3.3), but the fiducial localization error (FLE) is a significant factor for influencing the registration accuracy [44] [41]. Target registration error (TRE) measures the registration error in the image guided system. In [41], a statistical correlation was reported between FLE and TRE, based on their root mean square (RMS) values. It is defined as,

$$TRE_{RMS}{}^2 = \frac{1}{N}(1+\frac{1}{3}\sum_{k=1}^{3}\frac{d_k^2}{f_k^2})FLE_{RMS}^2. \tag{3.1}$$

Where N is the fiducial number, d_k is the distance of a point of interest from principal axis k of the fiducial configuration, and f_k^2 is the mean of the squared distances of the fiducials from that axis of the fiducial configuration. Due to the small tolerance of the robotic cochlear implant surgery, FLE is an important measurement for a successful surgery.

The purpose of our work is to develop a head motion detection method which can potentially work on CBCT imaging without any tracking device in order to find the correlation between the quantity of motion and the fiducial localization error (FLE) (see Figure 3.3). The

requirements of the proposed method are as follows. On the one hand, the method should be robust enough for CBCT imaging. On the other hand, it should easily and quickly filter out the unqualified (i.e. excessive) CBCT for surgical planning in the clinical workfolw.

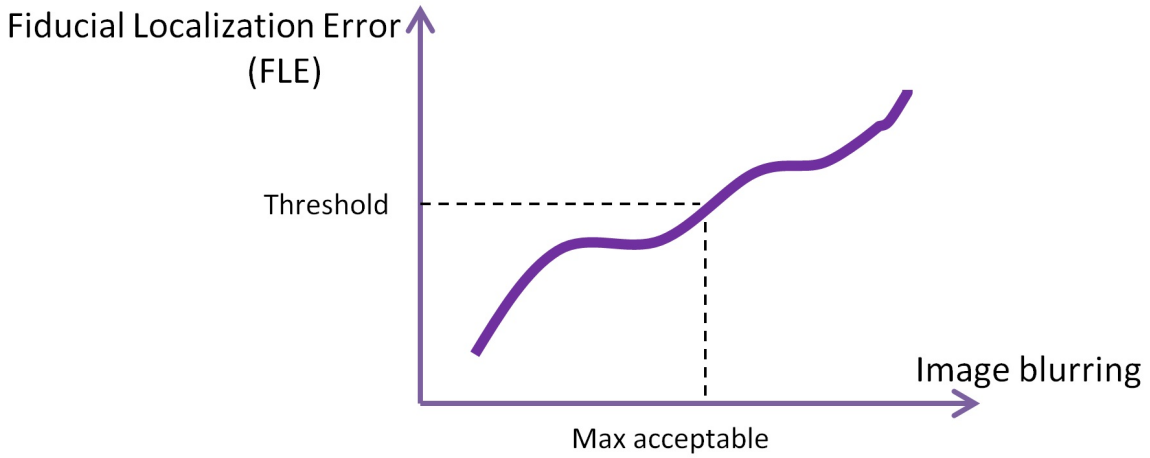


Figure 3.3: Sketch of the correlation between the quantity of motion and the fiducial localization error (FLE). The aim is to find the threshold that can evaluate if the blurred CBCT imaging is suitable for the surgical planning.

In this work, we started with phantom experiments using small rods of known densities in order to estimate the degree of motion. Later on, we will embed them into a wearable ear device during the pre-operative scanning. This chapter is organized as follows. Section 3.3 presents the simulated motion scanning system and the proposed method for head motion detection based on the phantom data. Section 3.4 describes the experimental evaluation, and section 3.5 ends with conclusions and discussion of results.

3.3 Materials and Methods

To detect patient head movement, we have developed an image-based registration pipeline, and tested it on experimental data, as shown in Figure 3.4. First, we extract the center line from the phantom block with and without motion. Then, we register the center line from these images and compute the Hausdorff distance, which describes the maximum distance between the surfaces of the rod phantoms with and without motion.

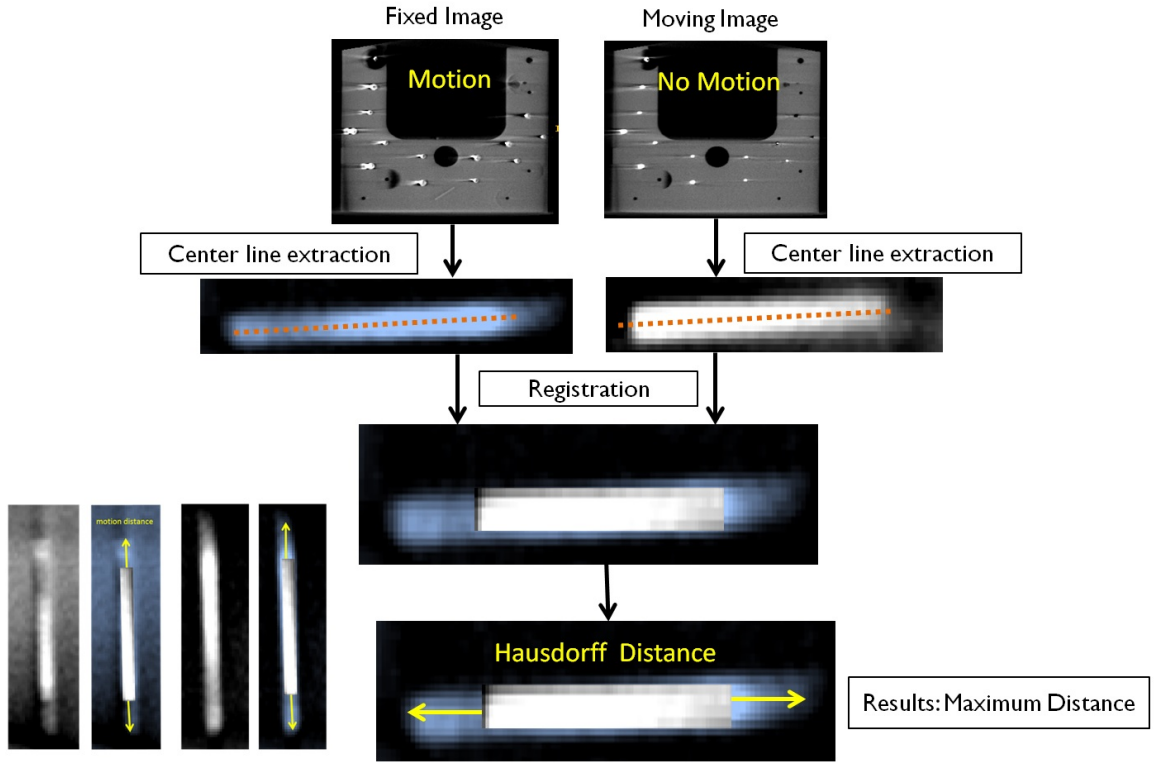


Figure 3.4: The experimental pipeline of CBCT movement detection. Center lines of rod phantoms are extracted, which are regarded as the input data for registration. The center line of the rod-phantom in the motion image is defined as the fixed image, and the center line of the rod-phantom in the no-motion image is considered as moving image. Next, the Hausdorff distance is employed to compute the maximum surface distance between the surface of rod phantoms with and without motion. The simulated head motion is quantified via Hausdorff distance calculation.

3.3.1 Phantom data sets

In order to simulate potential patient head movement during the CBCT scanning process, the following devices were employed to acquire experimental data: an imaging phantom ($90 \times 90 \times 35 mm^3$) with implanted fiducial screws and three implanted rod-phantoms, Planmeca 3D CBCT max imaging system, and the ARTORG Image Guided System (IGS) drilling robot [45]. As shown in Figure 3.5, a transparent block is attached to the drill robot arm.

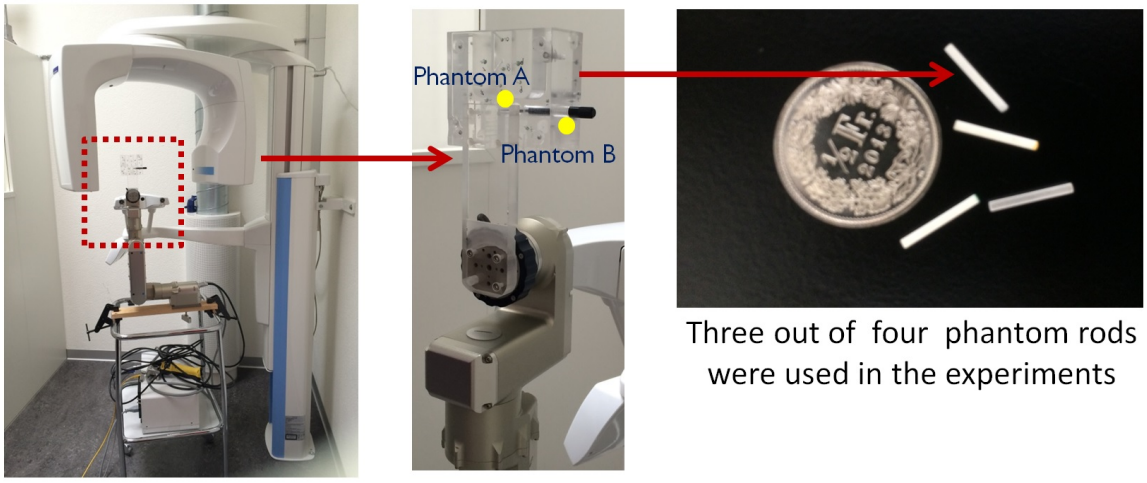


Figure 3.5: Simulated motion scanning system. From left to right: Planmeca 3D CBCT max imaging system and a phantom with the drilling robot, zoomed area for the scanned block, three tiny rod-phantoms attached on the scanned block.

In our study, we have compared three degrees of simulated motion including "slight motion", "moderate motion" and "severe motion". These motions are obtained from the rotation of the drilling robot at $0.75 degree$, $1.25 degree$ and $2.50 degree$, respectively. We have combined them with two motion patterns such as, sudden motion and continuous motion. Both robot and phantom block rotate simultaneously. The robot arm holds the phantom block horizontally, and then it returns to the original position. Figure 3.6 describes the sudden and continuous motion pattern at $1.25 degree$. In sudden motion, there is no motion in the first 5 seconds. Suddenly, the phantom moves $1.25 degree$ and remains still for 5 seconds. Then the phantom moves back to its initial position (i.e. $-1.25 degree$). Then, the phantom further moves by $-1.25 degree$ for 5 seconds. Finally, the phantom returns to the original place and holds the position. In continuous motion, the phantom moves constantly in the first 5 seconds and in 1.25 degrees. The phantom moves back to the original position at a constant speed. Then the phantom then further moves by $-1.25 degree$ for 5 seconds at a constant speed. At last, the phantom moves to the original position at a constant speed in the 20^{th} second.

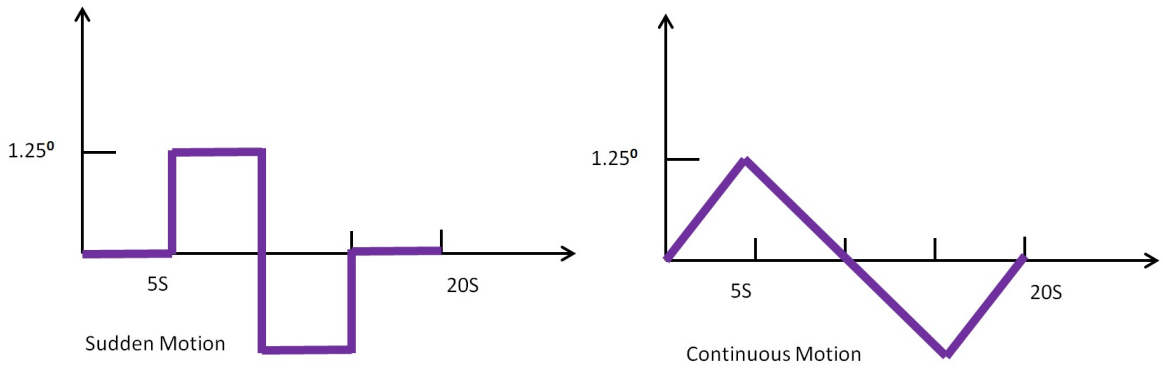


Figure 3.6: One example of motion types at 1.25*degree*. Left: sudden motion every 5 minutes to the left or the right, and return to the original position, then towards the opposite direction. Right: continuous motion in the same back-and-forth routine as the sudden motion.

We have first scanned the stationary phantom block without motion (i.e. 0*degree*) as reference. The different rotation modes of the phantom block are evaluated in relation to this position. We then scanned the moving phantom driven by the drill robot arm. Motion patterns (see Figure 3.7) are summarized as follows:

- **No motion** as reference for other motion patterns
- **Sudden motion (in degrees):** $0.75, 1.25, 2.50$
- **Continuous motion (in degrees):** $0.75, 1.25, 2.50$

The CBCT images were obtained with a Planmeca ProMax 3D Max scanner with $100 \times 90 mm^2 FOV$. For each set of data, a CBCT scan ($0.15 \times 0.15 \times 0.15 mm^3$) with a volume size of $668 \times 668 \times 668$ voxels was performed. Figure 3.8 shows one example of the scanned phantom block, which highlights two rod-phantoms that will be tested in the proposed approach. The zoomed area shows phantom B with and without motion artifacts. The zoomed area also shows three screws inside the block, and how they look like with and without motion.

3.3.2 Preprocessing

To detect motion in the simulated data, we have rigidly aligned the "no-motion" image to the "motion" image. In this work, we define the "motion" image as the fixed image, and the "no-motion" image as the moving image. Registration then brings the moving image into the alignment with the fixed image.

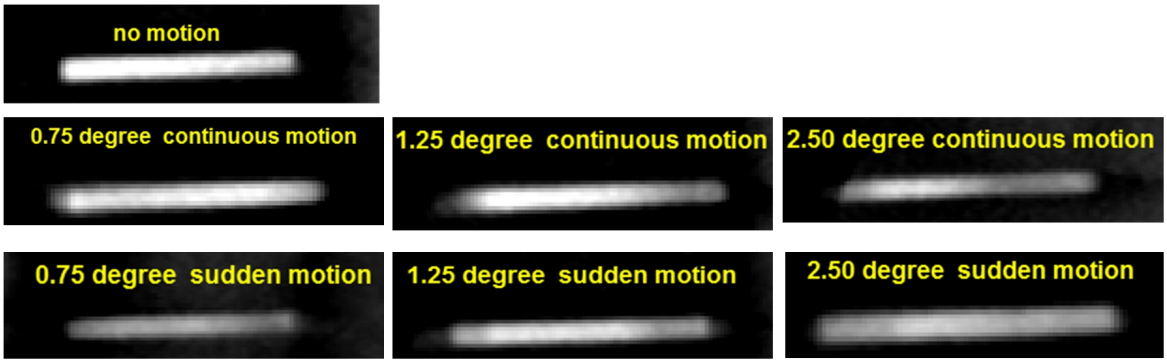


Figure 3.7: The motion patterns we have scanned. The no motion pattern image is the least blurred image. The outline of the rod-phantom can be easily recognized. When the rotation increases, the length of the rod-phantom appears longer.

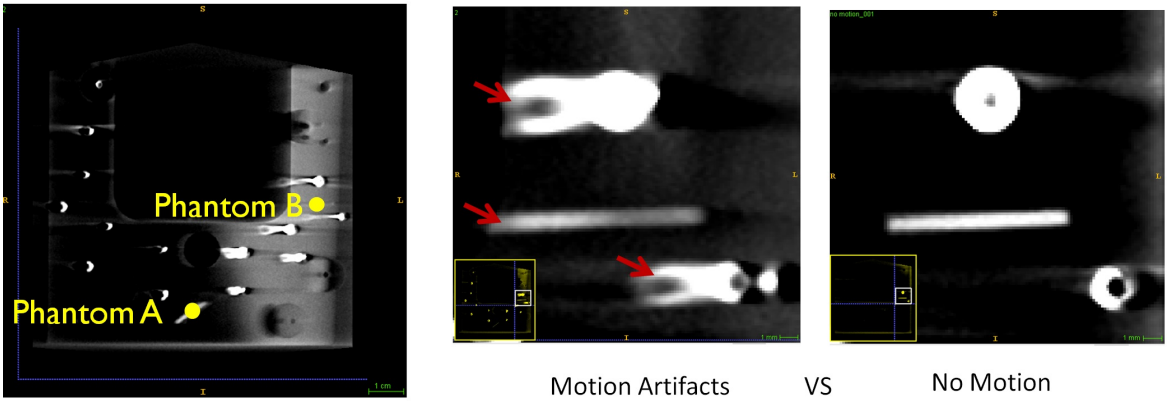


Figure 3.8: One example of the scanned phantom block. From left to right: the scanned block within two rod-phantoms (A and B), zoomed area showing the phantom block with motion artifacts, and zoomed area showing the phantom block without motion.

Pre-processing of CBCT Image with Motion

First, the **highest image intensity** is employed to represent the center pixel of the phantom with motion. $I = 0.95max(I_{image})$, in which 0.95 is a parameter obtained heuristically. This follows the assumption that the center of the object shows the least image distortion. The second step is to represent the original data through the center line of the rod-phantom. In this work, we chose a low-dimensional representation of the center line information. We used **singular value decomposition (SVD)** as a simple and effective way to fit a line in 3D space via eigenanalysis [138] [5][50]. It is defined as,

$$A_{mn} = U_{mm}S_{mn}V_{nn}^{T}. \tag{3.2}$$

In SVD, a rectangular $(m \times n)$ matrix A is decomposed into the product of three matrices, which includes an $(m \times m)$ orthogonal matrix U of eigen vectors ($UU^T = U^TU = I$), a diagonal $(m \times n)$ metric S of eigenvalues ($S = \begin{bmatrix} \sigma_0 & 0 & ... & 0 \\ 0 & \sigma_1 & ... & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & ... & \sigma_n \end{bmatrix}$, the singular values $\sigma_0 \geqslant \sigma_1 \geqslant ... \geqslant \sigma_n \geqslant 0$, where all the values in the diagonal S are positive and sorted from the largest value to the smallest) [40], and the transpose of an $(n \times n)$ orthogonal matrix V ($VV^T = V^TV = I$) [5][50]. The eigenvectors correspond to a set of orthogonal axes. Each eigenvector stands for one direction of variation. Moreover, each direction has a various scale which is provided by the eigenvalues [146]. Eigenvalues stand for the variance of the data points in each direction [146].

Pre-processing of CBCT Image without Motion

As the density of the four rod-phantoms is known, a simple image **threshold** [128] was used. The threshold chooses a range value of voxels as foreground, and defines the rest of voxels as the background. In this work, the thresholds correspond to the intensity of the rods $(0, 400, 800, 1200)$. Manual corrections were performed if needed. Three rod-phantoms were attached on the block $(400, 800, 1200)$. After obtaining binary images from the threshold method, a **distance map** [128] was computed for each thresholded rod phantom to extract its center line (i.e. skeletonization). The distance map is a method for generating a gray scale image from a binary digital image [128]. It defines the pixel values of the foreground as the value of the distance to the nearest background. In this work, the Chamfer distance is employed for computing the distance map. The Chamfer distance is similar to the Euclidean distance, but it computes faster than the Euclidean distance [22].

3.3.3 Alignments of Center Line

Once the center lines of rod-phantoms are extracted from the preprocessing step, they are regarded as the input data for registration. The registration method brings the center lines of rod phantoms into alignment (see Appendix 7.2), allowing the following step of measuring Hausdorff distance (see Figure 3.9). In this step, point set to point set registration method is applied at the center line of of rod-phantoms from the "motion" image and the "no-motion" image. It uses the fixed point set from the "motion" image as a reference and looks for a transformation that maps points from the space of the fixed point set to the space of the moving point set [72].

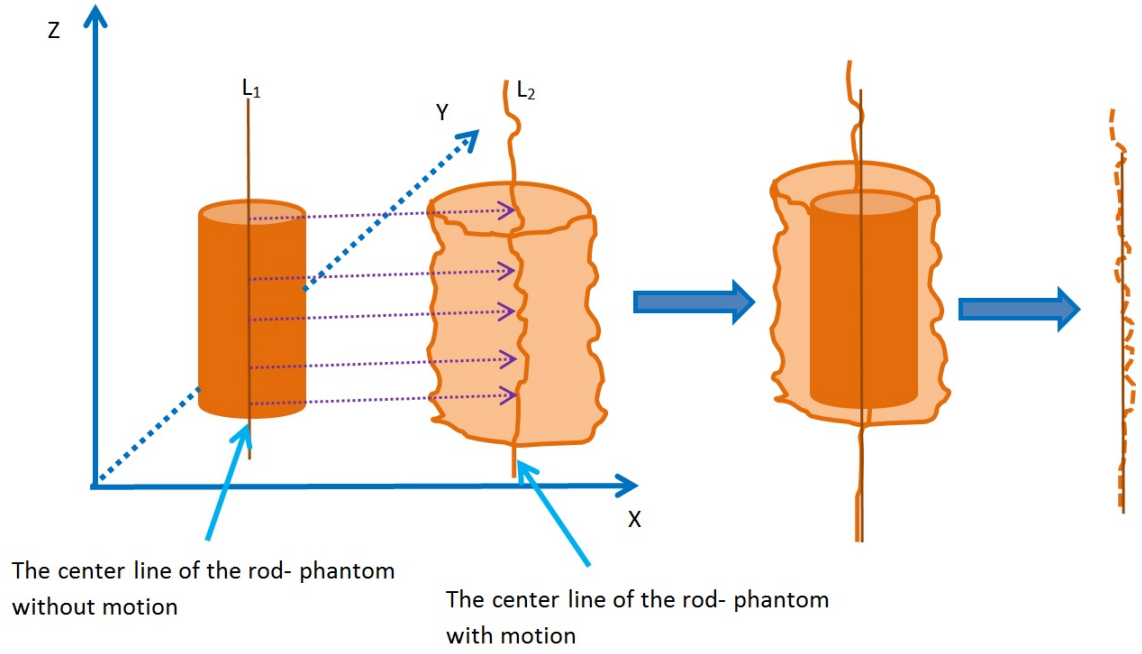


Figure 3.9: Sketch of fitting the center line of the rod-phantom for registration. During the registration, the moving point set of the center line of the rod-phantom from the "no-motion" image aligns to the fixed point set from the "motion" image.

