# AUTOMATIC IDENTIFICATION OF THE STRUCTURES OF THE EAR AND A NEW APPROACH FOR TUBULAR STRUCTURE MODELING AND SEGMENTATION

By

Jack H. Noble

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Electrical Engineering

May, 2011

Nashville, Tennessee

Approved by:

Professor Benoit M. Dawant

Professor J. Michael Fitzpatrick

Professor Bennett A. Landman

Professor Don M. Wilkes

Professor Robert F. Labadie

## ACKNOWLEDGEMENTS

I would like to thank Dr. Benoit Dawant for accepting me as a research assistant during my undergraduate and graduate careers at Vanderbilt. His invaluable guidance, dedication, and patience have profoundly influenced my education and research path. He never hesitated to help at every opportunity and always made sure I set high goals.

I would also like to thank Dr. Mike Fitzpatrick for his guidance early in my graduate career. His dedication to academic rigor has been nothing short of inspirational.

I would like to specially thank Dr. Robert Labadie for welcoming me into the CAOS lab and for always making the time to provide medical expertise. His innovations in the field of ear surgery led to most of the work in this dissertation and have provided a practical outlet for the work herein.

I would also like to thank Drs. Landman and Wilkes for their committee duties and their ideas that have contributed to this work, and I would also like to thank all of the people with whom I have had the pleasure of working with at Vanderbilt University.

Validation of the methods presented in this work would not have been possible without the assistance of many medical specialists who provided their time and expert anatomical knowledge, and each of them deserves acknowledgement.

I would also like to thank my family for a lifetime of their support. Finally, this work would not have been possible without the financial support from the NIH pre-doctoral training grant F31DC009791.

## TABLE OF CONTENTS

**Page**

## LIST OF FIGURES

**Page**

## LIST OF TABLES

**Page**

# CHAPTER I

# INTRODUCTION

This dissertation presents studies on, and from, the development of algorithms for the automatic segmentation of the structures of the ear. Many otological procedures would benefit from a system that automatically identifies anatomical structures of the ear in CT. Conventional (registration-based) segmentation techniques are suitable for identifying ear structures that have high intensity contrast in CT or those for which a high degree of accuracy is not necessary. For some ear structures, conventional techniques are inadequate, and other segmentation methods must be used or developed.

In this dissertation, approaches that permit the automatic identification of ear structures will be presented. This will include the ossicles (OS), external auditory canal (EAC), cochlea (CO), scala tympani (ST), scala vestibuli (SV), facial nerve (FN), chorda tympani (CH), semicircular canals (SC), and the carotid artery (CA). Each of these sensitive structures could be within millimeters of the surgical approach for various types of procedures.

Out of the above mentioned structures, six distinct tubular shaped structures are represented, each of which has unique properties that make detection difficult. This has led to the development of an algorithm for the segmentation of general tubular structures, which is also presented in this dissertation. This algorithm is validated on these ear structures as well as other tubular structures outside of the ear.

The results of all methods presented in this work are analyzed, and quantitatively compared to expert drawn or expert edited segmentations, which are treated as the gold standard.

# CHAPTER II

# BACKGROUND

## Introduction

In this chapter, the clinical impact of structure localization for otological procedures is introduced. Following this, background information is presented for image segmentation and specifically tubular structure segmentation. This is followed by introduction of work performed to identify the facial nerve and chorda, two tubular structures, and how that work led to the development of a general tubular structure localization algorithm.

## Ear Anatomy

Otologic surgeries, i.e., surgeries of the ear, require high precision and expert knowledge of complex anatomical relationships. Computer assistance systems are becoming common in many surgical fields, but are used only sparingly in the field of otology. A system for identification of a patient's ear anatomy would have several benefits. The anatomy could be reconstructed in a 3D rendering environment so that spatial relationships can be appreciated. This would be useful for surgeons to assess their surgical approach or as a teaching tool for residents who are trying to gain expert knowledge of the anatomy. Identification of anatomy is all but necessary for image-guided surgery procedures, which have recently been introduced to the otology community [ 25 ] [ 26 ].