3.3.4 Hausdorff Distance Calculation

To find the largest distance from the boundary of the phantom without motion to the boundary of the phantom with motion, the Hausdorff Distance (HSD) is employed. The following metric measures the Hausdorff distance [67] from the ground truth surface to its nearest neighbor in the segmented surface.

$$H(A,B) = max(h(A,B),h(B,A)), \tag{3.3}$$

where

$$h(A,B) = \max_{a\in A}\{\min_{b\in B}||a-b||\}. \tag{3.4}$$

The function $h(A,B)$ is the Hausdorff distance between two surfaces A (no motion image) and B (motion image). The term $||a-b||$ is the distance between point a and b (in our case the Euclidean distance). The underlying assumption is that the smaller the Hausdorff distance value, the smaller the amount of detected motion.

3.4 Experimental Design

Three rod-phantoms, with intensities 400, 800 and 1200, were attached to the rotating block and imaged with Planmeca CBCT. The rod-phantom with intensities 800 was positioned parallel to the x-ray beam, making it nearly imperceptible. Therefore, in our experiments, rod-phantoms with intensities 400 (Phantom B) and 1200 (Phantom A) were used. The no-motion images are used as a reference image during registration. Registration, singular value decomposition (SVD) and distance map were performed with the Insight-Toolkit version 4.4.1 [73]. The detail of the implementation is described in Appendix 7.2.

3.4.1 Evaluating detected motion with a geometrically generated Ground Truth

As ground truth data, we calculated the geometric distance, since the geometry of the phantom block is known. Figure 3.10 shows the geometric distance that can be calculated analytically, as follows,

$$L^2 = R^2 + R^2 - 2RRcos\theta, \tag{3.5}$$

where R is the length of the block, θ is the rotated degree that is controlled by the robot arm, L is the geometric distance.

3.4.2 Evaluation

In this section we present motion detection results for the two rod-phantoms we measured. The geometric distance was used to verify the proposed approach. Figure 3.15 shows the obtained measurement results.

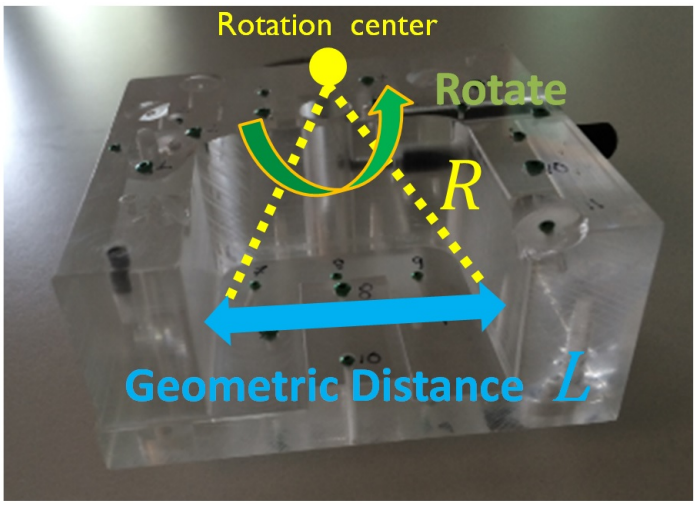
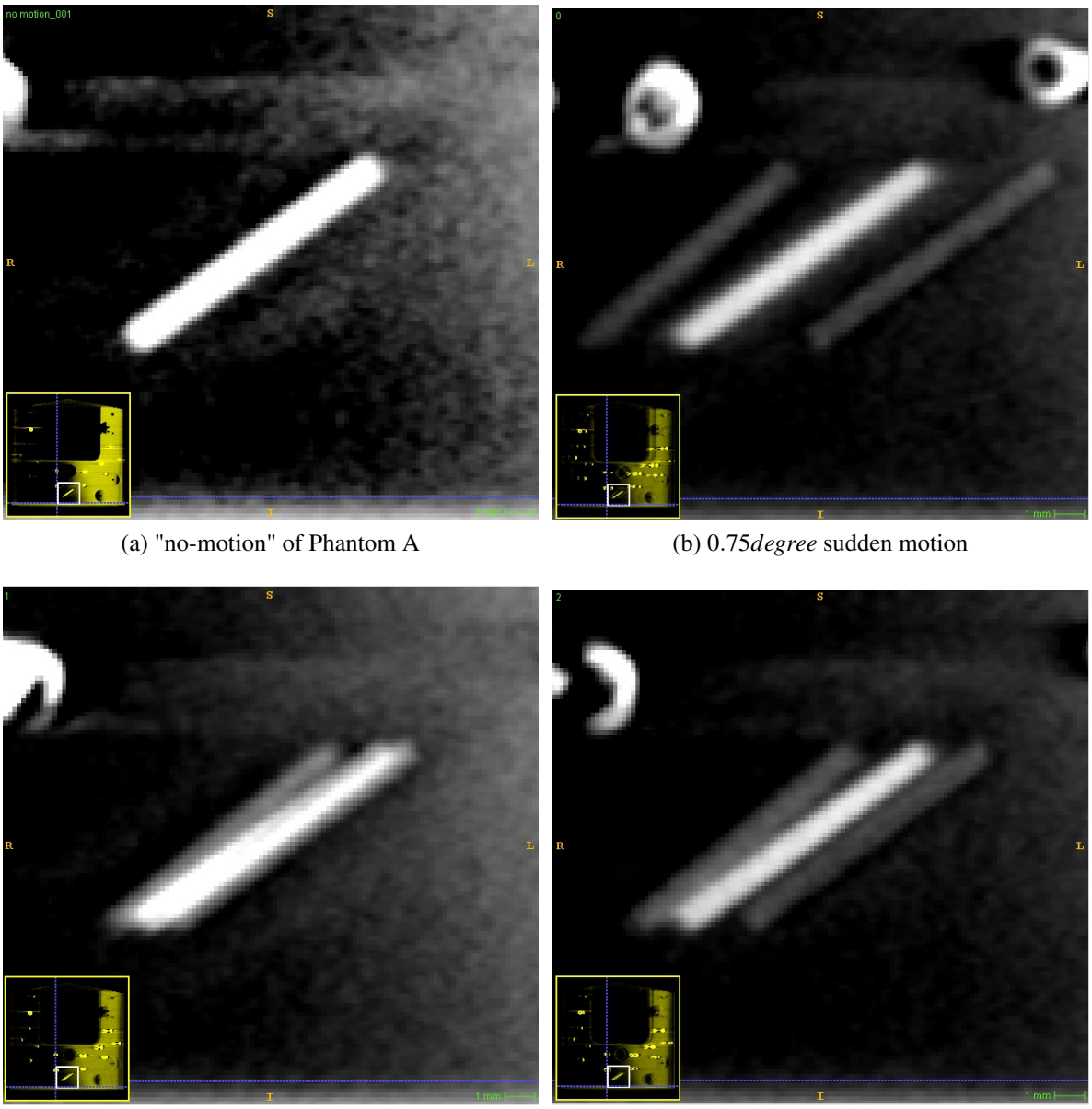


Figure 3.10: The phantom block used to simulate controlled motion patterns. The amount of motion can be calculated analytically as the dimensions of the rods and the magnitude of the motion are known.

1) Sudden Motion Study: We checked the quality of simulated sudden motion. We observed that "illusory" rod-phantoms of Phantom A appear on the CBCT in the sudden motion mode (see Figure 3.11). It is impossible to measure the amount of motion in these images, as multiple objects then appear on the image. The "0.75*degree*" sudden motion resulted in largest distances between two illusory rod-phantoms. Because the position has a big effect on Phantom A, it can not be measured by our proposed method. Nevertheless, the sudden motion patterns of Phantom B were the same as we expected (see Figure 3.12). They are able to be quantified by the proposed method. For the sudden motion experiments, we found that the positioning of the rod-phantom influences the CBCT scanning. Due to these issues for sudden motion patterns, we mainly concentrate on continuous motion experiments.

2) Continuous Motion Study: The performance of the simulated continuous motion patterns in Phantom A and Phantom B are consistent, as shown in Figure 3.13 and Figure 3.14. Figure 3.15 shows the blurred distances of Phantom A and B are larger than 2*mm* in 0.75*degree* continuous motion. As we know, at least 0.3*mm* distance between the drill and the facial nerve is needed in cochlear implant surgery. Therefore, motion blurred from 0.75*degree* exceeds the acceptable range of motion artifacts. Overall, and as expected the larger the amount of motion, the larger the Hausdorff distance. From the experiments, both the Hausdorff distance and the geometric distance are positively correlated, indicating the feasibility of using the Hausdorff distance as a metric for motion estimation.



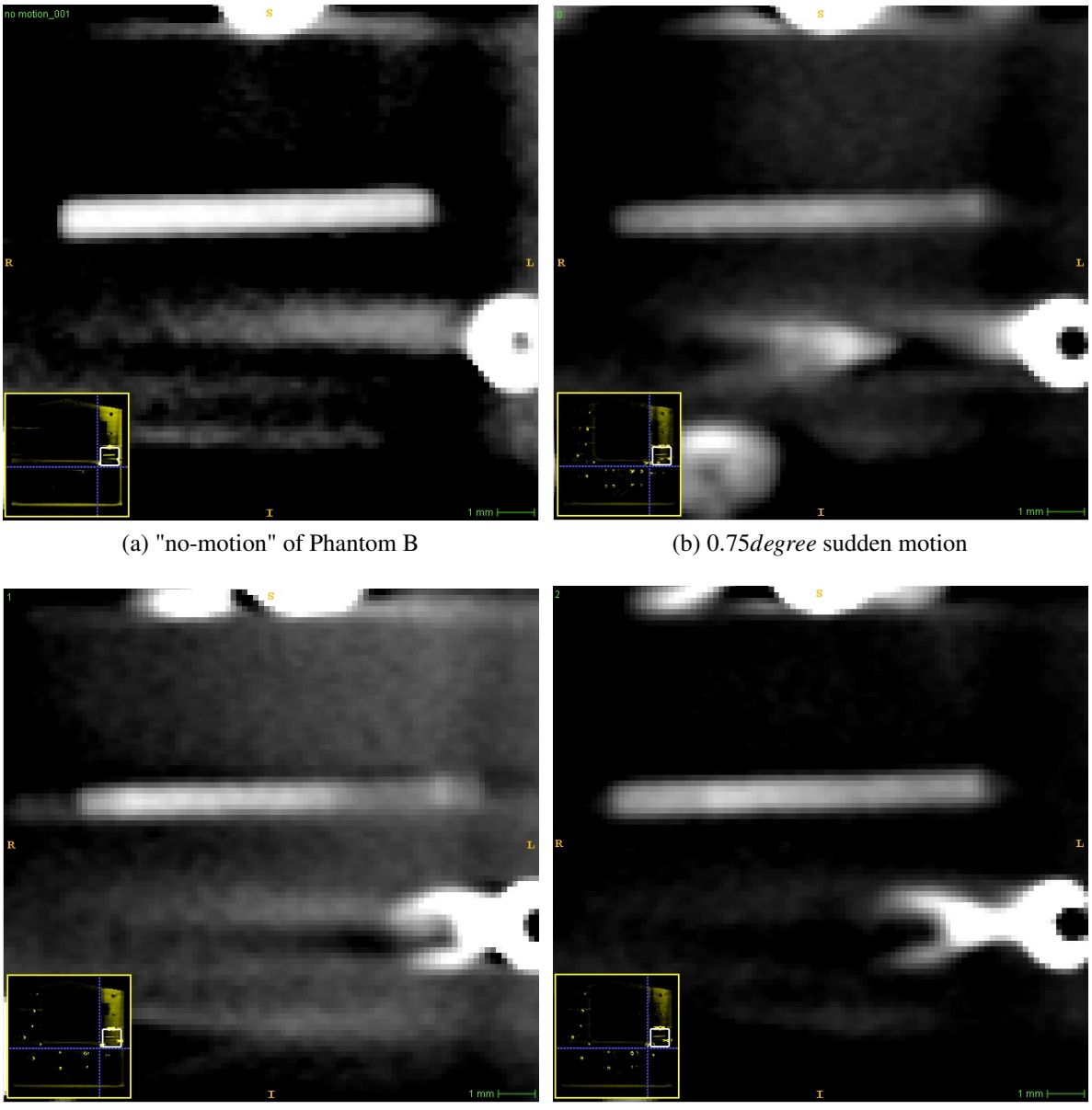
(a) "no-motion" of Phantom A

(b) 0.75*degree* sudden motion

(c) 1.25*degree* sudden motion

(d) 2.50*degree* sudden motion

Figure 3.11: One example of sudden motion of the Phantom A. "Illusory" rod-phantoms appear on the CBCT in the "0.75*degree*", "1.25*degree*", "2.50*degree*" sudden motion images, respectively. The distances among illusory rod-phantoms in the "0.75*degree*" sudden motion images are larger than the "2.50*degree*" one. The smallest distances are in the "1.25*degree*" sudden motion image.



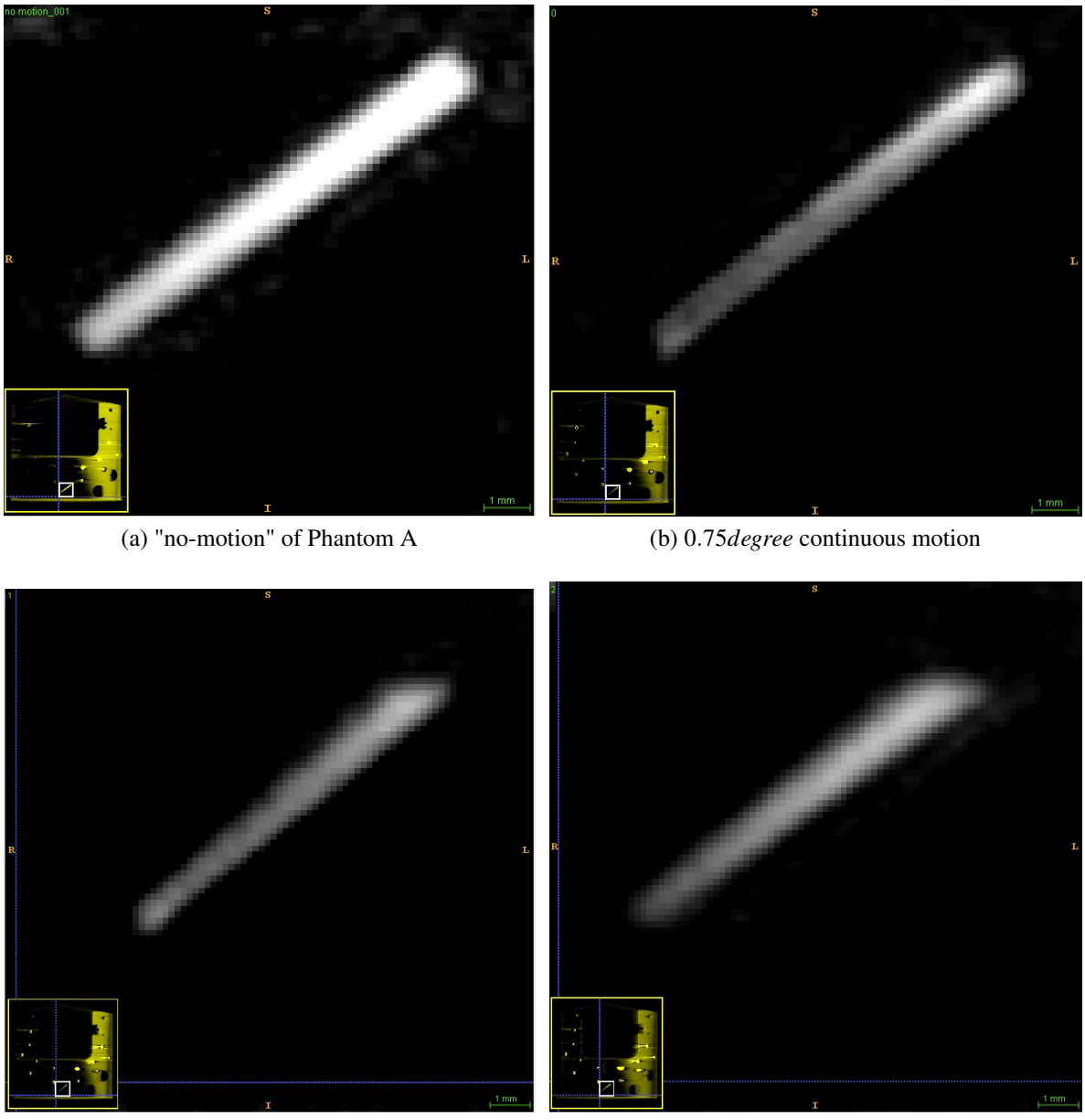
(a) "no-motion" of Phantom B

(b) 0.75*degree* sudden motion

(c) 1.25*degree* sudden motion

(d) 2.50*degree* sudden motion

Figure 3.12: One example of sudden motion of the Phantom B. "no-motion" pattern of Phantom B is easily distinguished. Different degrees sudden motion make phantom B blurred in various levels.



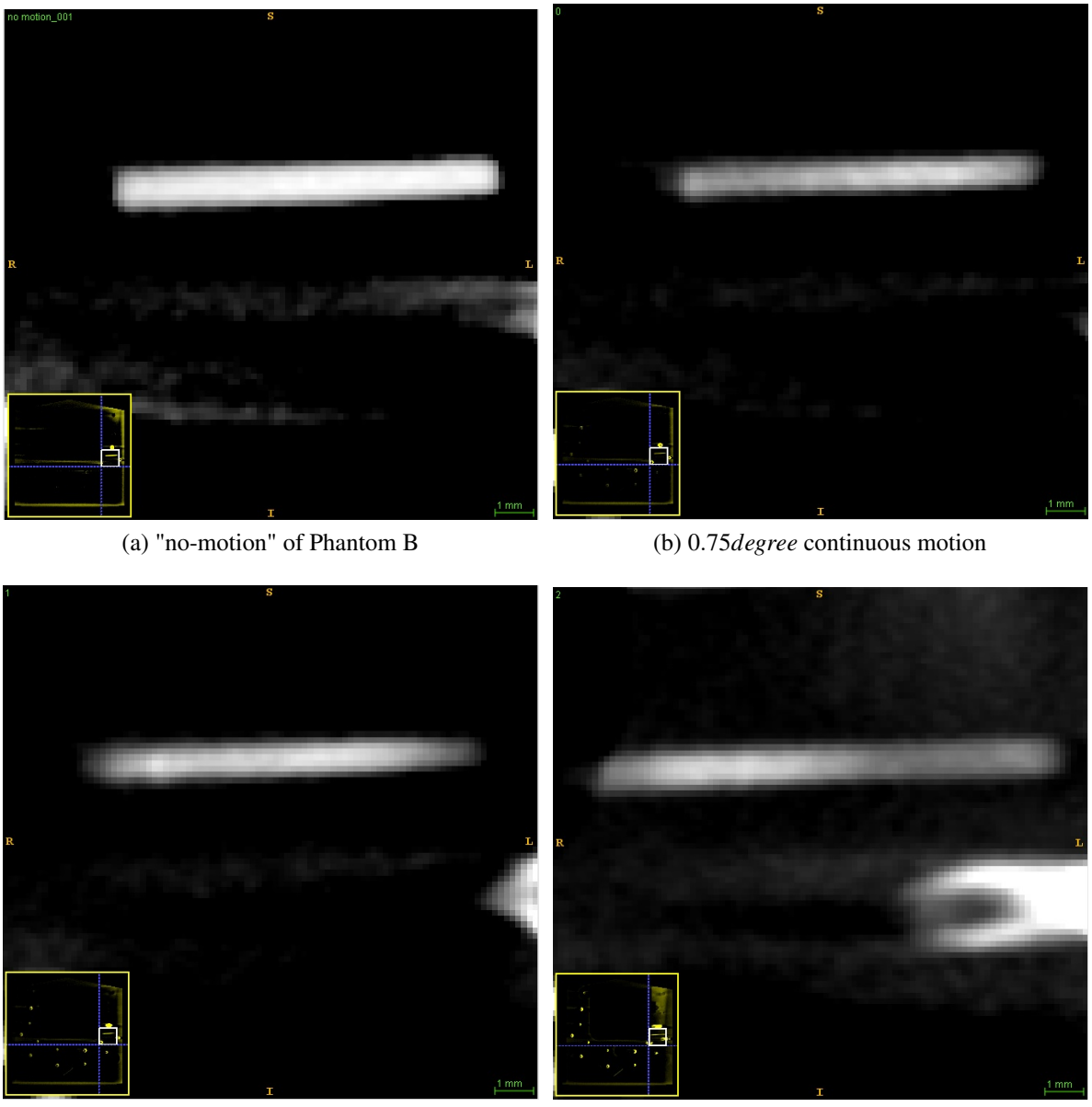
(a) "no-motion" of Phantom A

(b) 0.75*degree* continuous motion

(c) 1.25*degree* continuous motion

(d) 2.50*degree* continuous motion

Figure 3.13: One example of continuous motion of the Phantom A. "no-motion" pattern of Phantom A is the brightest among the rest of continuous motion pattern images. The continuous motion results in the uneven distributed intensities of Phantom A.