Cochlear implantation is a type of otologic surgery performed on individuals who experience profound to severe sensorineural hearing loss, i.e., individuals who are deaf. In

cochlear implant surgery, an electrode array is permanently implanted into the cochlea. The array is connected to a receiver mounted securely under the skin behind the patient's ear. When activated, the external processor senses sound, decomposes it into frequency information, and digitally reconstructs it before sending the signal through the skin to the internal receiver, which then activates the appropriate intracochlear electrodes causing stimulation of the auditory nerve and hearing. Current methods of performing the surgery require wide excavation of the mastoid region of the temporal bone. The preferred surgical approach is to gain access to the scala tympani along a path bounded posteriorly by the facial nerve, anteriorly and inferiorly by the chorda tympani and external auditory canal, and superiorly by the ossicles (see Figure 1). The scala tympani is one of the two principal internal cavities in the cochlea (the other is the scala vestibuli), and is the cavity in which the surgeon desires to implant the electrode array. The facial nerve, a tubular structure approximately 1.0-1.5 mm in diameter, is a highly sensitive structure that controls all movement of the ipsilateral face. If damaged, the patient may experience temporary or permanent facial paralysis. The chorda is a tubular structure approximately 0.3-0.5 mm in diameter. If the chorda is damaged, the patient may experience loss in the ability to taste. Damage to the external auditory canal can lead to a breach in sterility and open a potential avenue for future infection with possible implant extrusion. The ossicles are a set of tiny bones in the middle ear. Injury to the ossicles can result in damage to residual hearing. Using traditional approaches, the surgeon must excavate the temporal bone to pinpoint the location of each of these sensitive structures while approaching the cochlea. It is standard procedure for ear surgery to acquire a CT pre-operatively, so that the surgeon can assess the surgical approach. This requires the surgeon to visualize 3D anatomy by studying a volume of 2D radiograph images. Accurate mental reconstruction is clearly extremely difficult, if not impossible. A software

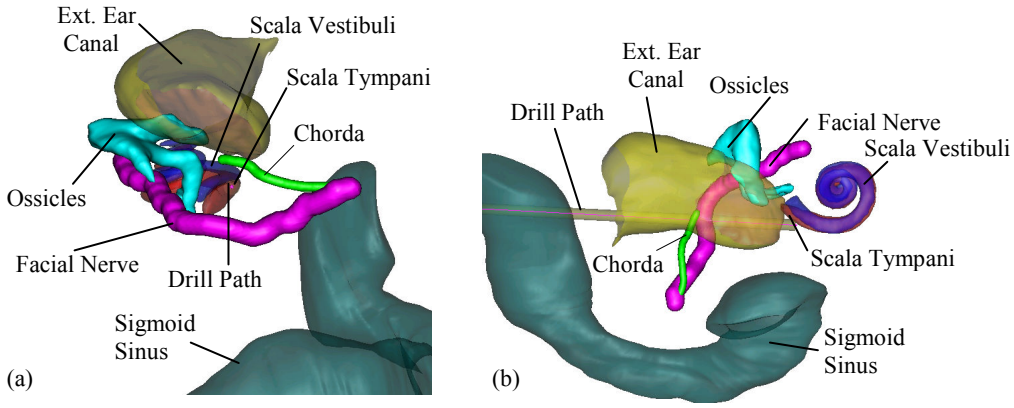


**Figure 1.** 3D rendering of the structures of the left ear. (a). Left-right, Posterior-Anterior view along the preferred cochlear implantation drilling trajectory. (b) Anterior-Posterior view.

package that could automatically identify ear anatomy would be highly beneficial in this application. Instead of just the 2D radiographs, the surgeon could also be presented with the segmented anatomical structures, which could be rendered in a 3D interactive visualization environment. This would allow very accurate spatial relationships to be easily appreciated. This may not only be helpful to the experts, but could also prove useful to students and residents who are trying to learn variations in the spatial relationships of ear anatomy. Furthermore, an image-guided approach to cochlear implant surgery is currently being developed [ 25 ] [ 26 ]. In this approach, a probabilistically safe drilling trajectory is automatically planned using an approach that requires segmentations of all these sensitive structures [ 37 ].

Access to the petrous apex is required for treatment of petrous apex lesions. The cochlea, semicircular canals, carotid artery, and facial nerve are all sensitive structures lying in close proximity to the preferred surgical approach. The carotid artery helps supply blood to the brain. If damaged, the patient could bleed to death. The semicircular canals are three canals connected to the cochlea. Damage to the canals could affect hearing and balance. Similarly to how

segmentation could be used as an aid for planning cochlear implant surgery, identification of these structures could be an aid for assessing the surgical approach to the petrous apex, and would allow automatic path planning for minimally invasive approaches [ 55 ].