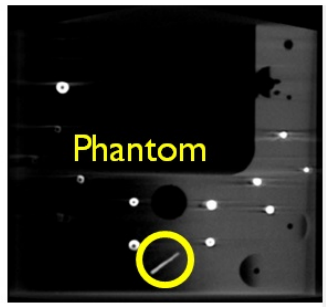
(a) "no-motion" of Phantom B

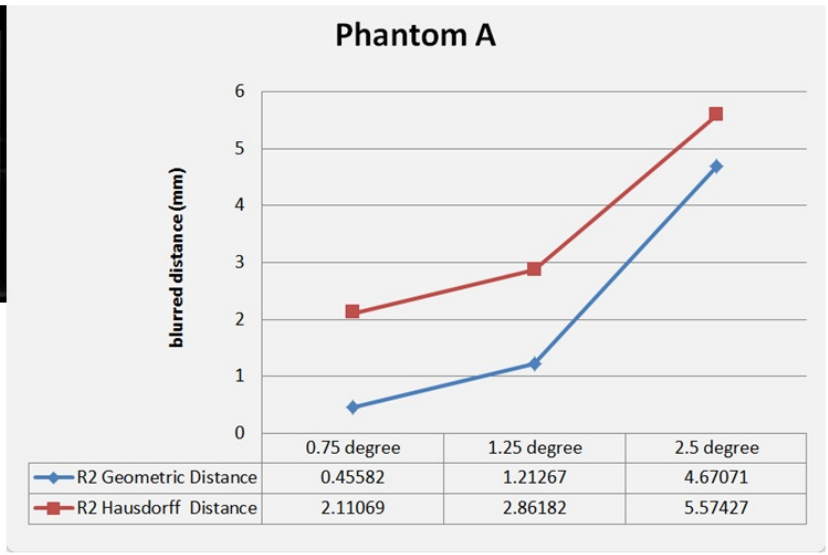
(b) 0.75*degree* continuous motion

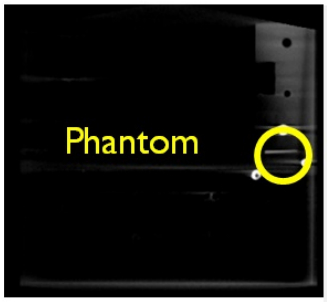
(c) 1.25*degree* continuous motion

(d) 2.50*degree* continuous motion

Figure 3.14: One example of continuous motion of the Phantom B. "no-motion" of Phantom B has the sharpest boundary among the continous motion patterns of the Phantom B. The larger motion rotation leads to the longer length of the Phantom B in the CBCT image.







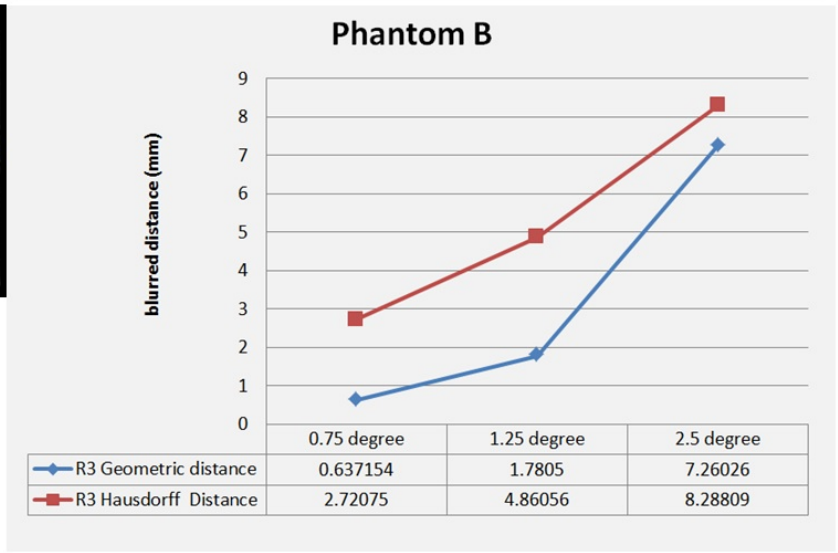


Figure 3.15: Correlation between ground truth (computed analytically) and estimated motion using the proposed image-based pipeline. First experiments on two phantom experiments show a good correlation, indicating the feasibility of the proposed approach to estimate head motion.

3.5 Discussions

In this preliminary study we have developed a practical motion detection method for CBCT image via Hausdorff distances based on image registration framework. We have simulated two different motion modes, sudden motion and continuous motion. In general, preliminary results indicate that the proposed method is able to reliably quantify motion. An acceptable degree of correlation between real motion (as calculated analytically) and amount of blurriness, quantified by the Hausdorff distance, has been found, suggesting the potential of this practical solution to detected motion in CBCT imaging.

In sudden motion study, we have observed that placing rod-phantoms with different orientation within the block have a major impact on the motion pattern. The reason is that the instant change at the exit angle between the rod-phantom and beam source from the Planmeca imaging system results in the detector receiving different amount of X-ray beams in the desired region. Two unexpected illusory rod-phantom shadows lead to conclude that the proposed method cannot be applied for sudden motions. In next step, we will explore the relationship between the phantom position and motion patterns, in order to find the optimal posture of the patient's head for CBCT scanning and correlate our measurements directly with FLE measurements.

In continuous motion study, the $2.50 degree$ robot-controlled motion has produced much more blurred image. The Hausdorff distance exceeded $5mm$ in both rod-phantoms, and it even reached around $8.3mm$ in one rod-phantom. Moreover, the smallest $0.75 degeree$ rotation produce a Hausdorff distance of more than $2mm$. It means that the patient head movement cannot rotate more than $0.75 degree$ in the cochlear implant surgery. However, this depends on the location of the rotation center.

The next step will be to improve the experimental setup by correlating Fiducial Localization Error (FLE), and Hausdorff distance, in order to determine an acceptable "blur" threshold, and create a template (i.e. wearable phantom) that the patient can wear during scanning. As second evaluation, we will choose digital cameras to capture head motion, similar to [113] [112] [114] [115]. Next, we will analyze the degree of motion in each direction and check if the detected motion is acceptable for cochlear imaging.

3.6 Conclusions

We have presented a practical and reliable motion detection method for estimating the degree of motion in simulated blurred CBCT images. This is the first step towards the clinical routine. In clinical workflow, it will be used to assess the degree of motion, deemed

acceptable for surgical planning of cochlear imaging. A simulated motion scanning system has been arranged, in which a robot controls the rotation modes for three rod-phantoms within a phantom block. The motion detection method is based on image registration and Hausdorff distance measurements. The results show that the proposed method is able to calculate the degree of motion approximately, and can be integrated in clinical practice.

Chapter 4

Facial Nerve Image Enhancement

This Chapter has been submitted as a conference article.[1]

4.1 Abstract

Facial nerve segmentation plays an important role in surgical planning of cochlear implantation. Clinically available CBCT images are used for surgical planning. However, its relatively low resolution renders the identification of the facial nerve difficult. In this work, we present a supervised learning approach to enhance facial nerve image information from CBCT. A supervised learning approach based on multi-output random forest was employed to learn the mapping between CBCT and micro-CT images. Evaluation was performed qualitatively and quantitatively by using the predicted image as input for a previously published dedicated facial nerve segmentation, and cochlear implantation surgical planning software, OtoPlan. Results show the potential of the proposed approach to improve facial nerve image quality as imaged by CBCT and to leverage its segmentation using OtoPlan.

4.2 Introduction

Cochlear implantation is a conventional treatment for patients suffering from profound hearing loss. The surgical operation for cochlear implant requires mastoidectomy, to access

[1] Ping Lu, Livia Barazzetti, Vimal Chandran, Kate Gavaghan, Stefan Weber, Nicolas Gerber, Mauricio Reyes, "Facial nerve image enhancement from CBCT using supervised learning technique", In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pp. 2964-2967. IEEE, 2015.

the cochlea and avoid critical anatomical structures. To minimize the invasiveness of the surgical operation, a surgical robot system with an associated planning tool, OtoPlan [45], has been developed. OtoPlan assists the robotic system to perform drilling for direct cochlear access [9]. One of the main challenges of this procedure is to avoid the facial nerve with a margin of at least $0.5mm$. Any damage of the facial nerve causes temporary or permanent paralysis in the ipsilateral face. Hence, an accurate facial nerve segmentation is a critical step for an effective surgical plan.

The surgical planning is performed on cone-beam computed tomography (CBCT) images. In clinical practice, CBCT images are acquired with reduced radiation dose to patients, which may result in low image quality and less clear structure border. The diameter of the facial nerve lies in the range of $0.8-1.7mm$ [126]. Accurate segmentation of the facial nerve from the acquired CBCT images is challenging, mainly in the border region. Image enhancement is hypothesized to enhance the overall image quality and thereby to improve facial nerve segmentation.

In recent years, several CBCT image enhancement algorithms based on deterministic models has been proposed [88] [92] [107]. However, they were built from a priori knowledge of the imaged anatomy or imaging process. Alternatively, supervised learning approach has been proposed to learn the relationship between the acquired low-resolution and corresponding high-resolution image [3]. In this work, we propose to apply supervised learning based on multi-output random regression forest to enhance the image quality, in order to obtain a faster and more reliable facial nerve segmentation in the framework of pre-operative cochlear planning.

Below, the proposed approach is described and a detailed description of the image enhancement process for facial nerve segmentation is presented. An initial evaluation of the approach performed on CBCT images of cadaveric specimen and segmentation results, as compared to ground truth micro-CT images, is presented.

4.3 Methods

In supervised learning, image features $x = (x_1, ...x_n) \in \mathbb{R}^n$ computed on clinical CBCT are mapped to the corresponding output response $y = (y_1, ...y_m) \in \mathbb{R}^m$ computed on micro-CT. The mapping is cast as a regression problem. Given a training set $\{\langle X_i, Y_i\rangle | i = 1, ..., N\}$ of CBCT and micro-CT aligned pairs of images, we extract from each i_{th} image, a feature vector $X_i = (x_1, ...x_C) \in \mathbb{X}$ and responses $Y_i = (y_1, ...y_C) \in \mathbb{Y}$ over a grid of C voxels. Then, a function $\hat{y} : \mathbb{X} \mapsto \mathbb{Y}$ from a space of features $\mathbb{X}$ to the space of responses $\mathbb{Y}$ that predicts the response for any new test image feature $X_{test} \in \mathbb{X}$ is constructed.

The complete pipeline is presented in Fig. 1, and is described below.

4.3.1 Feature Extraction

Input Features

For feature extraction, the CBCT image was rigidly registered and resampled to the micro-CT image, in order to capture the image mapping from low to high- resolution at the same spatial locations. A uniform sampling grid with isotropic grid spacing is defined over the CBCT and micro-CT images. At each node c_j of the grid $j = \{1, ...C\}$, a volume of interest (VOI) $5 \times 5 \times 5$ is extracted on which feature descriptors are computed.

We propose to use two family of features, intensity- and texture-based. The intensity-based features includes all the intensity values. The texture-based features includes first order statistical measures and the gray-level co-occurrence matrix (GLCM) [53] [116]. The list of texture features is presented in Table 4.1. For texture based features, a feature pooling was performed, followed by principle component analysis (PCA) to reduce dimensionality and redundancy of feature sets. This reduces the feature space from $\mathbb{R}^{214}$ to $\mathbb{R}^{34}$. Hence, the input feature set $x = (x_1, ...x_n)$ includes all the image intensities, first order statistics and mean and variance of all the pooled GLCM features (i.e. $n \in \{\mathbb{R}^{125} + \mathbb{R}^{34} = \mathbb{R}^{159}\}$)

Table 4.1: List of texture - based features computed at each grid node.

Texture Features	
1st Order Statistics	**GLCM**
Mean	Energy
Std.Dev	Entropy
Skewness	Correlation
Kurtosis	Inertia
Minimum	Cluster Shade
Maximum	Cluster Prominance
	Inverse Difference Moment
	Haralick Correlation

Output Response

A VOI of $3 \times 3 \times 3$ is extracted at each corresponding grid node c_j from micro-CT. The output response set $y = (y_1, ...y_{d2})$ includes all the intensity information ($d2 \in \mathbb{R}^9$).

4.3.2 Multi-Output Regression Model

Decision forests are a group of learning methods widely used for classification and regression tasks in machine learning and computer vision. An extension of decision forest with extra trees algorithm has been proposed to handle multi-ouput image classification [36, 93]. We adopted this technique as a regression approach for its ability to preserve local intensity patterns. During supervised learning, the algorithm randomly selects without replacement K input variables $\{v_1, ...v_k\}$ from the training data $D := \{\langle X_i, Y_i\rangle | i = 1, ..., N\}$. For each selected input variable, within the interval $[v_i^{min}, v_i^{max}]$ a cutpoint s_i was randomly defined, followed by spliting $[v_i < s_i]$. Among the K candidate splits, the best split was chosen via optimizating the L2 mean square error [93].

During testing, image features were extracted and passed through the regression random forest. The computed output corresponds to the intensity of the central voxel of the designed $3 \times 3 \times 3$ voxels window. No further post-processing was performed on the resulting image.

4.4 Results

We report in this study preliminary results obtained on a database of right and left ears from 4 cadaver heads following an approved clinical study. Pairs of CBCT isotropic ($0.15\times$ 0.15 $\times 0.15 mm^3$) and micro-CT ($0.018 \times 0.018 \times 0.018 mm^3$) images were acquired from the four heads. The images were rigidly aligned using a rigid registration transform, normalized cross-correlation and gradient descent optimization. For rigid registration, we defined CBCT as the moving image and micro-CT as the fixed one. Then we resampled CBCT with the micro-CT voxel spacing. From the resulting rigidly aligned images, image patches were extracted and used for the training phase.

For evaluation, we manually segmented the facial nerve from the micro-CT image (referred hereafter as groundtruth). We used the software OtoPlan [45] to perform segmentation of the facial nerve from the original CBCT image and its corresponding enhanced version. OtoPlan is a dedicated state-of-the-art software for cochlear implantation surgical planning.

We performed qualitative visual assessment of the resulting segmented facial nerve as well as a quantitative analysis of surface-to-surface distances to the ground-truth segmentation. For implementation, we use the scikit-learn package [120] and its ExtraTreesRegressor, which implements a multi output random forest regression algorithm. For model parameterization (i.e. number of estimators, number of trees, etc.) we adopted a leave-one-out strategy. Additionally, we investigated the influence of the window size on the prediction accuracy.

Figure 4.2 shows the results obtained after regression forest application to a CBCT image. Compared to the input CBCT image, the resulting enhanced image is much sharper and able to characterize image details not completely discernible from the CBCT image.

Following the recent findings from [3], we further explored the importance of the window size used to extract feature information during the training phase, Our results showed that using short range features (i.e. window size in our application smaller than $0.018mm$) yields

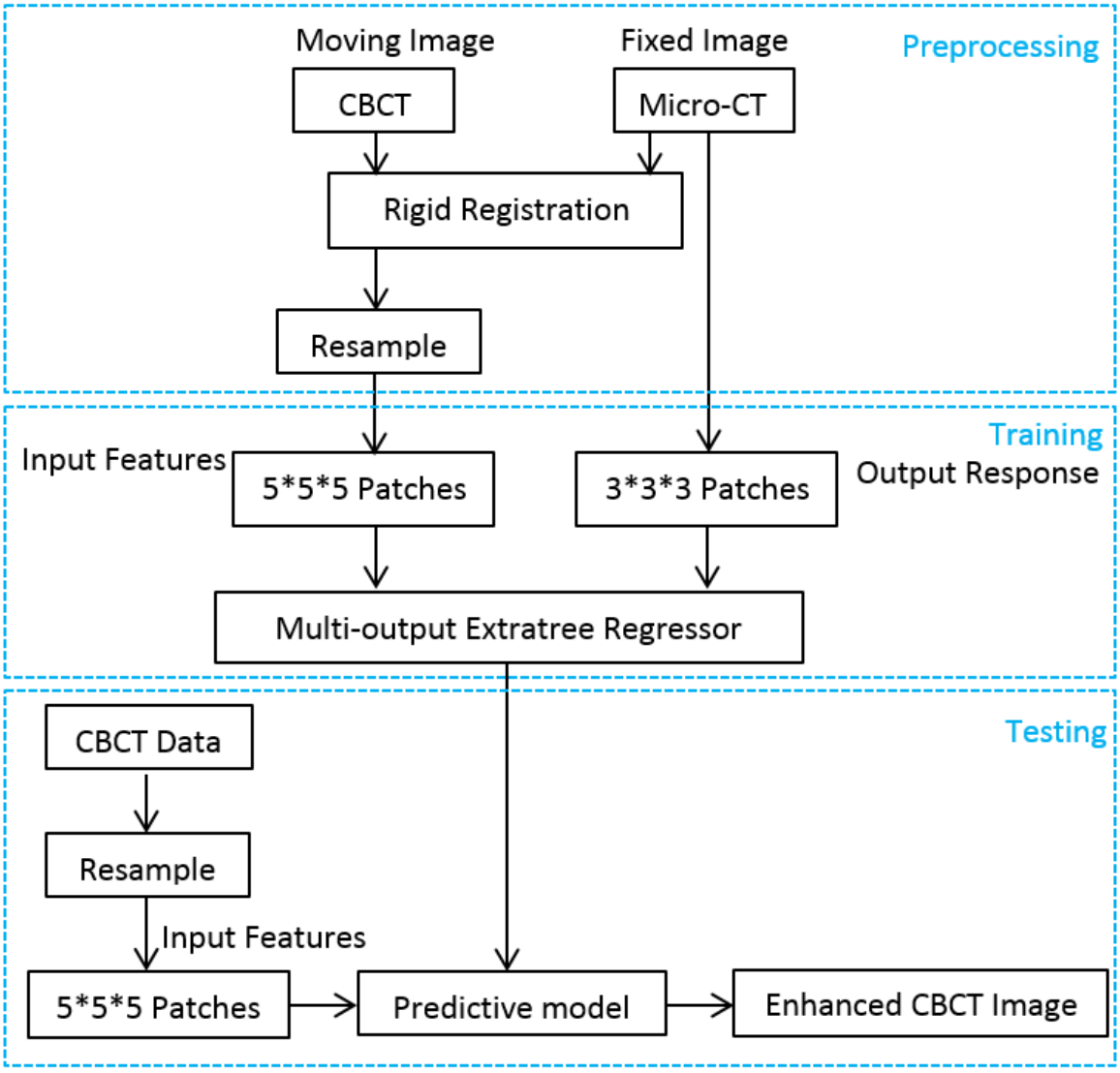
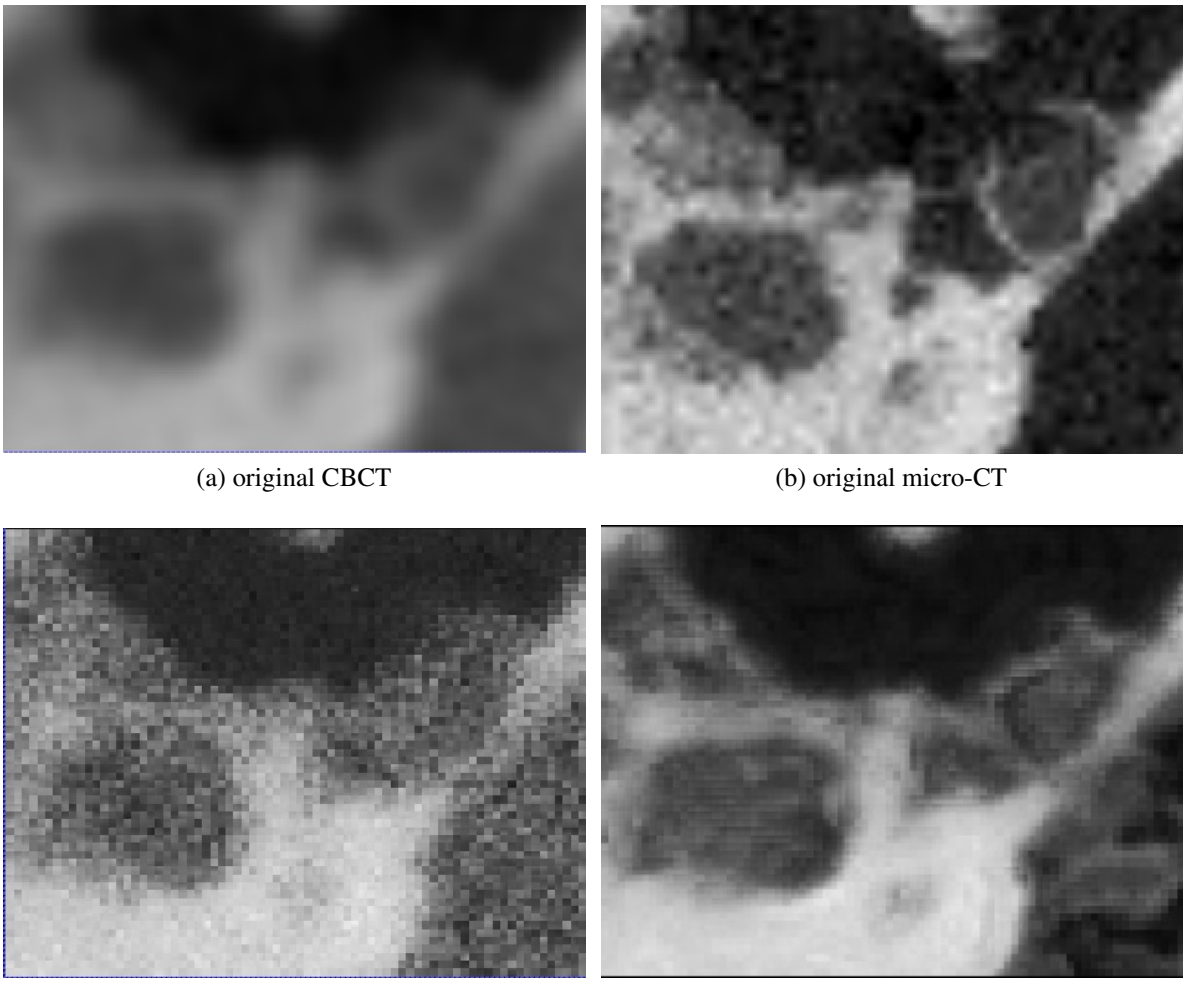


Figure 4.1: The complete pipeline of the proposed approach for enhancing CBCT image. During training, the original CBCT image is aligned to its related micro-CT image. Features are extracted from the CBCT by image patches, and intensities from related image patches from the micro-CT are provide. The mapping from image features to intensities is learned during the training phase.