A software package that is capable of identifying this set of anatomical structures, and that is validated in terms of accuracy and consistency would thus be useful and complete for the greater majority of ear surgery applications.

## Medical Image Segmentation

Accurately identifying 3D anatomy can be accomplished using a 3D image of the patient, e.g., a CT image. Automatic localization of structures in medical images has been a highly researched topic for over two decades. A typical method currently used in the field relies on a technique called image registration. A successful image registration is the process of finding a transformation that provides the best point-to-point mapping from one image to another. One example application that uses registration is the fusion of MR and CT data. Physicians often need to locate structures that are better resolved in CT and other structures that show up better in MR. But, if both MR and CT scans are performed on a particular patient, the data in the images will in general not be aligned. Registration can be used to find the transformation that aligns the two images, and the image data from both scans can be fused (see Figure 2). Once the CT and MR are fused into a single image, the physician can simultaneously locate and examine all the relevant structures. Automatic image registration is typically performed using intensity or surface based methods. Studies have demonstrated the superiority of intensity based method [ 59 ]. Intensity based registration is usually conducted by performing an iterative search to find a transformation that maximizes some similarity measure between the overlaid images. Studies

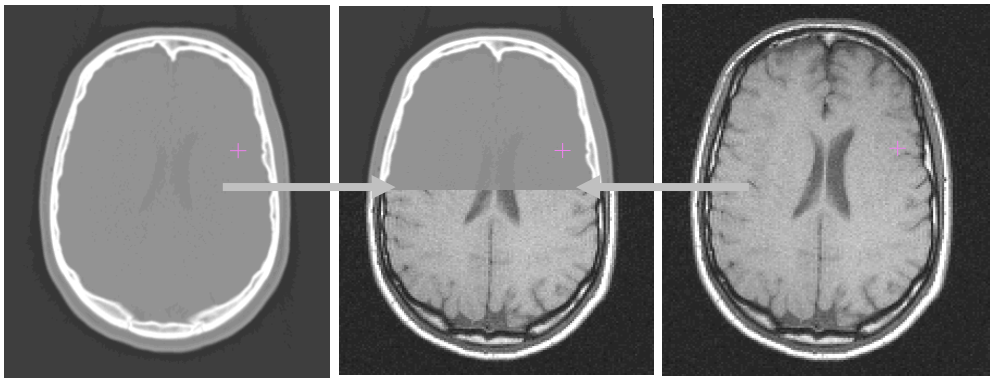

**Figure 2.** Image registration was performed between a CT (left) and MR (right) of the same patient. The registered images are fused in (middle).

have also shown that a statistical similarity measure called mutual information [ 30 ] [ 57 ] outperforms other intensity similarity measures [ 59 ]. While mutual information is a widely accepted measure of similarity, the types of transformations used to register images are many and include linear transformations, thin-plate splines, B-splines, and radial basis functions to name a few [ 43 ] [ 44 ] [ 45 ]. Registration can also be performed between images of different patients. This concept can be used to automatically localize structures in an unknown image. A particular structure can be manually localized in one image, the "atlas." Registration can be performed between this atlas and any unknown image. A perfect registration provides a perfect transformation from the atlas to the unknown image. Therefore, given knowledge of the structure location in the atlas, the location of the structure in the unknown image is provided by this transformation. This process is known as atlas-based segmentation and is a typical method for automatic structure localization (see Figure 3).

Although a perfect registration implies a perfect segmentation, registrations are, in general, imperfect. The greatest errors in atlas-based segmentation typically occur when attempting to localize small structures in a region where there are large inter-anatomical