(a) original CBCT

(b) original micro-CT

(c) "short range" prediction

(d) enhanced CBCT

Figure 4.2: Results of supervised-learning based CBCT image enhancement. Image features are extracted from (a) original CBCT image, and used to produce an enhanced version (d), that presents sharper and more clear structures, as compared to the original high-resolution micro-CT image (b). For demonstration purposes, we report results obtained using features extracted from a short range (i.e. small window size) (c), indicating the ability of the model to learn and utilize local structural information for the prediction process.

suboptimal prediction results. Conversely, employing an excessively large window size yields smoother but inaccurate prediction results. This result reflects the ability of the prediction model to learn local structural information that leverages the prediction of the image content at lower resolution scales. In the next section, we present preliminary results of segmentation of the facial nerve in the framework of minimally invasive cochlear implantation surgical planning, using the enhanced CBCT as input for the dedicated software tool OtoPlan.

4.4.1 Facial Nerve Segmentation

OtoPlan features a semi-automatic segmentation of the facial nerve. Based on a GUI-based tool, the user selects landmarks that approximately lie on the facial nerve's midline. A panoramic visualization is then constructed and displayed to the user, which corresponds to an "unfolding" of the facial nerve into one single view. The selected landmarks are displayed and used to cast intensity sampling lines that are perpendicular to the approximate midline of the facial nerve. A threshold based scheme is then used by OtoPlan to find the facial nerve wall as initialization prior to manual adjustments. Figure 4.3 illustrates this part of the segmentation process.

As the image contrast from CBCT is relatively poor, this process can be daunting and prone to errors, which in turn necessitates manual corrections. We performed a preliminary evaluation by comparing segmentation results obtained using the original and the enhanced CBCT image. OtoPlan can then generate and export a mesh representation of the segmented facial nerve. We measured surface-tosurface distances between each generated mesh and the corresponding ground-truth generated mesh. Figure 4.4 shows a particular example result where distances are color-coded to visualize deviations from the ground-truth segmentation. Employing the enhanced CBCT image provides OtoPlan with a sharper and better delineated facial nerve wall that yields a more precise segmentation of the facial nerve (i.e. lower surface-to-surface distances with respect to the groundtruth).

4.5 Conclusions

Good characterization of the facial nerve is of crucial importance for a safe planning of cochlear implantation interventions. Although CBCT-based imaging provides means to image the facial nerve in patients, its relatively low image contrast hinders a precisely definition of the facial nerve wall. In this work we proposed a machine-learning based approach that uses a supervised learning paradigm to learn the mapping between low-, and high-resolution imaging of the facial nerve. The approach relies on a multi-output

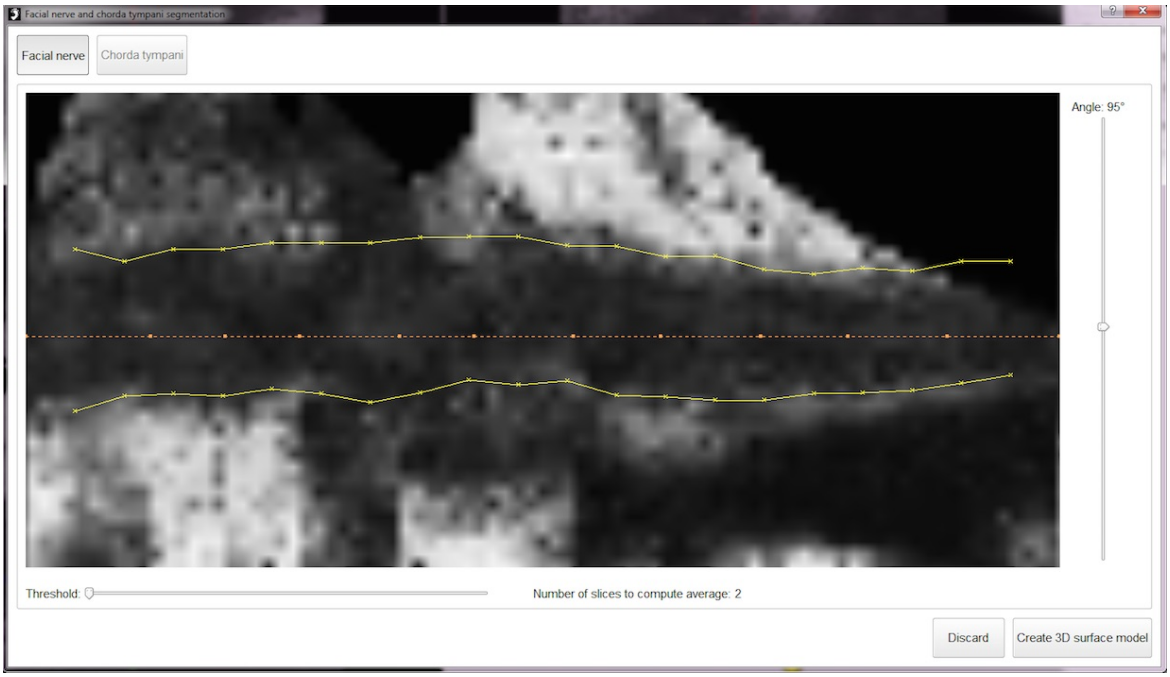
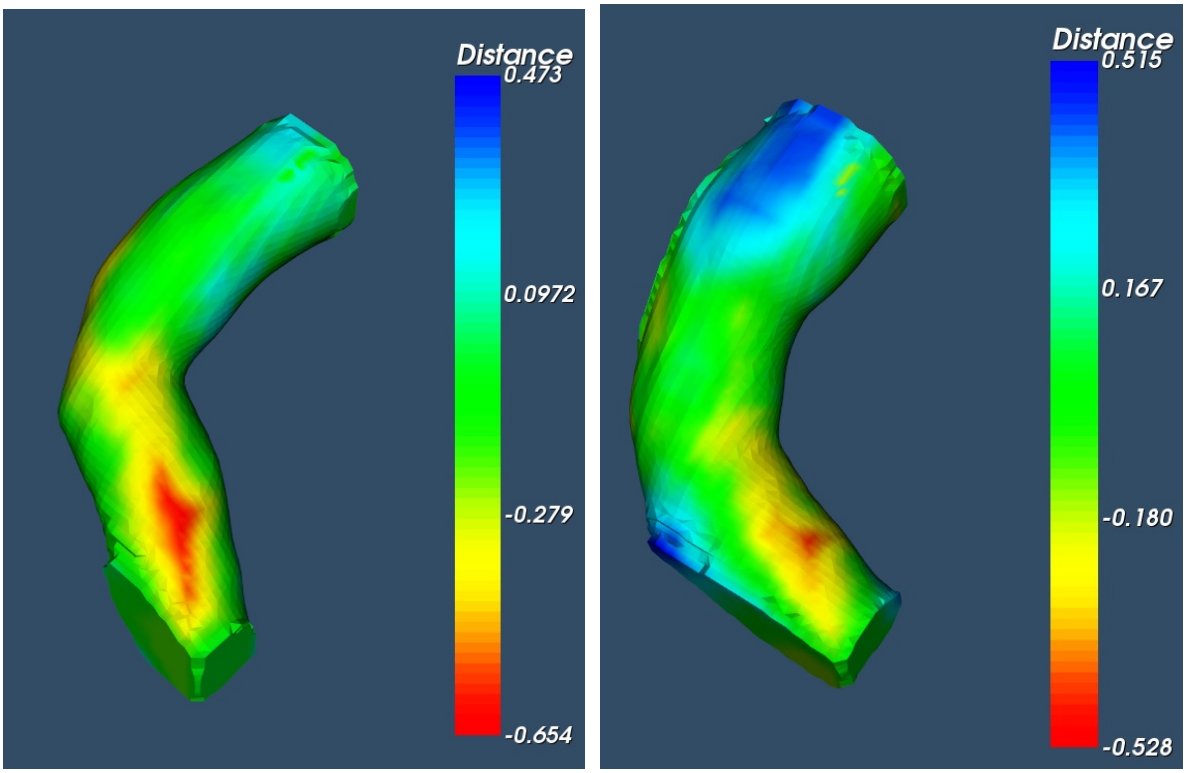


Figure 4.3: Panoramic view for semi-automatic segmentation of the facial nerve. The user selects a set of landmarks that approximately correspond to the middle line of the facial nerve. A threshold based scheme is then used to cast sampling perpendicular in order to find the facial nerve wall (above and below the middle line). Due to the low contrast quality of the CBCT image, manual correction of each point is commonly required. In this example, we illustrate the enhanced CBCT image.



(a) Original CBCT (b) Enhanced CT

Figure 4.4: Surface-to-surface distances from the ground-truth segmentation to OtoPlan segmentations generated using the original and enhanced CBCT image. Colormap encodes distances and best viewed in electronic version.

regression forest and image features, extracted from the clinically available CBCT image, to perform voxel intensity prediction at the micro-CT image resolution level. Preliminary results on CBCT images of the facial nerve show the ability of the proposed approach to enhance the imaging information by performing a prediction of voxel intensity information at the equivalent micro-CT resolution level. These first results also show the potential of the proposed approach to assist state-of-the-art cochlear surgical planing software, such as OtoPlan, to segment the facial nerve more precisely.

We report similar findings aligning with the literature on supervised-learning based image quality transfer [3], where local structural information from the low-resolution image was shown to convey information that can be used to predict localised voxel intensity information at the high-resolution image level. Our experiments also suggest the advantage of using a multi-output regression forest, in contrast to a single-output regression forest, in order to promote the learned local structural information.

The proposed approach presents some limitations. As for other supervised-learning based approaches, it is important to have a database of samples that characterises the expected variability of a population. Secondly, the mapping learned between low-, and high-resolution is specific to the imaging parameters used for the training database, and therefore a new model is potentially needed in case these parameters are modified. This can be circumvented, for instance, by designing imaging features that are independent of the energy parameters used for CT devices, or by designing compensation strategies based on an imaging phantom.

Further comprehensive evaluations and comparisons are needed, especially against other previously proposed super resolution approaches, as well as a more comprehensive quantitative evaluation on a larger dataset. Other future work will include an attempt to combine this approach with a fully automatic segmentation of the facial nerve that uses shape priors, as proposed by others [108], but that does not employ computationally expensive non-rigid registration techniques.

Chapter 5

Facial Nerve Segmentation

This Chapter has been submitted as a journal article.[1]

5.1 Abstract

Facial nerve segmentation is of considerable importance for pre-operative planning of cochlear implantation. However, it is strongly influenced by the relatively low resolution of the cone-beam computed tomography (CBCT) images used in clinical practice. In this paper, we propose a super-resolution classification method, which refines a given initial segmentation of the facial nerve to a sub-voxel classification level from CBCT/CT images. The super-resolution classification method learns the mapping from low-resolution CBCT/CT images to high-resolution facial nerve label images, obtained from manual segmentation on micro-CT images. We present preliminary results on dataset, 15 ex-vivo samples scanned including pairs of CBCT/CT scans and high-resolution micro-CT scans, with a Leave-One-Out (LOO) evaluation, and manual segmentations on micro-CT images as ground truth. Our experiments achieved a segmentation accuracy with a Dice coefficient of 0.818 ± 0.052, surface-to-surface distance of $0.121 \pm 0.030mm$ and Hausdorff distance of $0.715 \pm 0.169mm$. We compared the proposed technique to two other semi-automated segmentation software tools, ITK-SNAP and GeoS, and show the ability of the proposed approach to yield sub-voxel levels of accuracy in delineating the facial nerve.

[1]Ping Lu, Livia Barazzetti, Vimal Chandran, Kate Gavaghan, Stefan Weber, Nicolas Gerber, Mauricio Reyes, "Highly accurate Facial Nerve Segmentation Refinement from CBCT/CT Imaging using a Super Resolution Classification Approach", IEEE Transactions on Biomedical Engineering. 2017 Apr 25. doi: 10.1109/TBME.2017.2697916.

Keywords: Facial nerve, Segmentation, Cochlear implantation, Supervised Learning, Super-Resolution, CBCT, micro-CT.

5.2 Introduction

Cochlear implantation is a conventional treatment that helps patients with severe to profound hearing loss. The surgical procedure requires drilling of the temporal bone to access the cochlea. In the traditional surgical approach, a wide mastoidectomy is performed in the skull to allow the surgeon to identify and avoid the facial nerve, whose damage can cause temporal or permanent ipsilateral facial paralysis. In order to minimize invasiveness, a surgical robot system has been developed to perform highly accurate and minimally invasive drilling for direct cochlear access [9]. The associated planning software tool, OtoPlan [45], allows the user to semiautomatically segment structures of interest and define a safe drilling trajectory. The software incorporates a semiautomatic and dedicated method for facial nerve segmentation using interactive centerline delineation and curved planar reformation [45].

The surgical planning for minimally invasive cochlear implantation is affected by the relatively low resolution of the patient images. Imaging of the facial nerve is typically performed using CT or CBCT imaging with a resolution in the range of $0.15-0.3mm$ slice thickness, and a small field of view $80-100mm$ temporal bone protocol. This resolution is comparatively low in regards to the diameter of the facial nerve, which lies in the range of $0.8-1.7mm$.

Atlas-based approaches combined with level-set segmentation have been proposed before to segment the facial nerve in adults [108] and pediatric patients [126]. These methods automatically segment the facial nerve, with reported average and hausdorff accuracies in the ranges of $0.13-0.24mm$ and $0.8-1.2mm$, respectively. This reported accuracy is similar to other approaches, such as OtoPlan [45] or NerveClick [151], where a semi-automatic statistical model of shape and intensity patterns was developed with a reported RMSE accuracy of $0.28\pm0.17mm$. Since for the facial nerve a margin of up to $1.0mm$ is available and an accuracy of at least $0.3mm$ (depending on the accuracy of the navigation system) is required [131], an accurate facial nerve segmentation is crucial for an effective cochlear implantation surgical plan.

Super-resolution methods have been presented in computer vision related tasks to reach sub-voxel accuracy in regression problems, where the goal is to reconstruct a high-resolution image from low-resolution imaging information. Most of such methods employ linear or cubic interpolation [140], but are sub optimal for CBCT/CT images of the facial nerve, due to their SNR and local structural variability. In a recent study [132], a Random Forest

based regression model was used to perform upsampling of natural images. Similarly, in [3] a supervised learning algorithm was used to generate diffusion tensor images at super-resolution (i.e. upscaling from $2 \times 2 \times 2mm$ to $1.25 \times 1.25 \times 1.25mm$ resolution). Recently, in [35], a super-resolution convolutional neural network (SRCNN) learns an end-to-end mapping between the low- and high-resolution images. In [111] a super-resolution (SR) approach reconstructs high resolution 3D images from 2D image stacks for cardiac MR imaging, based on a convolutional neural network (CNN) model.

In the present clinical problem, we are concerned with the delineation of the facial nerve for cochlear implantation planning. Hence, as opposed to other super-resolution schemes, here we propose a super-resolution *classification* method for accurate segmentation refinment of the facial nerve.

We adopted a supervised learning scheme to learn the mapping between CBCT/CT images to high-resolution facial nerve label images, obtained from manual segmentations on micro-CT images. Here we coin the method super-resolution classification (SRC). The proposed approach then employs SRC to refine an initial segmentation provided by OtoPlan [45] to generate accurate facial nerve delineations.

In the following sections we present a description of the image data and the proposed algorithm, followed by segmentation results on test cases, and a comparison with two other general-purpose segmentation (ITK-SNAP, GeoS) software tools used to perform segmentation refinement.

5.3 Materials and Methods

This section describes the image data used to train and test the proposed SRC algorithm. Figure 5.1 shows an overview of the complete pipeline, composed of two phases for training and testing. During training, image upsampling, image preprocessing, registration to micro-CT images, and building of a classification model based on extracted features, are performed. During testing, the input CBCT image is upsampled, preprocessed, and features are extracted to perform super-resolution classification.

5.3.1 Image Data

We developed and tested our approach on a database of 15 patient cases, comprising 7 pairs of CBCT and micro-CT images, and 8 pairs of CT and micro-CT images of temporal bones. The CBCT temporal bones were extracted from four cadaver heads in the context of an approved clinical study on cochlear implantation [161]. The CBCT were obtained with a Planmeca

ProMax 3D Max with $100 \times 90 mm^2 FOV$. Micro-CT was performed with a Scanco Medical μCT 40 with $36.9 \times 80 mm^2 FOV$. For each sample, a CBCT ($0.15 \times 0.15 \times 0.15 mm^3$) and micro-CT ($0.018 \times 0.018 \times 0.018 mm^3$) scan was performed. The set of 8 pairs of CT and Xtreme CT images of temporal bones were obtained with a CT imaging (Siemens SOMATOM Definition Edge) and a temporal bone imaging protocol with parameters: $120 kVp$, $26 mA$, $80 mm FOV$. The spatial resolution of the scanned CT images was $0.156 \times 0.156 \times 0.2 mm^3$. The Xtreme CT ($0.0607 \times 0.0607 \times 0.0607 mm^3$) scans were obtained with a Xtreme CT imaging (SCANCO Medical). We note that cadaver images are similar to clinical images of the facial nerve, enabling the evaluation of our method with cadaver images. The image volume size ranges around $70 \times 80 \times 110$ voxels from CBCT and $60 \times 55 \times 60$ voxels from CT .

To create ground truth datasets, manual segmentations of the facial nerve on micro-CT images was performed following the segmentation protocol presented in [45], and were verified by experts using Amira 3D Software for Life Sciences version 5.4.4 (FEI, USA) [142]. The experts verifying the manual ground-truth are two senior biomedical engineers trained in the cochlear anatomy and with years of experience in manual segmentation of the cochlear structures.

5.3.2 Preprocessing

To create the supervised based machine learning model and to evaluate the approach on ground truth data, derived from micro-CT images, we rigidly aligned the pairs of CBCT and CT and micro-CT images using Amira (version 5.4.4) via the normalized mutual information method [143] and elastix [79] [137] with the advanced normalized correlation metric (see Appendix 7.3). For the sake of clarity, we refer for the rest of the paper to both CBCT and CT as CBCT/CT to indicate that the operations are applied to both.

Intensity normalization

We normalize the intensities of the CBCT/CT images by histogram matching, with a common histogram as a reference. Since we are computing the histogram only on a ROI (described below in section 5.3.2) to match the range of intensities being targeted for the facial nerve, we avoid the effect of background voxels and hence, there is no need to set an intensity threshold for the histogram matching.

CBCT/CT and micro-CT image alignment (only training phase)

In order to learn the mapping between CBCT/CT and micro-CT images we rigidly aligned the pairs of CBCT/CT and micro-CT images. Due to the fact that the diameter of the facial nerve lies in the range of $0.8-1.7mm$ and the facial nerve is only imaged across approximately at $5-11$ slices of CBCT/CT ($0.15\times0.15\times0.15mm^3$), we manually initialized a rigid registration based on landmarks defined by screws implanted in the specimens for patient-to-image registration, as presented in [161], followed by an automated rigid registration in Amira (version 5.4.4) using normalized mutual information metric. A second rigid registration was performed between the transformed micro-CT image and the CBCT image, using elastix with advanced normalized correlation metric. We observed that in practice this pipeline resulted in an improved robustness and accuracy, as opposed to performing a single registration. We also remark that no change of resolution is performed when registering the micro-CT image to the CBCT image (as typically is the case for image registration tasks). The set of sought transformations are then applied to the ground truth image in order to map them onto the CBCT image space.

Region of interest selection (ROI)

Since the main focus of the method is to obtain sub-voxel accuracy of the facial nerve border, and to reduce computational costs, we adopted a band-based region of interest selection strategy. Here we use the segmentation results from OtoPlan as initial segmentation to be refined through Super Resolution Classification. From the preliminary OtoPlan segmentation of the CBCT/CT image, a region-of-interest is created via a combination of erosion and dilation morphological operations. The region of interest, on which the super-resolution classification takes place, corresponds to the arithmetic difference between the dilated and eroded label images. In practice, a 16 and 24 voxel structuring element ($0.3mm$ and $0.4mm$ respectively on each side, which effectively translates as an additional two times magnitude of the accuracy error reported by other approaches)) was tested on the upsampled CBCT/CT images.

5.3.3 Super-Resolution Classification (SRC)

This section describes the steps for CBCT/CT upsampling, the feature extraction and the classification model building.

CBCT/CT upsampling

Similar to [132], we perform an upsampling of the CBCT/CT image to the target resolution in order to combine features extracted from the upsampled and the original image. In this study we employed a B-spline upsampling scheme. However, other interpolation schemes, such as linear or cubic, can be used since as classification results were not sensitive to this choice.

Feature Extraction

We employ texture-based features derived from first-order statistics, percentiles and Grey Level Co-occurrences Matrix (GLCM) [54] [117] [27], which are only extracted on the computed region of interest. First-order statistics [141] and percentile features [38] [47] are computed at original and upsampled resolutions, while GLCM features are computed only on patches from the upsampled image. This is supported by direct testing of GLCM features derived from both the original and upsampled images with poorer results (in terms of all evaluated metrics) than using only GCLM features extracted from the upsampled image. This can also be explained by the fact that the much larger size of voxel-wise GLCM features (in comparison to the other imaging features). We remark that through direct testing of GLCM features derived from both original and upsampled CBCT/CT images, GLCM features extracted from the original CBCT/CT image do not contribute as much as those extracted from the upsampled image.

First-order statistics Mean, standard deviation, minimum, maximum, skewness and kurtosis of voxel intensities are computed for each image patch of the CBCT/CT and upsampled CBCT/CT image.

Percentiles From each image patch of the CBCT/CT and upsampled CBCT/CT image, the 10th percentile, 25th percentile, 50th percentile, 75th percentile, 80th percentile, 95th percentile of the intensity distribution, are used as features.

The Grey Level Co-occurrences Matrix (GLCM) The Grey Level Co-occurrences Matrix (GLCM) is a second-order statistical texture that considers two-voxels relationship in an image. Following [27], we adopted 8 GLCM features: inertia, correlation, energy, entropy, inverse difference moment, cluster shade, cluster prominence, haralick correlation. Mean and variance of each feature with 13 independent directions in the center voxel of each image patch are calculated. Hence, 16 features of GLCM were calculated in the upsampled CBCT/CT image.

Classification model – Training phase

Given a training set $\{\langle X_i, Y_i\rangle | i = 1, ..., N\}$ of CBCT/CT and micro-CT aligned pairs of images, we extract from each i_{th} image patch, a feature vector $X_i = (v_1, ...v_n) \in \mathbb{X}$ and responses $y \in \{0, 1\}$, which describes the background/foreground label of the center voxel over a grid of C voxels. Then, a function $\hat{y} : \mathbb{X} \mapsto y$ from a space of features $\mathbb{X}$ to a space of responses y is constructed. The mapping is cast as a *classification* problem.