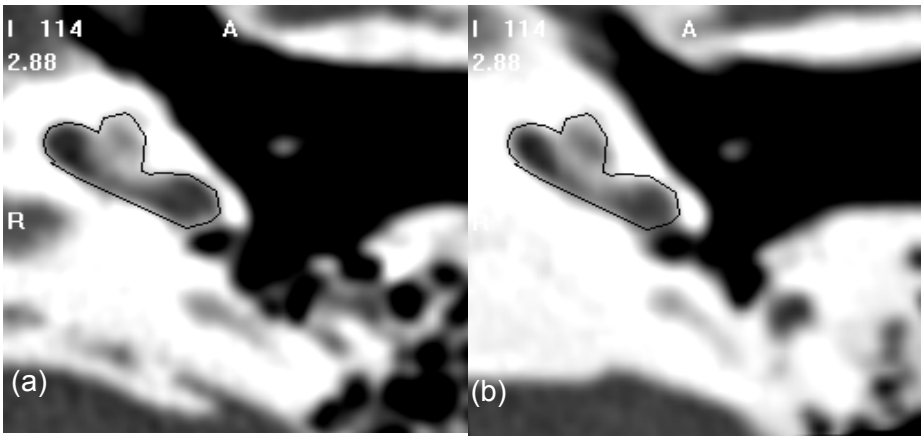


**Figure 3.** Example of atlas-based segmentation. The location of the cochlea is known in the atlas (a). After image registration, the cochlea surface can be projected onto the unknown image (b).

variations between subjects. In this case, the atlas is inherently different from the unknown image, and accurate registration is infeasible. Many segmentation algorithms have been developed that can be used when registration is inadequate. Particularly popular are snakes [ 24 ], active shape models [ 7 ], and level-set methods [ 47 ] for the segmentation of general structures. These algorithms all have strengths and weaknesses. Snakes and level-sets can impose smoothness and rigidity constraints on the shape of the structure, while optimizing an intensity based functional. Active shape models impose much more detailed constraints on the shape based on a training set. All of these approaches must be initialized, sometimes require interaction, and are subject to converging to local minima. In this dissertation, a combination of advanced segmentation techniques are used and developed for application to the specified ear anatomy in CT.

## Tubular Structure Segmentation

Much of the specified target anatomy is tubular in shape. Therefore a brief background on

tubular structure segmentation is presented here. Image registration, as discussed in the previous section, is commonly used to perform atlas-based segmentation. However, the greatest errors in atlas-based segmentation typically occur when attempting to localize small structures in a region where there are large inter-anatomical variations between subjects, which is quite often the case around tubular structures and is the case for many of the target structures specified in this dissertation. Therefore, other methods must be explored.

There is a large body of literature that has been published on algorithms developed specifically for the segmentation of tubular structures in medical images. For instance, Feng et. al. [ 13 ] propose a tubular structure deformable model. The model attempts to iteratively converge to the correct solution by minimizing an energy functional. For the energy function to be effective, the structure of interest must be enhanced, which requires a strong contrast between the structure and the background. Yim et. al. [ 61 ] propose another deformable model approach, but it relies on a manual initialization of the axis of the structure. Manniesing et. al. [ 32 ] propose a level-set based surface evolution algorithm. The authors construct a speed function that allows faster propagation along the axis of the structure. The evolving front can then be iteratively skeletonized and propagated. The speed function that the authors propose relies on contrast between the structure of interest and the background. Wesarg and Firle [ 58 ] use an iterative border detection scheme to approximate the surface of the structure and estimate the centerline. However, it is another approach that assumes the structure has been enhanced in the image. Santamaria-Pang et. al. [ 46 ] propose a probabilistic wave propagation approach to localize the structure's centerline, but assumes that the structure of interest is enhanced in the image. Olabarriaga et. al. [ 40 ] propose a minimal cost path approach for segmenting the coronary artery. Their approach relies on artery enhancement preprocessing procedure that

assumes the artery is brighter than the background. Hanssen et. al. [ 20 ] propose a semi-automatic approach for extracting nerve centerlines using fast marching methods. They also assume that the intensity value of the nerve contrasts with the background in the region of the nerve. Li and Yezzi [ 29 ] propose a 4-D minimal cost path approach to vessel segmentation, but the method also assumes consistent intensity characteristics along the length of the structure. Aylward et. al. [ 3 ] present an adaptive framework for tubular segmentation, but also assumes that the target anatomy is enhanced in the image. These approaches are examples of the state of the art of general approaches for tubular structure segmentations (Lesage et. al. [ 28 ] present a more comprehensive review of tubular structure segmentation methods), but these types of approaches are unlikely to work for tubular structures that are subject to variations in intensity along their length and lack of contrast with the background.

## The Facial Nerve and Chorda Tympani

The first tubular structures visited over the course of this work were the facial nerve and chorda tympani. These structures pose a problem for atlas-based segmentation and generic tubular structure segmentation approaches, because there is lack of local contrast and topological differences between volumes. Indeed, the facial nerve and chorda are surrounded by structures of comparable intensity values and pneumatized bone. Pneumatized bone appears as voids within the bone, which can vary in number and location across subjects. These characteristics are visualized in Figure 4 and Figure 5. In Figure 4, the facial nerve is identified by the dashed green arrow. Adjacent structures with boundaries that exhibit little contrast with the facial nerve are indicated by dotted red arrows. Several voids within the pneumatized bone (indicated by yellow solid arrows) of the subject in (a) do not appear in (c). The images in (b) and (d) show a portion

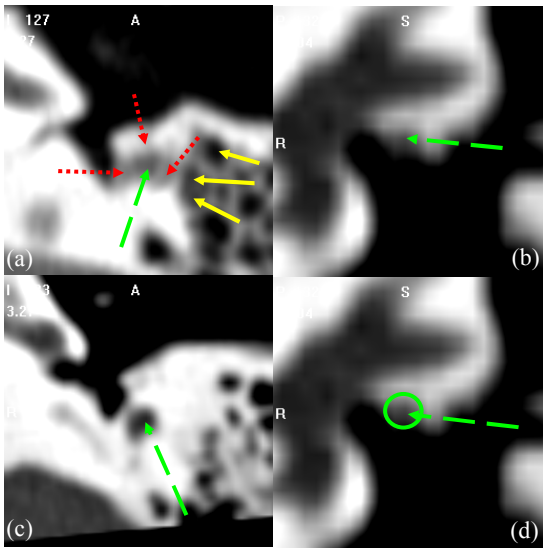


**Figure 4**. Illustration of the complexity of the problem. Panels (a) and (c) are axial slices of two different patients, and (b) and (d) show a coronal slice through the middle ear.

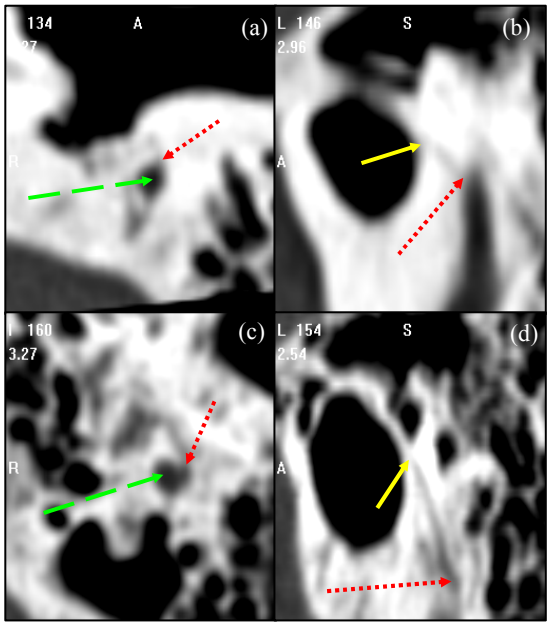


**Figure 5.** Illustration of the shape variation of the chorda between two patients with highly visible anatomy. Patient 1 (a,b) and 2 (c,d) in axial (a,c) and parasagittal (b,d) view.

of the facial nerve that lacks adequate contrast. Difficulties with the chorda are illustrated in Figure 5. The dotted red arrows indicate one extremity of the chorda, at its branch from the facial nerve (indicated by green dashed arrows). The solid yellow arrows indicate the opposing extremity of the structure, where it exits the temporal bone. The chorda can be seen as the dark rift of intensities between these two arrows. The chorda of the subject in the top row branches from the facial nerve at a more anterior orientation and at a position approximately 5 mm more superior relative to the anatomy of the subject in the bottom row. As a result, the latter chorda is much shorter and oriented at a different angle. Because of this lack of contrast, topological variation, and varying intensities, atlas-based or typical tubular structure based methods do not lead to results that are accurate enough. An example of an erroneous atlas-based segmentation of these structures is shown in Figure 6. In the figure, the atlas-based (indicated by "ab")