As classification model, we adopted extremely randomized trees (Extra-Trees) [46], which is an ensemble method that combines the predictions of several randomized decision trees to improve robustness over a single estimator. Extra-Trees have shown to be slightly more accurate than Random Forests (RF) and other tree-based ensemble methods [46]. During the training phase of Extra-Trees, multiple trees are trained and each tree is trained on all training data. Extra-Trees randomly selects without replacement, K input variables $\{v_1, ...v_k\}$ from the training data. Then, a cutpoint s_i is randomly selected, ruled by a splitting criteria $[v_i < s_i]$, for each selected feature within the interval $[v_i^{min}, v_i^{max}]$. Among the K candidate splits, the best split is chosen via normalization of the information gain [93]. We note that in our experiments, and in order to reduce irrelevant features [46], the number of input variables K is set to the size of the input feature vector n.

Classification model – Prediction phase

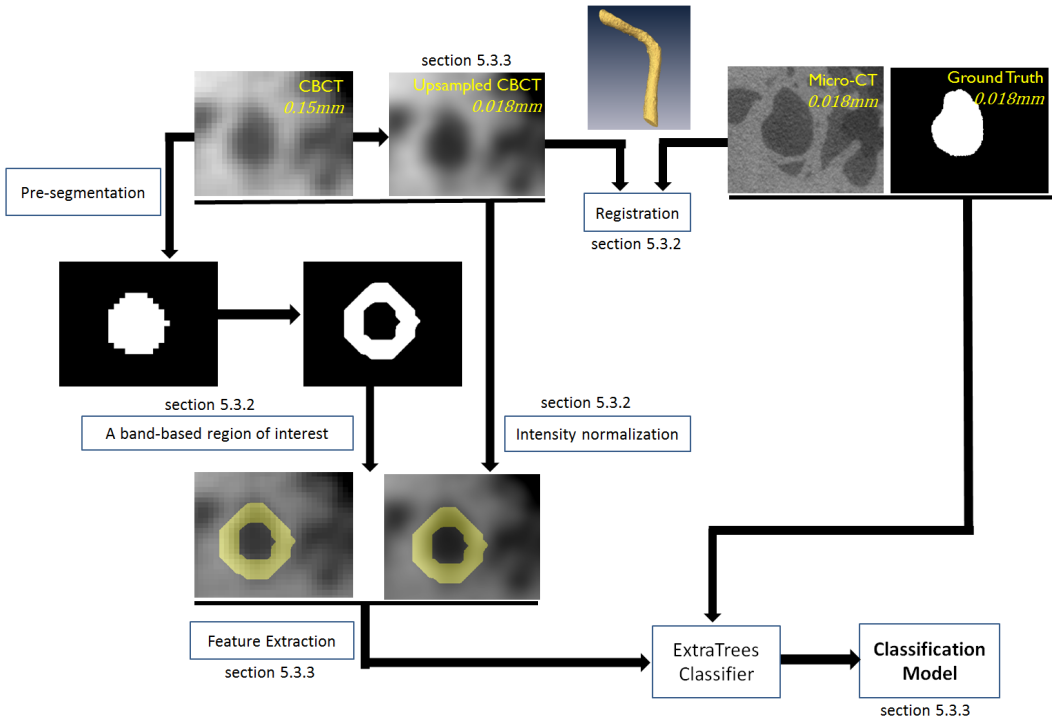
During testing, the CBCT/CT image is pre-processed through image intensity normalization (using the same reference image as for the training phase). Image features in a band of interest (results reported using a band size of 16 and 24 voxels) are extracted from the original and upsampled CBCT/CT image, and passed through the Extra-Trees classification model. The computed output corresponds to the label of the central voxel from the extracted patch.

Postprocessing

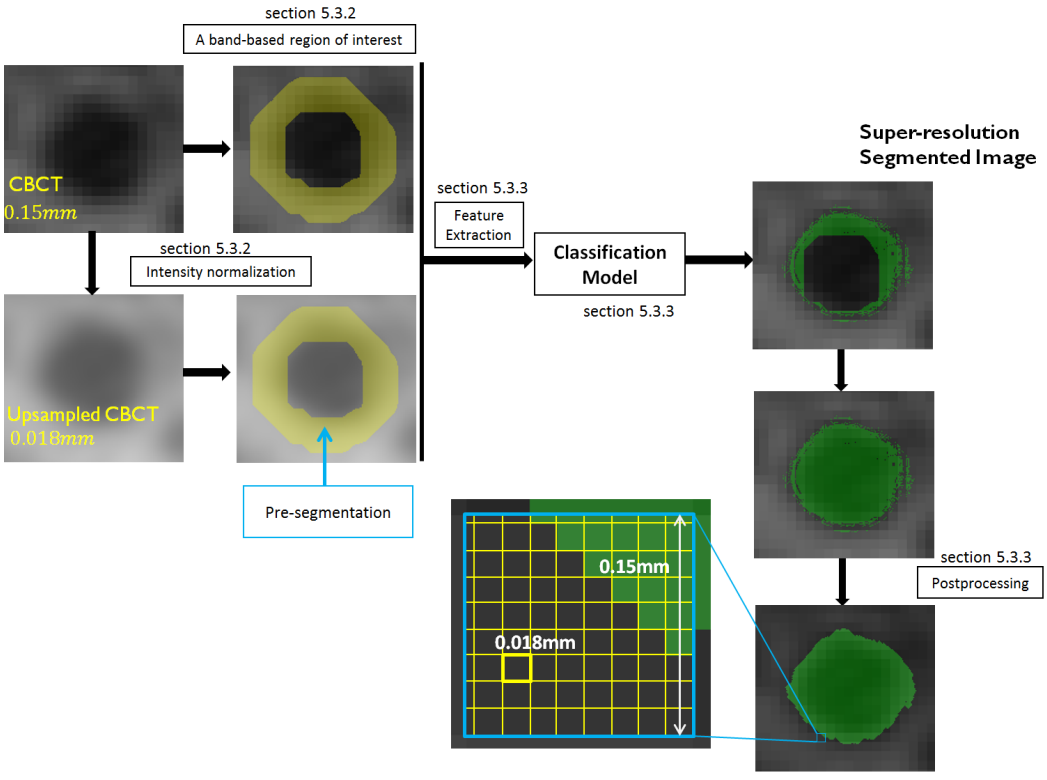
The refined segmentation is regularized in order to remove spurious and isolated segmented regions. In this study we adopted a basic regularization scheme based on erosion (kernel size=16 or 24) and dilation (kernel size=16 or 24) morphological operations.

5.4 Experimental Design

A Leave-One-Out (LOO) cross-validation study was carried out to evaluate the accuracy of the proposed super-resolution segmentation approach. The idea of LOO is to split data into train and test sets. One image data is chosen as a test set while the remaining data are used



(a) Training phase



(b) Testing phase

Figure 5.1: Proposed super-resolution classification (SRC) approach, described for the training (a), and testing phase (b). During training, the original CBCT/CT image is aligned to its corresponding micro-CT image. OtoPlan [45] is used to create an initial segmentation, from where a region-of-interest (ROI) band is created. From the original and upsampled CBCT/CT images, features are extracted from the ROI-band to build a classification model, which is used during testing to produce a final super-resolution segmented image. The zoomed square on the segmented super-resolution image shows on one voxel of the CBCT/CT image, the more accurate segmentation yielded by SRC.

for training. This method is repeated until every image data has been tested and evaluated using the following evaluation metrics.

5.4.1 Experimental detail

The upsampling of the CBCT/CT images was performed in Amira with a B-spline interpolation kernel. For computation of features, patches of size $5 \times 5 \times 5$ were extracted on the original and upsampled CBCT/CT images. Feature extraction, morphological operations to create the ROI, and intensity normalization was performed with the Insight-Toolkit version 4.4.1 [73], and classification was completed with Scikit-learn: Machine Learning in Python [120]. Default parameters were used for the Extra-Trees classifier (see Appendix 7.4).

5.4.2 Segmentation initialization with OtoPlan

As input to the segmentation refinement step with SRC we utilized OtoPlan [45] to obtain an initial segmentation from where the ROI bands are extracted, and then classified.

In OtoPlan, the centerline of the facial nerve is manually drawn and the borders are automatically defined by the tool. After this step, the facial nerve border can be manually modified by dragging contours of the facial nerve.

Based on this initial segmentation we extracted a ROI with two different sizes (section 5.3.2), referred as to band 16 and band 24, to indicate 16 and 24 voxels band size. The rationale behind is to analyze the sensitivity of SRC to different band sizes, as well as to analyze a potential dependency between the accuracy of the initial segmentation and the performance of SRC.

5.4.3 Evaluation metrics

1. **Hausdorff Distance (HSD)**
   This metric measures the Hausdorff distance [67] from the ground truth surface to its nearest neighbor in the segmented surface.

$$H(A,B) = max(h(A,B), h(B,A)), \tag{5.1}$$

   where

$$h(A,B) = \max_{a \in A} \{ \min_{b \in B} ||a - b|| \}. \tag{5.2}$$

The function $h(A,B)$ is the Hausdorff distance between two surfaces A (manually segmented ground truth image) and B (automatically segmented image). The term $||a-b||$ is the distance between point a and b (in our case the Euclidean distance). The smaller the Hausdorff distance value, the more accurate the segmentation performs in terms of facial nerve border definition. The Hausdorff distance is useful as an indicative of worst case scenario of the delineation.

2. **Root Mean Squared Error (RMSE)**
   The RMSE is calculated by the square root of the Mean Squared Error (MSE).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(a-b)^2}{m}}, \tag{5.3}$$

   where $a \in A$, $b = \min_{b^* \in B} ||a-b^*||$, and m corresponds to the number of surface points used to compute RMSE.

3. **Average Distance (AveDist)**

$$AveDist(A,B) = max(d(A,B), d(B,A)), \tag{5.4}$$

   where

$$d(A,B) = \frac{1}{m}\sum_{a \in A} \min_{b \in B} ||a-b||. \tag{5.5}$$

   The smaller the value of the average distance the better the accuracy of the facial nerve segmentation is.

4. **Positive Predictive Value (PPV)**

$$PPV = \frac{TP}{(TP+FP)}, \tag{5.6}$$

   where TP stands for true positive — the number of correctly segmented facial nerve voxels— and FP stands for false positive, the number of wrong segmented voxels (i.e. segmenting background voxels as facial nerve voxels).

5. **Sensitivity (SEN)**

$$SEN = \frac{TP}{(TP+FN)}, \tag{5.7}$$

   where FN stands for the number of wrong labeled background voxels (i.e., segmenting facial nerve voxels as background).

6. **Specificity (SPC)**

$$SPC = \frac{TN}{(TN+FP)}, \tag{5.8}$$

where TN stands for the number of correctly segmented background voxels.

7. **Dice Similarity Coefficients (DSC)**
This validation metric measures the accuracy of segmentation in spatial overlap. The DSC [33] is calculated via

$$DSC(A,B) = \frac{2|A \cap B|}{|A|+|B|}. \tag{5.9}$$

A DSC value of 1 indicates the segmentation fully overlaps with the ground truth (i.e. perfect segmentation), while a DSC value of 0 indicates no overlap beween the segmentation and the ground truth segmentation.

5.4.4 Evaluation

In this section we present segmentation results separately for CBCT and CT images with the intention to show the performance of the proposed approach on two different scan types. We remark that the adopted LOO evaluation strategy employs the entire set of clinical CBCT/CT scans for the training phase.

Experiment 1: Segmentation results on CBCT(training and testing with LOO)

We compared the proposed SRC method with the segmentation software GeoS and ITK-SNAP. We employed ITK-SNAP (version 3.4.0) [164] and its Random-Forest-based generation of speed images, which relies on defining brushes on the foreground and background areas of the facial nerve. The number of brushes was found empirically via trial-and-error with the main criteria of yielding robust segmentation results. In practice this resulted in approximately four brush strokes per image, and eight bubbles for contour initialization. No extensive search of optimal placement of brushes was conducted in order to keep the experiments to the typical usage scenario of the tool. Similar procedure was conducted for the GeoS tool (version 2.3.6) [30], a semi-automatic tool based on brush strokes and Random Forest supervised learning. On average over fifteen brush strokes were used for GeoS, with no further improvements observed beyond this number.

For both software tools, brush strokes were defined on the ROI-band (16 or 24 voxels) on background and foreground areas (c.f. section 5.3.2).

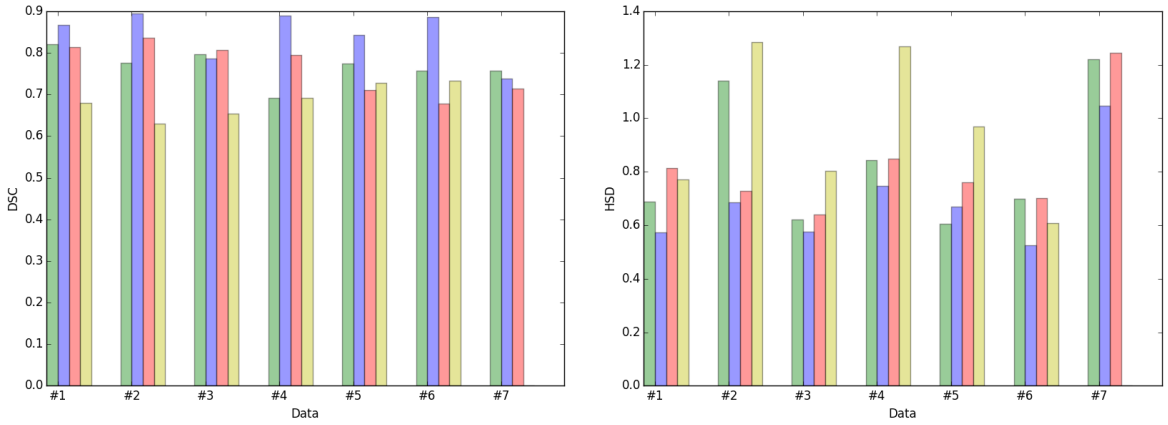
In order to compare DSC values among segmentation results and the ground truth (produced at micro-CT resolution), we resampled the results from the tools to the resolution of the ground truth using nearest interpolation. Second, we converted the segmentation results to surfaces [87] and computed average and Hausdorff distances.

Figure 5.2 shows the facial nerve segmentation results for each sample of the CBCT dataset for the proposed SRC method, and ITK-Snap and GeoS. From the DSC values and the Hausdorff distances (Figures 5.2a and 5.2b), it can be observed that the proposed method is robust and provides a higher average DSC and a lower Hausdorff distance than the other methods. From Figure 5.2d and Figure 5.2e it can be observed that that overall ITK-SNAP and GeoS tend to undersegment the CBCT cases. Conversely, SRC did not show a potential bias towards over-, or under-segmentation.

Table 5.1 summarizes the comparative results between the proposed approach and GeoS and ITK-SNAP, with two different band-based region of interest. It can be observed that in comparison to ITK-SNAP and GeoS, the proposed SRC method is more accurate and robust to an increase of the band size. Particularly, GeoS resulted to be less robust to an increase of the band size, as described by the increase variance of the metrics. The proposed SRC method achieved an average DSC value of 0.843, a mean Hausdorff distance of 0.689*mm*, and a sub-voxel average distance accuracy of 0.156*mm*. Regarding the tested segmentation tools, GeoS yielded the lowest DSC value among the evaluated approaches (average DSC of 0.686), followed by ITK-SNAP with an average DSC value of 0.765. In terms of distance metrics, the average Hausdorff metric for GeoS and ITK-SNAP was 0.951*mm* and 0.819*mm*, respectively. Using a two-tailed t-test and Wilcoxon signed ranks tests, statistically significantly greater results than GeoS and ITK-SNAP were obtained ($p < 0.05$, Bonferroni corrected) for the dice, average distance and RMSE metrics. Figure 5.3 shows an example result, put in the context of the original and high-resolution ground-truth, while Figure 5.4 shows example results for all tested approaches. It can be observed that the proposed approach yields a more precise delineation than the other tested methods. Particularly, the postprocessing step based on simple morphological removes any potential holes and isolated small regions.

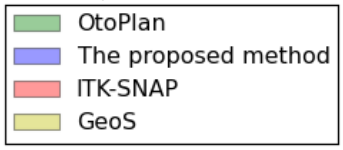
Experiment 2: Segmentation results on CT

Figure 5.5 and Table 5.2 summarize the comparative results on the CT database, between the proposed approach and GeoS and ITK-SNAP for two different band-based sizes. Similar to the results on CBCT cases, it is observed that the proposed SRC method is superior to ITK-SNAP and GeoS, and is more robust to the band size. The proposed method achieved an average DSC value of 0.797, a mean Hausdorff distance of 0.739*mm*, and a sub-voxel average distance accuracy of 0.129*mm*. GeoS yielded the lowest DSC value among the

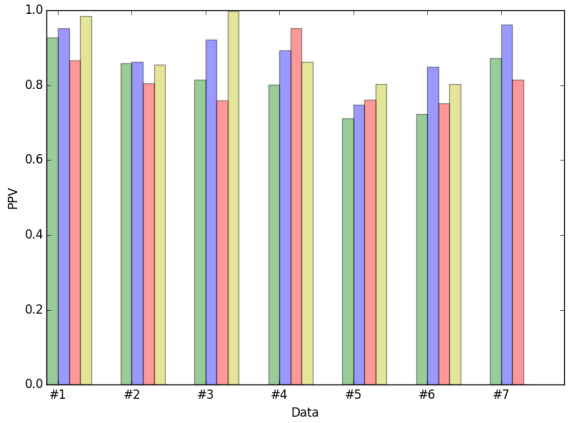


(a) Dice Similarity Coefficients

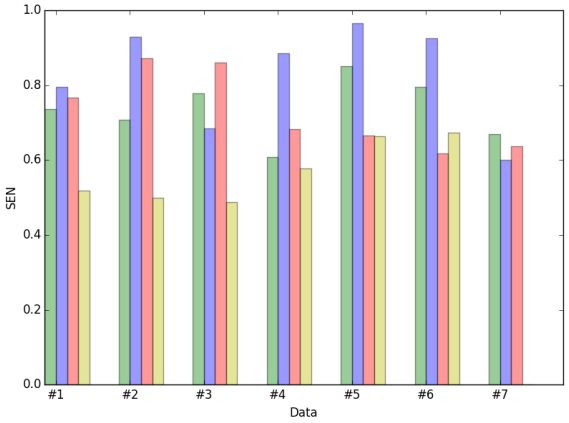
(b) Hausdorff Distance



(c)



(d) Positive Predictive Value



(e) Sensitivity

Figure 5.2: Evaluation on CBCT cases between the proposed super-resolution segmentation method and GeoS (version 2.3.6) and ITK-SNAP (version 3.4.0). Average values for DSC (a), Hausdorff (b), Positive Predictive Value (d), and Sensitivity (e). Case number 7 could not be segmented via GeoS. Note: best seen in colors.

	CBCT dataset with band16			CBCT dataset with band24		
Method	Proposed method (SRC)	GeoS	ITK-SNAP	Proposed method (SRC)	GeoS	ITK-SNAP
Dice	**0.843±0.055(0.866)**∗	0.686±0.037(0.686)	0.765±0.058(0.795)	**0.822±0.062(0.847)**	0.578±0.123(0.547)	0.732±0.099(0.777)
AveDist	**0.112±0.034(0.100)**∗	0.296±0.058(0.193)	0.196±0.052(0.205)	**0.131±0.032(0.121)**	0.436±0.137(0.461)	0.197±0.064(0.164)
RMSE	**0.156±0.038(0.144)**∗	0.345±0.062(0.367)	0.247±0.063(0.248)	**0.186±0.034(0.180)**	0.493±0.134(0.526)	0.237±0.062(0.212)
Hausdorff	**0.689±0.163(0.670)**∗	0.951±0.253(0.886)	0.819±0.185(0.760)	**0.747±0.117(0.736)**	1.309±0.207(1.364)	0.744±0.090(0.708)

Table 5.1: Quantitative comparison on CBCT cases between our method and GeoS and ITK-SNAP, for band sizes 16 (left) and 24 (right). Dice and surface distance errors (in *mm*). The measurements are given as mean ± standard deviation (median). The best performance is indicated in boldface. The '∗' indicates that SRC results are statistically significantly greater ($p < 0.05$) than GeoS and ITK-SNAP using a two-tailed t-test and Wilcoxon signed ranks tests.

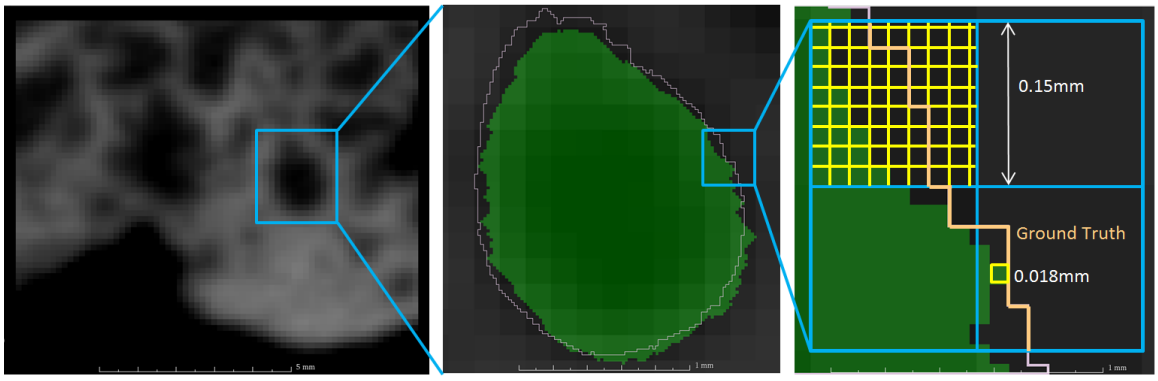


Figure 5.3: Example results for the proposed super-resolution segmentation approach. From left to right: Original CBCT image with highlighted (in blue) facial nerve, resulting segmentation and ground truth delineation (orange contour), and zoomed area describing SRC results on four corresponding CBCT voxels.

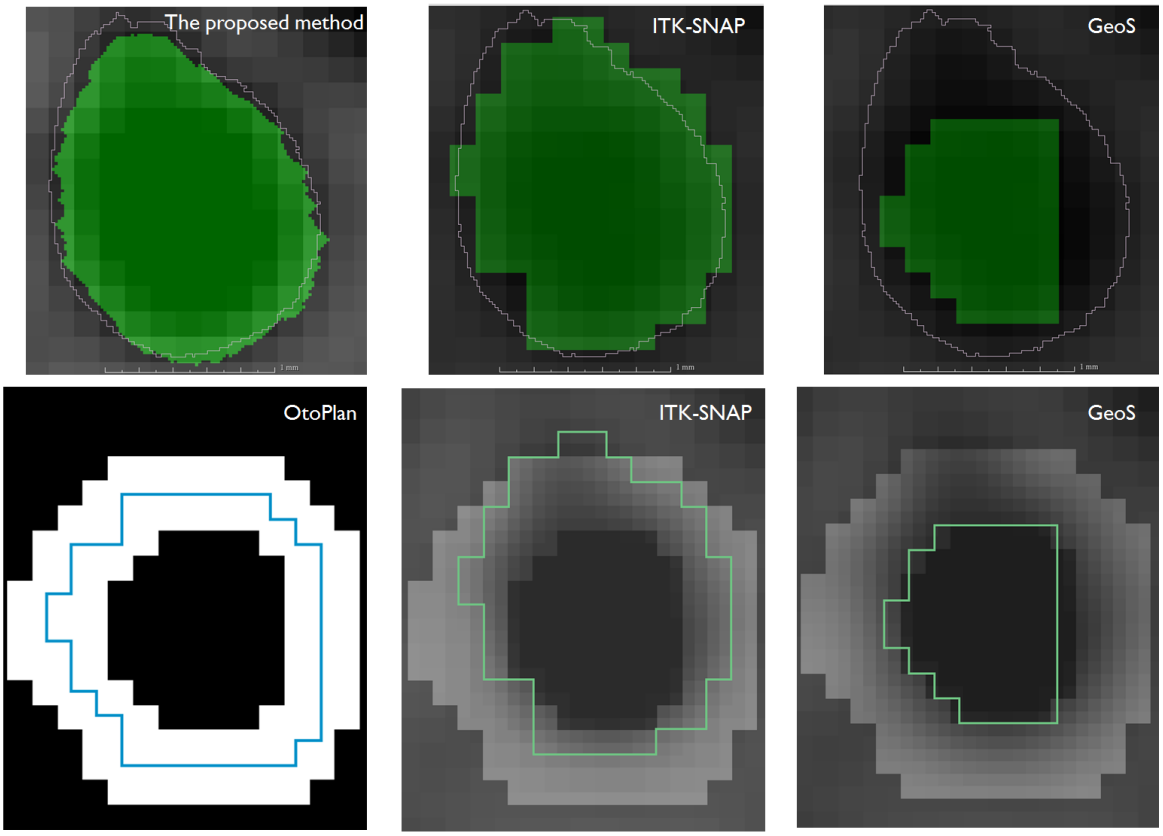
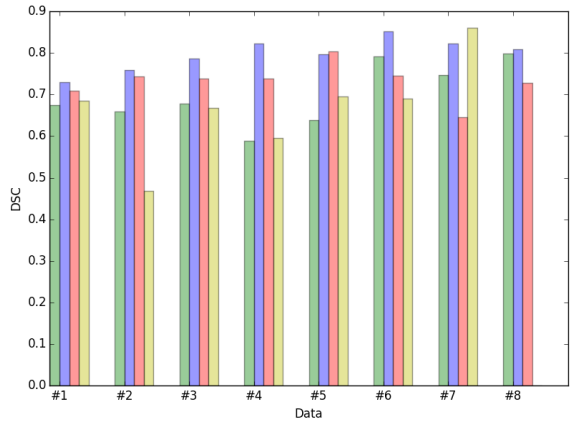
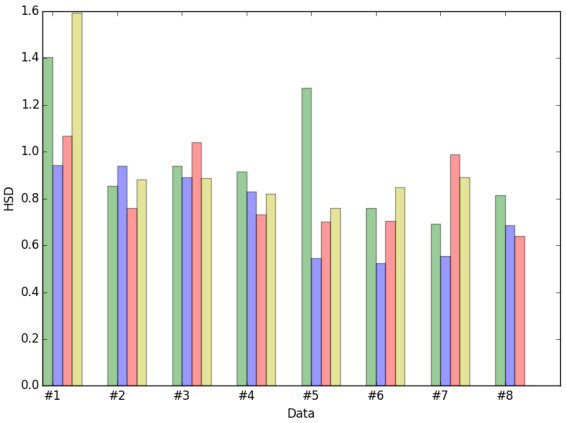


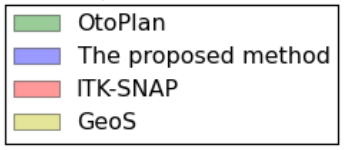
Figure 5.4: The facial nerve segmentation comparison on the original CBCT image between the proposed SRC method and other segmentation software — ITK-SNAP and GeoS. The ROI selection via band 16 from OtoPlan initial segmentation.



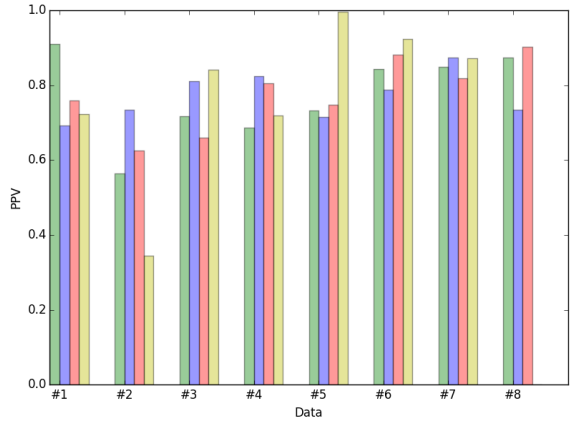
(a) Dice Similarity Coefficients



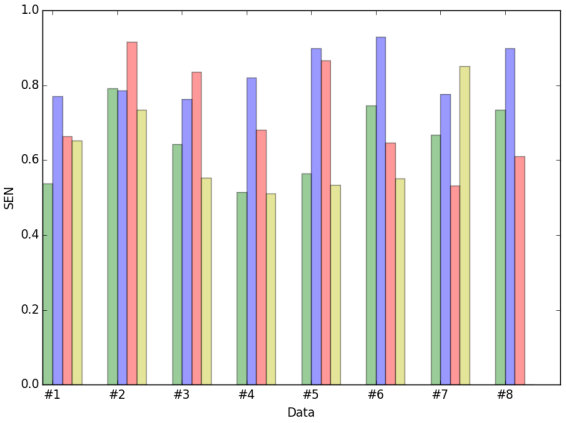
(b) Hausdorff Distance



(c)



(d) Positive Predictive Value



(e) Sensitivity

Figure 5.5: Evaluation on CT cases between the proposed super-resolution segmentation method and GeoS (version 2.3.6) and ITK-SNAP (version 3.4.0). Average values for DSC (a), Hausdorff (b), Positive Predictive Value (d), and Sensitivity (e). Case number 8 could not be segmented via GeoS. Note: best seen in colors.

	CT dataset with band16			CT dataset with band24		
Method	Proposed method (SRC)	GeoS	ITK-SNAP	Proposed method (SRC)	GeoS	ITK-SNAP
Dice	**0.797±0.036(0.802)**∗	0.666±0.109(0.685)	0.731±0.041(0.738)	**0.749±0.048(0.759)**	0.549±0.084(0.545)	0.744±0.049(0.734)
AveDist	**0.129±0.023(0.125)**∗	0.292±0.021(0.292)	0.194±0.036(0.204)	**0.156±0.027(0.155)**	0.484±0.181(0.433)	0.176±0.061(0.163)
RMSE	**0.177±0.033(0.178)**∗	0.353±0.030(0.341)	0.243±0.045(0.251)	**0.216±0.037(0.226)**	0.593±0.247(0.477)	0.229±0.073(0.229)
Hausdorff	**0.739±0.171(0.758)**∗	0.954±0.264(0.883)	0.829±0.162(0.746)	**0.854±0.168(0.901)**	1.524±0.732(1.087)	0.812±0.221(0.848)

Table 5.2: Quantitative comparison on CT cases between our method and GeoS and ITK-SNAP, for band sizes 16 (left) and 24 (right). Dice and surface distance errors (in *mm*). The measurements are given as mean ± standard deviation (median). The best performance is indicated in boldface. The '∗' indicates that SRC results are statistically significantly greater ($p < 0.05$) than GeoS and ITK-SNAP using a two-tailed t-test and Wilcoxon signed ranks tests.

evaluated approaches, followed by ITK-SNAP with an average DSC value of 0.731. In terms of distance metrics, the average Hausdorff metric for GeoS and ITK-SNAP was 0.954*mm* and 0.829*mm*, respectively. Using a two-tailed t-test and Wilcoxon signed ranks tests, statistically significantly greater results than GeoS and ITK-SNAP were obtained ($p < 0.05$, Bonferroni corrected) for the dice, average distance and RMSE metrics, see Table 5.2

5.5 Discussions

In this study we developed an automatic super-resolution facial nerve segmentation via a random forest Extra-Trees based classification framework that refines an initial segmentation of the facial nerve to sub-voxel accuracy. To our knowledge, this is the first attempt to perform super-resolution classification for facial nerve segmentation in CBCT/CT images by exploting imaging modalities featuring different resolution levels. Preliminary results, based on a leave-one-out evaluation on fifteen ex-vivo cases, suggest that the proposed method is able to classify the facial nerve with high accuracy and robustness. On a standard desktop computer, the learning phase is the most time-consuming part, requiring for our set-up around 2 hours. The testing phase (running on a new case) takes only 9 minutes. Given an input CBCT or CT image, the proposed pipeline start with an initial segmentation of the facial nerve region, which in this study was obtained via OtoPlan [45]. However, we remark that other approaches can be used to yield the initial segmentation (e.g. [108], ITK-SNAP [164]). A band ROI is then created from this initial segmentation and used by SRC to attain a highly accurate segmentation of the facial nerve in an automated fashion. Comparison with

other available segmentation tools, ITK-SNAP and GeoS, confirms the higher accuracy and robustness of the proposed SRC approach.

According to our experiments, better segmentation results are obtained when the features computed on both the original and upsampled CBCT images than with features extracted only from the original CBCT image. This is in agreement with recent findings in semi-supervised regression based image upscaling where features extracted from an initial upsampling has shown to yield better estimates of the sought high-resolution image [132]. This is motivated by the fact that the training phase is enriched by including model samples that stem from micro-CT labels (i.e. from the micro-CT ground-truth image) and corresponding imaging features approximated at micro-CT level by the upsampling step on the CBCT/CT images.

As described, GLCM features extracted from the original CBCT/CT image do not contribute as much as those extracted from the upsampled image. This behavior can be conceptually explained since GLCM features computed on the upsampled image describe textural patterns on a much localized 5^3 patch size that better correlates to the label of the central voxel, extracted from the micro-CT image. Conversely, GLCM features computed on the original CBCT/CT image covers a much larger spatial extent, and hence the described textural information correlates less to the label of the central voxel at micro-CT resolution. Interestingly, the role of features from first-order statistics and percentiles provide benefits on both original and upsampled CBCT/CT images. First-order statistics and percentiles computed on the original CBCT/CT image improve the positive predictive value, but yields to a blocky effect in the segmentation result when not used in combination with first-order statistics and percentiles computed on the upsampled CBCT/CT image. We also checked (not reported here) the accuracy of the general-purpose segmentation tools on the upsampled CBCT images. Obtained results suggests that these general-purpose tools do not benefit from an upsampling of the CBCT image. On the contrary, worse results were obtained, with an average worsening on the dice scores of 80.1% and 62.3% for ITK-SNAP and GeoS, respectively. We refrained from further investigating the reasons as to why of this behavior due to the lack of implementation details of the tools.

The proposed SRC approach can also be used on pathological anatomies as it does not rely on shape priors. For instance, in case of bony dehiscence of the fallopian canal. In facial nerve dehiscence the nerve is uncovered in the middle ear cavity, leading to proximity of air voxels to the facial nerve. As the proposed approach uses a band surrounding the facial nerve, it already includes air voxels labeled as background to train the model. Therefore, it is expected that the proposed method can handle these cases. However, due to the absence of this type of cases in our database, we were not able to test this point in this study. In the context of the required accuracy for an effective and safe cochlear implantation planning of

at least 0.3*mm* [131], analysis of the RMSE error (suitable for this clinical scenario as large errors are to be penalized), the proposed SRC approach is the only one yielding RMSE errors with ranges not surpassing the required accuracy for the tested CBCT/CT cases (Table 5.1 & 5.2).

There are some limitations in this study. First, the approach relies on aligned pairs of CBCT/CT and micro-CT images, which are not readily available on all centers. A potential solution to this limitation, is the use of synthetically-generated images from a phantom of known geometry. Similarly, our short-term goal is to prepare a data descriptor in order to make the datasets in this study available for research purposes. Secondly, the learned mapping between clinical and high resolution imaging is specific for the corresponding imaging devices used to generate training data. However, as technical specifications of CBCT/CT imaging devices among different vendors do not differ substantially for facial nerve imaging of cochlear patients, we hypothesize that utilization of an existing super-resolution classification model to a different CBCT/CT vendor might require slight adaptations related to straightforward intensity normalization and histogram matching operations. In this direction, future work includes evaluation of the approach on a large dataset including images from different CBCT/CT devices in order to produce a more generally applicable algorithm. Future work also consider a larger dataset of cochlear image datasets including pathological cases to further validate the approach. The segmentation method was made specific to the task of super-resolution segmentation of the facial nerve. Our next step is to extend it to the segmentation of the chorda tympani by creating a dedicated model for it. In order to share the data with the scientific community and to foster future research in this and other related research lines, a data descriptor and open repository will be released.

Another limitation is the computational cost needed to extract features on the upsampled CBCT/CT image (hour range depending on the length of the facial nerve), which is expected to be improved through a pyramidal upsampling scheme, on which features are progressively extracted on each resolution level and concatenated, similar to the pyramid approach recently proposed in [165].

In this study we employed an ad-hoc regularization post-processing of the resulting segmentation based on morphological operations, aiming at removing isolated small regions and holes in the segmentation. Future work includes the use of a regularization component based on a conditional random field, similar to [103]. In practice, the postprocessing step had a larger impact on the Hausdorff distance metric, as single and isolated voxels outside of the facial nerve region would be used to compute it.

We anticipate that the proposed approach can be seamlessly applied as well to pediatric cases, because it does not rely on shape priors as it is the case of atlas-based methods.

Moreover, as demonstrated, this approach can be applied to other image modalities for super-resolution image segmentation, particularly for CT, which is an imaging modality often employed for bone imaging.

5.6 Conclusions

We have presented an automatic random forest based super-resolution classification (SRC) framework for facial nerve segmentation from CBCT/CT images, which refines a given initial segmentation of the facial nerve to sub-voxel accuracy. Preliminary results on seven 3D CBCT and eight 3D CT ex-vivo datasets suggests that the proposed method achieves accurate segmentations at sub-voxel accuracy.

Chapter 6

Conclusion and Outlook

6.1 Conclusion

In this thesis, an approach for motion detection and two approaches for accurate image segmentation of the facial nerve have been presented. For motion detection, a practical solution is not required external devices or access to projected data, enabling its use in the work flow. For facial nerve segmentation, two approaches are proposed. The first one aims to enhance the facial nerve image based on a supervised learning method, and then segment the facial nerve image with the surgical planning software, OtoPlan. The second approach aims to segment the facial nerve based on a supervised learning method at sub-voxel accuracy.

The following will summarize key assessments of the proposed method from a bird's eye view.

6.1.1 Motion Detection: a registration based approach

The motion detection approach, discussed in Chapter 3, is a practical solution to evaluate the amount of motion in a CBCT image. As a first proof-of-concept, we have tested it on the simulated motion using a scanning system, and a robot arm to produce known motion patterns. The proposed motion detection approach is based on the expected correlation between motion and image blurriness. In the experiments, we have investigated two different simulated motion modes, sudden motion and continuous motion. Moreover, we have explored three rod-phantom with different intensities which were attached to the rotating block. We found out that positioning of rods in a phantom for sudden motion, with respect to x-ray beams, is very important to obtain appropriate phantom images. Due to the "illusory" phantom pattern produced in sudden motion study, we primarily focused on a continuous motion study. Preliminary results show that the proposed method is able to detect subtle

sudden, as well as continuous motion patterns. Results also confirm the hypothesis relating motion and image blurriness, through the Hausdorff distance metric. To our knowledge, this is the first direct method for head motion detection in CBCT image and the first prototype is estimated on phantom CBCT data, and robot-controlled motion patterns.

6.1.2 Facial Nerve Image Enhancement: a machine learning based approach

The facial nerve image enhancement approach presented in Chapter 4, has been developed based on supervised regression using multiple output extremely randomized trees. This approach learns the mapping between the low resolution CBCT image and the high resolution micro-CT image through features extraction from the CBCT image. Compared to single output trees, we have observed that multiple output trees reduce the computational time and consider the correlation among voxels in each output image response. In addition, the size of image patches have been investigated for the effects of mapping in the training phase. We found that excessively large size of image patches produce smoother but inaccurate prediction results. On the contrary, the relatively small size of image patch cannot capture enough information. We also observed that $5 \times 5 \times 5$ is the optimal size of image patches for image features. The experimental results show that the proposed method can enhance image information, which can then also be used to obtain an improved segmentation of the facial nerve, applied in the cochlear surgical planning software, OtoPlan. Using OtoPlan, the facial nerve segmentation highly depends on the user interaction. Therefore, segmentation results suffer from the inter- and intra- observer variability and could not be standardized. However, due to a lack of sufficient robustness of the regression, we moved on to a classification strategy so to focus on directly delineating the facial nerve, which is described in Chapter 5.

6.1.3 Facial Nerve Segmentation: a Super-Resolution Classification (SRC) machine learning based approach

The facial nerve segmentation approach, presented in Chapter 5, has been developed based on supervised classification using extremely randomized trees. The band-based region of interest speeds up the mapping process. Our implementation has achieved subvoxel level. Compared to other segmentation tools, the proposed SRC approach yield higher accuracy and robustness. This general approach can be applied to other clinical cases where the super resolution is required for a safe preoperative planning. The following three points should be mentioned, which have significant effects on the final prediction results of the proposed

method. First, the facial nerve is a very tiny structure, therefore the ratio of the foreground (facial nerve) and the background samples is considerable small. In the experiments, we selected the ratio to be 1 : 1, in order to compensate the bias of the training data. Secondly, the intensity normalization is an important step. Because its absence might lead to poor model's performance. Thirdly, the proposed method works on aligned image pairs. The registration error influences the mapping process in the training phase. If the registration error is large, the resulting classification model will include noise and lose predictive power. Registration is time consuming. However, it is performed off-line only once. Once the images are aligned, we do not need registration any more. To our knowledge, this is the first attempt to perform super-resolution classification for facial nerve segmentation in CBCT/CT images by exploring imaging modalities featuring different resolution levels.

6.2 Limitations of the Work

6.2.1 Motion Detection

In this work, three rod-phantoms have been prepared for motion detection. One rod-phantom could not be calculated due to the position at the block phantom. Therefore, only two rod-phantoms were evaluated. An optimal positioning of the rods for a wearable phantom needs to be integrated in the future. With the best position of the wearable phantom for the patient, the acceptable motion can be easily estimated for the surgical planning. The second limitation lies in the simulated motion patterns. Although we have simulated two motion patterns, they are only approximate of the patient's motion patterns. Head movements of the patient might produce different motion patterns. This preliminary study did not consider motion patterns from a real patient or subject. Up to now, it only roughly estimates motion patterns and measures the degree of the motion.

6.2.2 Facial Nerve Image Enhancement

In this thesis, a large database is expected to contain different kinds of ear images, including facial nerve dehiscence and facial nerve disorder (see section 1.1.4). The various image information of different facial nerve cases, which will provide more diverse patterns in the training data. It results in highly accurate and more robust machine learning model. This work is based on a registration framework. It requires CBCT and micro-CT be well registered.

6.2.3 Facial Nerve Segmentation

In this work, in order to obtain a more accurate predictive model, large volume of training date is required to ensure a good performance. Hence, more image data are expected to be evaluated. Besides, one needs to analyze some pathological cases such as facial nerve dehiscence and facial nerve disorder (see section 1.1.4). The second limitation lies in the computational cost of extracting features from the upsampled CBCT/CT image. The consuming time depends on the length of the facial nerve. The fourth limitation is that the prediction results rely on the registration of the CBCT and micro-CT at the preprocessing step.

6.3 Outlook

This research has contributed to the fields of motion detection and a highly accurate facial nerve segmentation for surgical planning. The final goal of the surgical planning is to obtain an optimal drill trajectory for cochlear implant surgery without damaging the facial nerve. Such surgical planning will lead a robotic image guided system to carry out a minimally invasive implantation.

6.3.1 Motion Detection

For motion detection, future research will involve evaluating the relationship between the position and pattern with more test samples. The aim is to find the optimal posture of the patient's head for CBCT scanning. In order to find the acceptable degree of the image blurring and measure Fiducial Local Error (FLE), we need to measure a large number of simulated motion images with different degrees, especially with the degree less than $0.75 degree$ in our cases. For measuring the real motion patterns of patients, we will capture a head motion with digital camera, similar to [113] [112] [114] [115]. After that, the improved simulated motion will be tested based on the proposed motion detection strategy. The future work will also include testing of the proposed approach on ex-vivo or in-vivo CBCT datasets. Finally, a motion protocol needs to be created and validated. Such protocol could include a 3D setup to quantify motion.