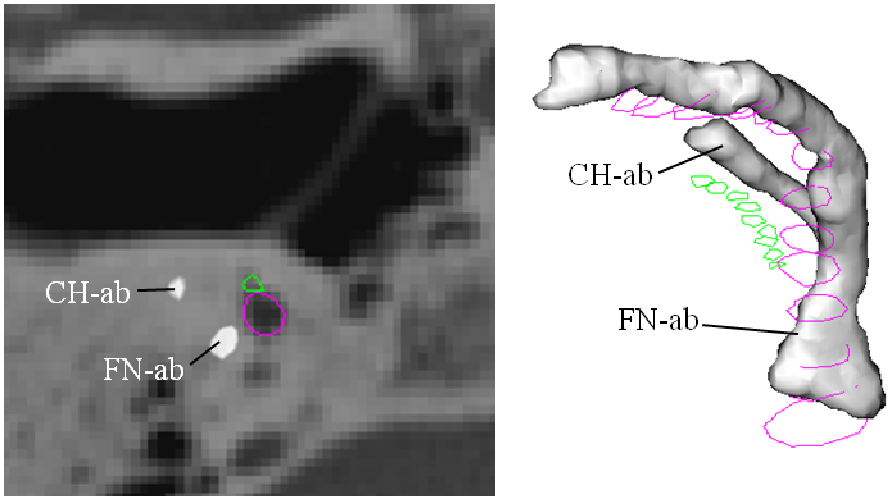


**Figure 6.** Example of atlas-based segmentation of the facial nerve and chorda versus gold standard delineations.

segmentations are shown as the white surfaces, and the manual contours of the chorda and facial nerve are shown in green and purple. As can be seen in the figure, atlas-based segmentation alone is not sufficient for segmenting this type of structure; however, atlas-based methods do consistently get close to the correct result. This led us to approaches for which we can use some information from registration for segmentation with another type of algorithm.

The problems that these structures pose for segmentation are (1) changing intensity characteristics along their length, (2) changing topology of surrounding pneumatized bone and background structures of similar intensity, (3) partial volume effects due to very small diameter, (4) areas with low contrast borders, and (5) for the chorda, large variations in shape and position relative to its size. Therefore, an algorithm must be developed that is not sensitive to these properties. Because all of these issues give rise to local minima, algorithms that perform segmentation by converging to local minima will be problematic. One class of tubular structure localization approaches is based on minimal cost paths. Generally, when this type of algorithm is

applied to medical image segmentation, the image voxels are treated as neighboring nodes in a graph, and a path between two endnodes, representing the centerline of the structure, is found that minimizes a cost function using a shortest path finding algorithm, such as [ 11 ]. The great strength of this approach is that it finds a path that globally minimizes a cost function. Because this type of approach is insensitive to local minima, a minimal cost path-based approach was chosen to create a solution for this segmentation problem.

The drawback of the algorithm is that the graph (this includes the nodes, edges connecting neighboring nodes, and costs associated with those edges) must be static to guarantee a global optimum. Therefore, no dynamic changes can be made during the search, and the graph cannot be dependent on the search. Because of this, there are issues that need to be addressed: (1) There is no standard way to account for locally changing intensity characteristics along the structure length, and (2) there is no standard way to incorporate shape information into the cost function. The solution that will be presented in Chapter 8 is to create a structure model from training samples that encodes local intensity and shape information, and to use image registration to roughly align this model with the target image to provide local priors. This approach will be shown to be powerful enough to achieve accurate results for these structures. Later, the same framework is used to accurately identify the optic nerves and chiasm in MR/CT, followed by the optic tracts in MR, demonstrating that this type of approach may be widely applicable.

## A General Tubular Structure Localization Algorithm

To use a minimal cost path algorithm for general tubular structure segmentation, several drawbacks need to be addressed: (1) Similarly to some other types of algorithms, the framework itself makes it inherently difficult to apply local prior knowledge, or to use training-based

knowledge of the structure of interest. Thus, it must be assumed that the structure of interest can be enhanced in the image prior to segmentation. This is not always possible as many tubular structures have varying intensity characteristics and borders that lack contrast. The algorithm developed for the facial nerve and chorda addressed this by relying on non-rigid registration to align a model of the local average characteristics of the structure of interest; however, a non-rigid registration to an atlas that is sufficiently accurate may not be available, and such a model may not be detailed enough for many applications. (2) The algorithm identifies a curve, rather than the 3D structure. Further processing must be applied to identify the full structure. (3) This algorithm is biased with respect to length, i.e., because the algorithm finds the path composed of edges with the minimum cumulative costs, it tends to favor paths composed of fewer edges. So, it is biased towards shorter distance, or straight, paths in the image. (4) The algorithm requires manual decision about what detection features should be used and tuning of parameters that weigh those features.