6.3.2 Facial Nerve Image Enhancement

The idea of the image enhancement is to obtain higher quality of the ear image and then segment facial nerve on the enhanced image. The final goal is facial nerve segmentation. We

will segment facial nerve directly without an enhancement step. Therefore, our following research concentrates on facial nerve segmentation.

6.3.3 Facial Nerve Segmentation

For facial nerve segmentation, future work will consider a more robust and general approach, which will be tested on a large medical image dataset from different CBCT/CT vendors. The cochlear image data should contain pathological cases to further evaluate this approach. This segmentation method can be applied to delineation of chorda tympani. In order to remove the small isolated regions and holes in the segmentation, future work can employ a regularization component based on a conditional random field, similar to [103]. The proposed method in this thesis can also be applied to other nerves and in general to any microscopic structure of the human body (e.g. blood vessels), provided that pair of high- and low- image resolution dataset are available . In a next step, deep learning approaches can be tested and compared to the proposed Super-Resolution Classification (SRC) method. Convolutional Neural Networks (CNN), also named as Convolutional Networks, have been very popular in recent years for image classification [23] [55] [84] and other computer vision applications. The advantages of CNN [35] are as followed: the training phase works on powerful GPUs; the Rectified Linear Units(ReLU) produces much faster training results and still maintain the accuracy [105]. A CNN model could be employed to carry out facial nerve segmentation in the sub-voxel resolution as the SRC. To our knowledge, there is no neural network method applicable to facial nerve segmentation for cochlear implantation.

Chapter 7

Appendices

7.1 Pseudo-code of the Extra-tree algorithm [1]

Build an extra tree ensemble (S).
Input: a training set S.
Output: a tree ensemble $T = \{t_1, ..., t_M\}$;
for $i \leftarrow 1$ **to** M
do Generate a tree: $t_i =$ **Build an extra tree** (S);
return (T).

Build an extra tree (S).
Input: a training set S.
Output: a tree t.
if $|S| < n_{min}$, **or**
all candidate attributes are constant in S, **or**
the output variable is constant in S
then Return a leaf labeled by class frequencies (or average output, in regression) in S
else
1. Select randomly K attributes, $\{a_1, ..., a_k\}$, without replacement, among all candidate attributes.
2. Generate K splits $\{s_1, ..., s_k\}$, where $s_i =$ **Pick a random split** $(S, a_i), \forall i = 1, ..., K$;
3. Select a split s_* such that $Score(s_*, S) = max_{i=1,...,K} Score(s_i, S)$;
4. Split S into subsets S_l and S_r according to the test s_*;
5. Build $t_l =$ **Build an extra tree** (S_l) and $t_l =$ **Build an extra tree** (S_r) from these subsets;

[1] Modified from [46]

6. Create a node with the split s_*, attach t_l and t_r as left and right subtrees of this node and return the resulting tree t.

Pick a random split(S, a).
Input: a training set S and an attribute a.
Output: a split.
if the attribute a is numerical:
then {
Compute the maximal and minimal value of a in S, denoted respectively by a^S_{min} and a^S_{max};
Draw a cut-point a_c uniformly in $[a^S_{min}, a^S_{max}]$;
return the split $[a < a_c]$;

if the attribute a is categorical (denote by A its set of possible values):
then {
Compute A_s the subset of A of values of a that appear in S;
Randomly draw a proper non empty subset A_1 of A_S and a subset A_2 of A/A_S;
Return the split $[a \in A_1 \cup A_2]$.

7.2 Point Set to Point Set Registration Method

We provide a short C++ code for the point set to point set registration method, which brings the center lines of rod phantoms into alignment.

```
//registration
//------------------------------------------------------------
// Set up the Metric
//------------------------------------------------------------
 typedef itk::EuclideanDistancePointMetric<
                                  PointSetType,
                                  PointSetType>
                                                MetricType;
 typedef MetricType::TransformType              TransformBaseType;
 typedef TransformBaseType::ParametersType      ParametersType;
 typedef TransformBaseType::JacobianType        JacobianType;
 MetricType::Pointer  metric = MetricType::New();
//------------------------------------------------------------
// Set up a Transform
//------------------------------------------------------------
 typedef itk::Euler3DTransform< double >      TransformType;
 TransformType::Pointer transform = TransformType::New();
 // Optimizer Type
 typedef itk::LevenbergMarquardtOptimizer OptimizerType;
 OptimizerType::Pointer      optimizer     = OptimizerType::New();
 optimizer->SetUseCostFunctionGradient(false);
 // Registration Method
 typedef itk::PointSetToPointSetRegistrationMethod<
                                        PointSetType,
                                        PointSetType >
                                                RegistrationType;
 RegistrationType::Pointer   registration  = RegistrationType::New();
 // Scale the translation components of the Transform in the Optimizer
 OptimizerType::ScalesType scales(transform->GetNumberOfParameters());
 const double translationScale = 1000.0; // dynamic range of ↩
    translations
 const double rotationScale = 1.0;  // dynamic range of rotations
 scales[0] = 1.0 / rotationScale;
 scales[1] = 1.0 / rotationScale;
 scales[2] = 1.0 / rotationScale;
 scales[3] = 1.0 / translationScale;
 scales[4] = 1.0 / translationScale;
 scales[5] = 1.0 / translationScale;
```

```
38 unsigned long   numberOfIterations =  200;
39 double          gradientTolerance  =  1e-4;   // convergence criterion
40 double          valueTolerance     =  1e-4;   // convergence criterion
41 double          epsilonFunction    =  1e-5;   // convergence criterion
42 optimizer->SetScales( scales );
43 optimizer->SetNumberOfIterations( numberOfIterations );
44 optimizer->SetValueTolerance( valueTolerance );
45 optimizer->SetGradientTolerance( gradientTolerance );
46 optimizer->SetEpsilonFunction( epsilonFunction );
47 // Start from an Identity transform (in a normal case, the user
48 // can probably provide a better guess than the identity...
49 transform->SetIdentity();
50 registration->SetInitialTransformParameters(transform->GetParameters()↩
       );
51 //-----------------------------------------------------------
52 // Connect all the components required for Registration
53 //-----------------------------------------------------------
54 registration->SetMetric( metric );
55 registration->SetOptimizer( optimizer );
56 registration->SetTransform( transform );
57 registration->SetFixedPointSet( fixedPointSet );
58 registration->SetMovingPointSet( movingPointSet );
59 try
60   {
61   registration->Update();
62   }
63 catch( itk::ExceptionObject & e )
64   {
65   std::cout << e << std::endl;
66   return EXIT_FAILURE;
67   }
68 std::cout << "Solution = " << transform->GetParameters() << std::endl;
```

Listing 7.1: Point set to point set registration method

```
const     unsigned int    Dimension = 3;
typedef   unsigned char PixelType;
typedef itk::Image<PixelType, Dimension > FixedImageType;
typedef itk::Image<PixelType, Dimension > MovingImageType;
typedef itk::ImageFileReader<FixedImageType> resampleReaderType;
typedef itk::ResampleImageFilter< MovingImageType,
                                  FixedImageType > ResampleFilterType;

resampleReaderType::Pointer fixedReader=resampleReaderType::New();
fixedReader->SetFileName("D:\\Programming Task 2014\\↩
    PhantomRegistration 5_05_2014\\PhantomNoMotion11_gblur.nii");
fixedReader->Update();

resampleReaderType::Pointer movingImageReader =resampleReaderType::New↩
    ();
movingImageReader->SetFileName("D:\\Programming Task 2014\\↩
    PhantomRegistration 5_05_2014\\PhantomNoMotion_intensity.mhd");
movingImageReader->Update();

ResampleFilterType::Pointer resampler = ResampleFilterType::New();
resampler->SetInput( movingImageReader->GetOutput() );
resampler->SetTransform( transform->GetInverseTransform()  );
FixedImageType::Pointer fixedImage = fixedReader->GetOutput();
resampler->SetSize( fixedImage->GetLargestPossibleRegion().GetSize());
resampler->SetOutputOrigin( fixedImage->GetOrigin() );
resampler->SetOutputSpacing( fixedImage->GetSpacing() );
resampler->SetOutputDirection( fixedImage->GetDirection() );
resampler->SetDefaultPixelValue(0);
resampler->Update();
```

Listing 7.2: Interpolated moving image with the optimal transform matric

7.3 A Protocol Description of the Registration Pipeline

1. Registration with Amira:

   First, manual alignment of the CBCT/CT image and the corresponding micro-CT image based on manually placed landmarks (from 4 landmarks). Second, first rigid registration with normalized mutual information in Amira (version 5.4.4).

2. Registration with elastix:

   First, the rigidly transformed micro-CT image is defined as the moving image, and the CBCT/CT image is defined as the fixed image. In order to preserve the resolution of the micro-CT image and transform it to the CBCT/CT image space, the CBCT/CT image is resampled to micro-CT resolution. Default registration parameters taken from **http://elastix.bigr.nl/wiki/index.php/Default0**.

   The resulting non-rigid trasnform parameters is used to transform the ground-truth label image using nearest interpolation. The resulting trasnformed ground-truth image is then used during training of the SRC approach.

7.4 Employed parameters of the ExtraTreesClassifier

Parameters		
n_estimators	10	the number of tress
criterion	default='gini'	the Gini impurity
max_feature	default='auto'	sqrt(the number of features)
max_depth	default='None'	nodes are expanded until all leaves are pure or until all leaves contain less than min_samples_split samples.

Table 7.1: Employed parameters of the ExtraTreesClassifier in sklearn.

Bibliography

[1] X-ray Imaging and Computed Tomography. [online] https://www.imt.liu.se/edu/courses/TBMT02/ct/X-Ray.pdf. Accessed: 2017-1-27.

[2] Artifacts and partial-volume effects. [online] http://www.ctlab.geo.utexas.edu/about-ct/artifacts-and-partial-volume-effects/. Accessed: 2017-1-2.

[3] Daniel C Alexander, Darko Zikic, Jiaying Zhang, Hui Zhang, and Antonio Criminisi. Image Quality Transfer via Random Forest Regression : Applications in Diffusion MRI. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014*, 8675:225–232, 2014. URL http://link.springer.com/chapter/10.1007/978-3-319-10443-0_29.

[4] K Araki, K Maki, K Seki, K Sakamaki, Y Harata, R Sakaino, T Okano, and K Seo. Characteristics of a newly developed dentomaxillofacial X-ray cone beam CT scanner (CB MercuRay™): system configuration and physical properties. *Dentomaxillofacial Radiology*, 2014.

[5] Kirk Baker. Singular value decomposition tutorial. *The Ohio State University*, 24, 2005.

[6] Julia F Barrett and Nicholas Keat. Artifacts in CT: recognition and avoidance 1. *Radiographics*, 24(6):1679–1691, 2004.

[7] Don Beddard and William H Saunders. Congenital defects in the fallopian canal. *The Laryngoscope*, 72(1):112–115, 1962.

[8] Brett Bell, Christof Stieger, Nicolas Gerber, Andreas Arnold, Claude Nauer, Volkmar Hamacher, Martin Kompis, Lutz Nolte, Marco Caversaccio, and Stefan Weber. A self-developed and constructed robot for minimally invasive cochlear implantation. *Acta oto-laryngologica*, 132(4):355–360, 2012.

[9] Brett Bell, Nicolas Gerber, Tom Williamson, Kate Gavaghan, Wilhelm Wimmer, Marco Caversaccio, and Stefan Weber. In Vitro Accuracy Evaluation of Image-Guided Robot System for Direct Cochlear Access. *Otology Neurotology*, 2013.

[10] Brett Bell, Tom Williamson, Nicolas Gerber, Kate Gavaghan, Wilhelm Wimmer, Martin Kompis, Stefan Weber, and Marco Caversaccio. An image-guided robot system for direct cochlear access. *Cochlear implants international*, 2014.

[11] Kunwar Bhatia, Kevin P Gibbin, Thomas P Nikolopoulos, and Gerard M O'Donoghue. Surgical complications and their management in a series of 300 consecutive pediatric cochlear implantations. *Otology & neurotology*, 25(5):730–739, 2004.

[12] Ujjal Bhowmik, M Zafar Iqbal, and Reza Adhami. Mitigating motion artifacts in FDK based 3D Cone-beam Brain Imaging System using markers. *Open Engineering*, 2(3): 369–382, 2012.

[13] Christoph Bodensteiner, Cristina Darolti, H Schumacher, Lars Matthäus, and Achim Schweikard. Motion and positional error correction for cone beam 3D-reconstruction with mobile C-arms. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 177–185. Springer, 2007.

[14] Joel D Boerckel, Devon E Mason, Anna M McDermott, and Eben Alsberg. Microcomputed tomography: approaches and applications in bioengineering. *Stem cell research & therapy*, 5(6):144, 2014.

[15] V Bolón-Canedo, E Ataer-Cansizoglu, D Erdogmus, Jayashree Kalpathy-Cramer, O Fontenla-Romero, A Alonso-Betanzos, and MF Chiang. Dealing with inter-expert variability in retinopathy of prematurity: a machine learning approach. *Computer methods and programs in biomedicine*, 122(1):1–15, 2015.

[16] Verónica Bolón-Canedo, Beatriz Remeseiro, Amparo Alonso-Betanzos, and Aurélio Campilho. Machine learning for medical applications.

[17] Boundless. Facial (VII) Nerve. [online] https://www.boundless.com/physiology/textbooks/boundless-anatomy-and-physiology-textbook/peripheral-nervous-system-13/cranial-nerves-131/facial-vii-nerve-704-6544/. Accessed: 2017-2-07.

[18] Steven K Boyd. Micro-computed tomography. In *Advanced Imaging in Biology and Medicine*, pages 3–25. Springer, 2009.

[19] Katharina Braun, Frank Böhnke, and Thomas Stark. Three-dimensional representation of the human cochlea using micro-computed tomography data: presenting an anatomical model for further numerical calculations. *Acta oto-laryngologica*, 132(6): 603–613, 2012.

[20] David J. Brenner and Eric J. Hall. Computed Tomography — An Increasing Source of Radiation Exposure. *New England Journal of Medicine N Engl J Med*, 357(22): 2277–2284, 2007. doi: 10.1056/nejmra072149.

[21] Jason Brownlee. Supervised and Unsupervised Machine Learning Algorithms. [online] http://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/. Accessed:2017-2-06.

[22] M Akmal Butt and Petros Maragos. Optimum design of chamfer distance transforms. *IEEE Transactions on Image Processing*, 7(10):1477–1484, 1998.

[23] Jinzheng Cai, Le Lu, Zizhao Zhang, Fuyong Xing, Lin Yang, and Qian Yin. Pancreas segmentation in mri using graph-based decision fusion on convolutional neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 442–450. Springer, 2016.

[24] The center for hearing and balance disorders. Surgical management of hearing loss. [online] http://www.stlouisear.com/hearing-loss.html. Accessed: 2017-3-13.

[25] Juan Cerrolaza, Sergio Vera, Alexis Bagué, Mario Ceresa, Pablo Migliorelli, Marius George Linguraru, and Miguel Ángel González Ballester. Hierarchical shape modeling of the cochlea and surrounding risk structures for minimally invasive cochlear implant surgery. In *Workshop on Clinical Image-Based Procedures*, pages 59–67. Springer, 2014.

[26] Kai-Chieh Chan, Pa-Chun Wang, Yen-An Chen, and Che-Ming Wu. Facial Nerve Dehiscence at Mastoidectomy for Cholesteatoma. *Int Adv Otol*, 7:311–316, 2011.

[27] Vimal Chandran, Philippe Zysset, and Mauricio Reyes. Prediction of Trabecular Bone Anisotropy from Quantitative Computed Tomography Using Supervised Learning and a Novel Morphometric Feature Descriptor. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015*, pages 621–628. Springer, 2015.

[28] Taylor P Cotton, Todd M Geisler, David T Holden, Scott A Schwartz, and William G Schindler. Endodontic applications of cone-beam volumetric tomography. *Journal of endodontics*, 33(9):1121–1132, 2007.

[29] Antonio Criminisi and Jamie Shotton. *Decision forests for computer vision and medical image analysis*. Springer Science & Business Media, 2013.

[30] Antonio Criminisi, Toby Sharp, and Andrew Blake. Geos: Geodesic image segmentation. In *Computer Vision–ECCV 2008*, pages 99–112. Springer, 2008.

[31] Rueckert Daniel. Machine learning meets medical imaging: From signals to clinically useful information. [online] https://www.youtube.com/watch?v=7vtpWbrVdDY. Accessed: 2016-12-18.

[32] Jens De Cock, Koen Mermuys, Jean Goubau, Simon Van Petegem, Brecht Houthoofd, and Jan W Casselman. Cone-beam computed tomography: a new low dose, high resolution imaging technique of the wrist, presentation of three cases with technique. *Skeletal radiology*, 41(1):93–96, 2012.

[33] Lee R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.

[34] I Diogo, U Walliczeck, J Taube, N Franke, A Teymoortash, J Werner, and C Güldner. Possibility of differentiation of cochlear electrodes in radiological measurements of the intracochlear and chorda-facial angle position. *Acta Otorhinolaryngologica Italica*, 36(4):310, 2016.

[35] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2016.

[36] Marie Dumont, Raphaël Marée, Louis Wehenkel, and Pierre Geurts. Fast multi-class image annotation with random subwindows and multiple output randomized trees. In *Proc. International Conference on Computer Vision Theory and Applications (VISAPP)*, volume 2, pages 196–203, 2009.

[37] Georg Eggers, Hitomi Senoo, Gavin Kane, and Joachim Mühling. The accuracy of image guided surgery based on cone beam computer tomography image data. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology*, 107(3): e41–e48, 2009.

[38] Pierre Elbischger, Stig Geerts, Kathrin Sander, Gerda Ziervogel-Lukas, and P Sinah. Algorithmic framework for HEp-2 fluorescence pattern classification to aid auto-immune diseases diagnosis. In *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 562–565. IEEE, 2009.

[39] Halit Eren and John G Webster. *Telehealth and Mobile Health*. CRC Press, 2015.

[40] Paolo Favaro. Advanced Topics in Machine Learning. University of Bern, 2017.

[41] J Michael Fitzpatrick. Fiducial registration error and target registration error are uncorrelated. In *SPIE Medical Imaging*, pages 726102–726102. International Society for Optics and Photonics, 2009.

[42] Alexander Graham Bell Association for the Deaf and Hard of Hearing. How Hearing Works. [online] https://www.agbell.org/Document.aspx?id=138. Accessed: 2017-1-23.

[43] Robert L Galloway Jr. The process and development of image-guided procedures. *Annual Review of Biomedical Engineering*, 3(1):83–108, 2001.

[44] Nicolas Gerber. *Computer Assisted Planning and Image Guided Surgery for Hearing Aid Implantation*. PhD thesis, University of Bern, 2013.

[45] Nicolas Gerber, Brett Bell, Kate Gavaghan, Christian Weisstanner, Marco Caversaccio, and Stefan Weber. Surgical planning tool for robotically assisted hearing aid implantation. *International Journal of Computer Assisted Radiology and Surgery*, 9: 11–20, 2014. ISSN 18616410. doi: 10.1007/s11548-013-0908-5.

[46] Pierre Geurts, Damien Ernst, and Louis Wehenkel. Extremely randomized trees. *Machine learning*, 63(1):3–42, 2006.

[47] Subarna Ghosh and Vipin Chaudhary. Feature analysis for automatic classification of HEp-2 florescence patterns: Computer-aided diagnosis of auto-immune diseases. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 174–177. IEEE, 2012.

[48] Ben Glocker, Olivier Pauly, Ender Konukoglu, and Antonio Criminisi. Joint classification-regression forests for spatially structured multi-object segmentation. In *European Conference on Computer Vision*, pages 870–881. Springer, 2012.

[49] Lee W. Goldman. Principles of CT: radiation dose and image quality. *Journal of Nuclear Medicine Technology*, 35:213–225, 2007.

[50] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[51] Maria Eugenia Guerrero, Reinhilde Jacobs, Miet Loubele, Filip Schutyser, Paul Suetens, and Daniel van Steenberghe. State-of-the-art on cone beam CT imaging for preoperative planning of implant placement. *Clinical oral investigations*, 10(1):1–7, 2006.

[52] Sachin Gupta, Francine Mends, Mari Hagiwara, Girish Fatterpekar, and Pamela C Roehm. Imaging the facial nerve: a contemporary review. *Radiology research and practice*, 2013, 2013.

[53] Robert M Haralick and K Shanmugam. Textural Features for Image Classification. *IEEE TSMC-3*, 6:610–621, 1973. doi: 10.1109/TSMC.1973.4309314.

[54] Robert M Haralick, Karthikeyan Shanmugam, and Its' Hak Dinstein. Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, (6): 610–621, 1973.

[55] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. In *European Conference on Computer Vision*, pages 346–361. Springer, 2014.

[56] healthline. Facial nerve, . [online] http://www.healthline.com/human-body-maps/facial-nerve. Accessed: 2017-2-07.

[57] healthline. Semicircular canals, . [online] http://www.healthline.com/human-body-maps/semicircular-canals. Accessed: 2017-1-23.

[58] hear-it.org. The ear - a magnificent organ, . [online] http://www.hear-it.org/The-ear---a-magnificent-organ. Accessed: 2017-1-23.

[59] hear-it.org. The inner ear, . [online] http://www.hear-it.org/The-inner-ear-1. Accessed: 2017-1-23.

[60] hear-it.org. The outer ear, . [online] http://www.hear-it.org/The-outer-ear. Accessed: 2017-1-23.

[61] hear-it.org. The middle ear, . [online] http://www.hear-it.org/The-middle-ear-1. Accessed: 2017-1-23.

[62] hearnet. About Hearing Loss. [online] http://www.hearnet.com/at_risk/risk_aboutloss.shtml. Accessed: 2017-1-23.

[63] Tin Kam Ho. Random decision forests. In *Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on*, volume 1, pages 278–282. IEEE, 1995.

[64] Christof Holberg, Stefanie Steinhäuser, Phillip Geis, and Ingrid Rudzki-Janson. Cone-beam computed tomography in orthodontics: benefits and limitations. *Journal of Orofacial Orthopedics/Fortschritte der Kieferorthopädie*, 66(6):434–444, 2005.

[65] University Hospital Southampton NHS Foundation Trust. Facial Nerve Disorders. [online] http://www.uhs.nhs.uk/OurServices/Brainspineandneuromuscular/TheWessexFacialNerveCentre/Typesoffacialnervedisorders.aspx. Accessed: 2016-9-5.

[66] Walter Huda, Kristin A. Lieberman, Jack Chang, and Marsha L. Roskopf. Patient size and x-ray technique factors in head computed tomography examinations. II. image quality. *Med. Phys. Medical Physics*, 31(3):595, 2004. doi: 10.1118/1.1646233.

[67] Daniel P Huttenlocher, Gregory Klanderman, William J Rucklidge, et al. Comparing images using the Hausdorff distance. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 15(9):850–863, 1993.

[68] Toronto Hear Clinic Inc. Combined electric and acoustic stimulation. [online] http://www.hearlifeclinic.com/ca/show/index/id/86/title/EAS-HEARING-IMPLANTS. Accessed: 2017-3-15.

[69] Business Insider UK. Restore Hearing. [online] http://www.businessinsider.com/brain-implants-will-give-us-superpowers-2014-4?IR=T. Accessed: 2017-1-24.

[70] Lorenz Jäger, Harald Bonell, Martin Liebl, Sudesh Srivastav, Viktor Arbusow, Martin Hempel, and Maximilian Reiser. CT of the Normal Temporal Bone: Comparison of Multi–and Single–Detector Row CT 1. *Radiology*, 235(1):133–141, 2005.

[71] Robert J. Witte John I. Lane. *The Temporal Bone*. Springer-Verlag Berlin Heidelberg, 2010.

[72] Hans J Johnson, Matt McCormick, and Luis Ibanez. The itk software guide third edition updated for itk version 4.5. 2013.

[73] Hans J. Johnson, Matthew M. McCormick, and Luis Ibanez. *The ITK Software Guide Book 1: Introduction and Development Guidelines - Volume 1*. Kitware, Inc., USA, 2015. ISBN 1930934270, 9781930934276.

[74] Micheline Kamber, Jiawei Han, and Jian Pei. *Data mining: Concepts and techniques*. Elsevier, 2012.

[75] A Khursheed, M C Hillier, P C Shrimpton, and B F Wall. Influence of patient age on normalized effective doses calculated for CT examinations. *The British Journal of Radiology BJR*, 75(898):819–830, 2002. doi: 10.1259/bjr.75.898.750819.

[76] JH Kim, Johan Nuyts, A Kyme, Z Kuncic, and R Fulton. A rigid motion correction method for helical computed tomography (CT). *Physics in medicine and biology*, 60 (5):2047, 2015.

[77] JH Kim, Tao Sun, AR Alcheikh, Z Kuncic, Johan Nuyts, and R Fulton. Correction for human head motion in helical x-ray CT. *Physics in medicine and biology*, 61(4):1416, 2016.

[78] Namkeun Kim, Yongjin Yoon, Charles Steele, and Sunil Puria. Cochlear anatomy using micro computed tomography (μCT) imaging. In *Biomedical Optics (BiOS) 2008*, pages 68421A–68421A. International Society for Optics and Photonics, 2008.

[79] Stefan Klein, Marius Staring, Keelin Murphy, Max A Viergever, and Josien PW Pluim. Elastix: a toolbox for intensity-based medical image registration. *Medical Imaging, IEEE Transactions on*, 29(1):196–205, 2010.

[80] Darius Kohan and Daniel Jethanamest. Image-guided surgical navigation in otology. *The Laryngoscope*, 122(10):2291–2299, 2012.

[81] Igor Kononenko. Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in medicine*, 23(1):89–109, 2001.

[82] Stilianos E Kountakis. *Encyclopedia of otolaryngology, head and neck surgery*. Springer, 2013.

[83] Louis B Kratchman, Grégoire S Blachon, Thomas J Withrow, Ramya Balachandran, Robert F Labadie, and Robert J Webster. Design of a bone-attached parallel robot for percutaneous cochlear implantation. *IEEE Transactions on Biomedical Engineering*, 58(10):2904–2910, 2011.

[84] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[85] Liang Li, Zhiqiang Chen, Xin Jin, Hengyong Yu, and Ge Wang. Experimental measurement of human head motion for high-resolution computed tomography system design. *Optical Engineering*, 49(6):063201–063201, 2010.

[86] Chi Liu, Adam M Alessio, and Paul E Kinahan. Respiratory motion correction for quantitative PET/CT using all detected events with internal—external motion correlation. *Medical physics*, 38(5):2715–2723, 2011.

[87] William E Lorensen and Harvey E Cline. Marching cubes: A high resolution 3D surface construction algorithm. In *ACM siggraph computer graphics*, volume 21, pages 163–169. ACM, 1987.

[88] Yifei Lou, Tianye Niu, Xun Jia, Patricio a. Vela, Lei Zhu, and Allen R. Tannenbaum. Joint CT/CBCT deformable registration and CBCT enhancement for cancer radiotherapy. *Medical Image Analysis*, 17(3):387–400, 2013. ISSN 13618415. doi: 10.1016/j.media.2013.01.005. URL http://dx.doi.org/10.1016/j.media.2013.01.005.

[89] Ping Lu. Rotation invariant Registration of 2D aerial images using local phase correlation, 2013.

[90] George D Magoulas and Andriana Prentza. Machine learning in medical applications. In *Machine Learning and its applications*, pages 300–307. Springer, 2001.

[91] Omid Majdani, Thomas S Rau, Stephan Baron, Hubertus Eilers, Claas Baier, Bodo Heimann, Tobias Ortmaier, Sönke Bartling, Thomas Lenarz, and Martin Leinung. A robot-guided minimally invasive approach for cochlear implant surgery: preliminary results of a temporal bone study. *International journal of computer assisted radiology and surgery*, 4(5):475–486, 2009.

[92] T E Marchant, C J Moore, C G Rowbottom, R I MacKay, and P C Williams. Shading correction algorithm for improvement of cone-beam CT images in radiotherapy. *Physics in medicine and biology*, 53(2008):5719–5733, 2008. ISSN 0031-9155. doi: 10.1088/0031-9155/53/20/010.

[93] Raphaël Marée, Louis Wehenkel, and Pierre Geurts. Extremely randomized trees and random subwindows for image classification, annotation, and retrieval. In *Decision Forests for Computer Vision and Medical Image Analysis*, pages 125–141. Springer, 2013.

[94] MED-EL. Cochlear implant system, . [online] http://www.medel.com/us/image-gallery-usa/. Accessed: 2017-3-15.

[95] MED-EL. What is a degree of hearing loss?, . [online] http://www.medel.com/blog/degree-of-hearing-loss/. Accessed: 2017-1-23.

[96] MED-EL. How hearing works, . [online] http://www.medel.com/int/how-hearing-works/. Accessed: 2017-3-13.

[97] MED-EL. Anatomy of the Ear, . [online] http://www.medel.com/anatomy-of-the-ear/. Accessed: 2017-1-23.

[98] MED-EL. How Hearing Works, . [online] http://www.medel.com/how-hearing-works/. Accessed: 2017-1-23.

[99] MED-EL. How The Ear Works, . [online] http://www.medel.com/blog/how-the-ear-works/. Accessed: 2017-1-25.

[100] MED-EL. What is Sound?, . [online] http://www.medel.com/blog/what-is-sound/. Accessed: 2017-1-23.

[101] MED-EL. The benefits of complete cochlear coverage (ccc), . [online] http://www.medel.com/technology-complete-cochlear-coverage/. Accessed: 2017-3-13.

[102] MedicineNet. Facial Nerve Problems and Bell's Palsy (Bell Palsy). [online] http://www.medicinenet.com/facial_nerve_problems/article.htm. Accessed: 2017-2-07.

[103] Raphael Meier, Venetia Karamitsou, Simon Habegger, Roland Wiest, and Mauricio Reyes. Parameter learning for crf-based tissue segmentation of brain tumors. In *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 156–167. Springer, 2015.

[104] P Mozzo, C Procacci, A Tacconi, P Tinazzi Martini, and IA Bergamo Andreis. A new volumetric CT machine for dental imaging based on the cone-beam technique: preliminary results. *European radiology*, 8(9):1558–1564, 1998.

[105] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814, 2010.

[106] Ichiro Nishimura. Three-dimensional bone-implant integration profiling using micro-computed tomography. 2005.

[107] Tianye Niu and Lei Zhu. Overview of x-ray scatter in cone-beam computed tomography and its correction methods. *Current Medical Imaging Reviews*, 6(2):82–89, 2010.

[108] Jack H Noble, Frank M Warren, Robert F Labadie, and Benoit M Dawant. Automatic segmentation of the facial nerve and chorda tympani in CT images using spatially dependent feature values. *Medical physics*, 35(12):5375–5384, 2008.

[109] Jack H Noble, Benoit M Dawant, Frank M Warren, and Robert F Labadie. Automatic identification and 3-D rendering of temporal bone anatomy. *Otology & neurotology: official publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*, 30(4):436, 2009.

[110] Jack H Noble, Omid Majdani, Robert F Labadie, Benoit Dawant, and J Michael Fitzpatrick. Automatic determination of optimal linear drilling trajectories for cochlear access accounting for drill-positioning error. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 6(3):281–290, 2010.

[111] Ozan Oktay, Wenjia Bai, Matthew Lee, Ricardo Guerrero, Konstantinos Kamnitsas, Jose Caballero, Antonio de Marvao, Stuart Cook, Declan O'Regan, and Daniel Rueckert. Multi-input Cardiac Image Super-Resolution Using Convolutional Neural Networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 246–254. Springer, 2016.

[112] Oline V Olesen, MR Jorgensen, and Rasmus R Paulsen. Structured light 3d tracking system for measuring motions in pet brain imaging. In *Proc. SPIE*, volume 7625, page 76250X, 2010.

[113] Oline Vinter Olesen, Rasmus R Paulsen, Liselotte Højgaard, Bjarne Roed, and Rasmus Larsen. Motion tracking in narrow spaces: a structured light approach. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 253–260. Springer, 2010.

[114] Oline Vinter Olesen, Rasmus R Paulsen, Liselotte Hojgaard, Bjarne Roed, and Rasmus Larsen. Motion tracking for medical imaging: a nonvisible structured light tracking approach. *IEEE transactions on medical imaging*, 31(1):79–87, 2012.

[115] Oline Vinter Olesen, Jenna M Sullivan, Tim Mulnix, Rasmus R Paulsen, Liselotte Hojgaard, Bjarne Roed, Richard E Carson, Evan D Morris, and Rasmus Larsen. List-mode pet motion correction using markerless head tracking: Proof-of-concept with scans of human subject. *IEEE transactions on medical imaging*, 32(2):200–209, 2013.

[116] Andrés Ortiz, Antonio A Palacio, Juan M Górriz, Javier Ramírez, and Diego Salas-González. Segmentation of brain MRI using SOM-FCM-based method and 3D statistical descriptors. *Computational and mathematical methods in medicine*, page 638563, 2013. ISSN 1748-6718. doi: 10.1155/2013/638563. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3666364&tool=pmcentrez&rendertype=abstract.

[117] Andrés Ortiz, Antonio A Palacio, Juan M Górriz, Javier Ramírez, and Diego Salas-González. Segmentation of brain MRI using SOM-FCM-based method and 3D statistical descriptors. *Computational and mathematical methods in medicine*, 2013, 2013.

[118] Olivier Pauly. *Random Forests For Medical Application*. PhD thesis, Technischen Universität München, 2012.

[119] Ruben Pauwels, Jilke Beinsberger, Bruno Collaert, Chrysoula Theodorakou, Jessica Rogers, Anne Walker, Lesley Cockmartin, Hilde Bosmans, Reinhilde Jacobs, Ria Bogaerts, et al. Effective dose range for dental cone beam computed tomography scanners. *European journal of radiology*, 81(2):267–271, 2012.

[120] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(Oct):2825–2830, 2011.

[121] Terry M Peters. Image-guidance for surgical procedures. *Physics in medicine and biology*, 51(14):R505, 2006.

[122] pinterest. Facial nerve. [online] https://www.pinterest.com/explore/facial-nerve/. Accessed: 2017-3-15.

[123] Milan Profant, Zuzana Kabátová, and Lukáš Varga. Otosclerosis and Cochlear Implantation. In *Surgery of Stapes Fixations*, pages 105–112. Springer, 2016.

[124] quizlet. Scalp and face. [online] https://quizlet.com/96888444/scalp-and-face-flash-cards/. Accessed: 2017-3-15.

[125] Mark A Rafferty, Jeffrey H Siewerdsen, Yvonne Chan, Michael J Daly, Douglas J Moseley, David A Jaffray, and Jonathan C Irish. Intraoperative cone-beam CT for guidance of temporal bone surgery. *Otolaryngology—Head and Neck Surgery*, 134(5): 801–808, 2006.

[126] Fitsum A Reda, Jack H Noble, Alejandro Rivas, Theodore R McRackan, Robert F Labadie, and Benoit M Dawant. Automatic segmentation of the facial nerve and chorda tympani in pediatric CT scans. *Medical physics*, 38(10):5590–5600, 2011.

[127] Joana Ruivo, Koen Mermuys, Klaus Bacher, Rudolf Kuhweide, Erwin Offeciers, and Jan W Casselman. Cone beam computed tomography, a low-dose imaging technique in the postoperative assessment of cochlear implantation. *Otology & neurotology*, 30 (3):299–303, 2009.

[128] John C Russ. *The image processing handbook*. CRC press, 2016.

[129] William C Scarfe and Allan G Farman. What is cone-beam CT and how does it work? *Dental Clinics of North America*, 52(4):707–730, 2008.

[130] Rob Schapire. Theoretical Machine Learning. [online] http://www.cs.princeton.edu/courses/archive/spr08/cos511/scribe_notes/0204.pdf. Accessed: 2016-12-18.

[131] Jörg Schipper, Antje Aschendorff, Iakovos Arapakis, Thomas Klenzner, Christian Barna Teszler, Gerd Jürgen Ridder, and Roland Laszig. Navigation as a quality management tool in cochlear implant surgery. *The Journal of Laryngology & Otology*, 118(10):764–770, 2004.

[132] Samuel Schulter, Christian Leistner, and Horst Bischof. Fast and accurate image upscaling with super-resolution forests. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3791–3799, 2015.

[133] ScienceNews. Artificial hearing has come a long way since 1960s. [online] https://www.sciencenews.org/article/artificial-hearing-has-come-long-way-1960s. Accessed: 2017-1-24.

[134] scikit learn. Decision Trees, . [online] http://scikit-learn.org/stable/modules/tree.html#tree-multioutput. Accessed: 2017-2-23.

[135] scikit learn. Ensemble methods, . [online] http://scikit-learn.org/stable/modules/ensemble.html#forest. Accessed: 2017-1-10.

[136] seelio. Shape memory polymer cochlear implant insertion device. [online] https://seelio.com/w/2hui/shape-memory-polymer-cochlear-implant-insertion-device?student=zlsiegel. Accessed: 2017-3-15.

[137] Denis P Shamonin, Esther E Bron, Boudewijn PF Lelieveldt, Marion Smits, Stefan Klein, and Marius Staring. Fast parallel image registration on CPU and GPU for diagnostic classification of alzheimer's disease. *Frontiers in Neuroinformatics, 7, 2014*, 2014.

[138] Jonathon Shlens. A tutorial on principal component analysis. *arXiv preprint arXiv:1404.1100*, 2014.

[139] SIEMENS. Degrees of Hearing Loss. [online] https://usa.bestsoundtechnology.com/hearing-loss-and-tinnitus/understanding-hearing-loss/degrees-of-hearing-loss/. Accessed: 2017-1-23.

[140] Milan Sonka, Vaclav Hlavac, and Roger Boyle. *Image processing, analysis, and machine vision*. Cengage Learning, 2014.

[141] GN Srinivasan and G Shobha. Statistical texture analysis. In *Proceedings of world academy of science, engineering and technology*, volume 36, pages 1264–1269, 2008.

[142] Detlev Stalling, Malte Westerhoff, and Hans-Christian Hege. Amira: a highly interactive system for visual data analysis, 2005.

[143] Colin Studholme, Derek LG Hill, and David J Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern recognition*, 32(1):71–86, 1999.

[144] P Sukovic. Cone beam computed tomography in craniofacial imaging. *Orthodontics & craniofacial research*, 6(s1):31–36, 2003.

[145] Kenji Suzuki. Pixel-based machine learning in medical imaging. *Journal of Biomedical Imaging*, 2012:1, 2012.

[146] Elham Taghizadeh and Reyes Mauricio. Medical Image Analysis. University of Bern, 2014.

[147] Mahdieh Taleb Mehr. Usefulness of dental cone beam computed tomography (CBCT) for detetion of the anatomical landmarks of the external, middle and inner ear. 2013.

[148] Klaus D. Toennies. *Guide to Medical Image Analysis: Methods and Algorithms*. Springer, 2012.

[149] Klaus D Toennies. *Guide to medical image analysis: methods and algorithms*. Springer Science & Business Media, 2012.

[150] Kostas Tsiklakis, Catherine Donta, Sophia Gavala, Kety Karayianni, Vasiliki Kamenopoulou, and Costas J Hourdakis. Dose reduction in maxillofacial imaging using low dose Cone Beam CT. *European journal of radiology*, 56(3):413–417, 2005.

[151] Eduard HJ Voormolen, Marijn van Stralen, Peter A Woerdeman, Josien PW Pluim, Herke Jan Noordmans, Max A Viergever, Luca Regli, and Jan Willem Berkelbach van der Sprenkel. Determination of a facial nerve safety zone for navigated temporal bone surgery. *Operative Neurosurgery*, 70:ons50–ons60, 2012.

[152] Guotai Wang, Maria A Zuluaga, Rosalind Pratt, Michael Aertsen, Tom Doel, Maria Klusmann, Anna L David, Jan Deprest, Tom Vercauteren, and Sébastien Ourselin. Dynamically Balanced Online Random Forests for Interactive Scribble-Based Segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 352–360. Springer, 2016.

[153] Frank M Warren, Ramya Balachandran, J Michael Fitzpatrick, and Robert F Labadie. Percutaneous cochlear access using bone-mounted, customized drill guides: demonstration of concept in vitro. *Otology & Neurotology*, 28(3):325–329, 2007.

[154] Stefan Weber, Brett Belle, Peter Brett, Xinli Du, Marco Caversaccio, David Proops, Chris Coulson, and Andrew Reid. Minimally invasive, robot assisted cochlear implantation. In *Proceedings of the 3rd joint workshop on new technologies for computer/robot assisted surgery (CRAS)*, pages 134–137, 2013.

[155] Louis Wehenkel. On uncertainty measures used for decision tree induction. In *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, page 6, 1996.

[156] Miles N Wernick, Yongyi Yang, Jovan G Brankov, Grigori Yourganov, and Stephen C Strother. Machine learning in medical imaging. *IEEE signal processing magazine*, 27 (4):25–38, 2010.

[157] Wikipedia. Cochlea, . [online] https://en.wikipedia.org/wiki/Cochlea. Accessed: 2017-1-23.

[158] Wikipedia. Ear, . [online] https://en.wikipedia.org/wiki/Ear. Accessed: 2017-1-23.

[159] Wikipedia. Facial nerve, . [online] https://en.wikipedia.org/wiki/Facial_nerve. Accessed: 2017-2-07.

[160] Tom Williamson. *Integrated Sensing Control For Robotic Microsurgery on the Lateral Skull Base*. PhD thesis, University of Bern, 2015.

[161] Wilhelm Wimmer, Brett Bell, Markus E Huth, Christian Weisstanner, Nicolas Gerber, Martin Kompis, Stefan Weber, and Marco Caversaccio. Cone beam and microcomputed tomography validation of manual array insertion for minimally invasive cochlear implantation. *Audiology and Neurotology*, 19(1):22–30, 2013.

[162] Yong Yang, Eduard Schreibmann, Tianfang Li, Chuang Wang, and Lei Xing. Evaluation of on-board kV cone beam CT (CBCT)-based dose calculation. *Physics in medicine and biology*, 52(3):685, 2007.

[163] Ziv Yaniv and Kevin Cleary. Image-guided procedures: A review. *Computer Aided Interventions and Medical Robotics*, 3:1–63, 2006.

[164] Paul A. Yushkevich, Joseph Piven, Heather Cody Hazlett, Rachel Gimpel Smith, Sean Ho, James C. Gee, and Guido Gerig. User-guided 3D Active Contour Segmentation of Atomical Structures: Significantly Improved Efficiency and Reliability. *Neuroimage*, 31(3):1116–1128, 2006.

[165] Qiang Zhang, Abhir Bhalerao, Edward Dickenson, and Charles Hutchinson. Active Appearance Pyramids for Object Parametrisation and Fitting. *Medical Image Analysis*, 2016.

[166] Qinghui Zhang, Yu-Chi Hu, Fenghong Liu, Karyn Goodman, Kenneth E Rosenzweig, and Gig S Mageras. Correction of motion artifacts in cone-beam CT using a patient-specific respiratory motion model. *Medical physics*, 37(6):2901–2909, 2010.

[167] Barbara Zitova and Jan Flusser. Image registration methods: a survey. *Image and vision computing*, 21(11):977–1000, 2003.

[168] Jing Zou, Markus Hannula, Kalle Lehto, Hao Feng, Jaakko Lähelmä, Antti S Aula, Jari Hyttinen, and Ilmari Pyykkö. X-ray microtomographic confirmation of the reliability of CBCT in identifying the scalar location of cochlear implant electrode after round window insertion. *Hearing research*, 326:59–65, 2015.

Declaration of Originality

Last name, first name: Lu Ping
Matriculation number: 12-137-733

I hereby declare that this thesis represents my original work and that I have used no other sources except as noted by citations.
All data, tables, figures and text citations which have been repreduced from any other source, including the internet, have been explicitly acknowledged as such.
I am aware that in case of non-compliance, the Senate is entitled to withdraw the doctorate degree awarded to me on the basis of the present thesis, in accordance with the "Statut der Universität Bern (Universitätsstatut; UniSt)", Art. 69, of 7 June 2011.

Bern, 30 May 2017

Ping Lu