An analysis of these drawbacks has lead to a solution presented in this dissertation, which is based on high-dimensional, uniquely structured graphs that account for the limitations of the traditional 3D solution.

## Conclusions

In this chapter, background was presented on the target anatomy, on image segmentation methods, and on the general tubular structure localization algorithm to be presented. In Chapter 3, the details of the general registration algorithms that are used throughout this work are provided. Techniques that are used throughout this work to validate segmentation accuracy are presented in Chapter 4. A segmentation approach for several ear structures based entirely on

registration is presented in Chapter 5. This is followed by a technique for identifying intra-cochlear structures using an approach based on active shape models in Chapter 6. In Chapter 7, the minimal cost path algorithm, which will be used through the rest of the dissertation, is presented. In Chapter 8, the approach originally developed to identify the facial nerve and chorda tympani is presented. That method was modified for the segmentation of the optic nerves and chiasm, as well as the optic tracts, as presented in Chapters 9 and 10. The general algorithm for the localization of tubular structures developed for this dissertation and its application to the rest of the structures of the ear are presented in Chapter 11. Finally, Chapter 12 concludes the dissertation.

# CHAPTER III

# REGISTRATION ALGORITHMS

## Introduction

Image registration will be used for several purposes in this dissertation. This will include standard atlas-based registration, segmentation initialization, and as an aid for training segmentation algorithms. Image registration is the process of finding a transformation that provides the best point-to-point mapping from one image to another. Intensity-based registration is usually conducted by performing an iterative search to find a transformation that maximizes some similarity measure, like mutual information, between the overlaid images. While mutual information is a widely accepted measure of similarity, the types of transformations used to register images are many and include linear transformations, thin-plate splines, B-splines, and radial basis functions to name a few [ 43 ] [ 44 ] [ 45 ]. Affine and radial basis function type transformations are used for registrations in this dissertation.

## Algorithm Details

For affine registration, the approach suggested by Maes et. al. [ 30 ] is used. In this approach, transformations are computed by optimizing 12 parameters (translation, rotation, scaling, and skew) using Powell's direction set method and Brent's line search algorithm [ 42 ] to maximize the mutual information [ 30 ] [ 57 ] between the two images, where the mutual information between image $A$ and $B$ is computed as

$$MI(A,B) = H(A) + H(B) - H(A,B), \quad (1)$$

where $H(.)$ is the Shannon entropy in one image and $H(.\ ,\ .)$ is the joint entropy between these images. The entropy in the images is estimated as follows:

$$H(A) = -\sum_{i \in A} p(i) \log p(i), \tag{2}$$

in which $p(i)$ is the intensity probability density function, which throughout this work is estimated using intensity histograms with 64 bins.

Elastic registration is performed using an intensity-based algorithm developed at Vanderbilt called the adaptive bases algorithm (ABA) [ 43 ]. This algorithm models the deformation field that registers the two images as a linear combination of radial basis functions with finite support,

$$\boldsymbol{v}(\boldsymbol{x}) = \sum_{i=1}^{N} \boldsymbol{c}_i \boldsymbol{\Phi}(\boldsymbol{x} - \boldsymbol{x}_i), \tag{3}$$

where $\boldsymbol{x}$ is a coordinate vector in $\mathbb{R}^d$, with $d$ being the dimensionality of the images. $\Phi$ is one of Wu's compactly supported positive radial basis functions [ 60 ], and the $\boldsymbol{c}_i$'s are the coefficients of these basis functions. Normalized mutual information [ 51 ] is defined as

$$NMI(A,B) = \frac{H(A) + H(B)}{H(A,B)}. \tag{4}$$

The $\boldsymbol{c}_i$'s, which maximize the normalized mutual information between the images, are computed through an optimization process, which combines steepest gradient descent and line minimization. The steepest gradient descent algorithm determines the direction of the optimization. The line minimization calculates the optimal step in this direction. The algorithm is applied using a multiscale and multiresolution approach. The resolution is related to the spatial resolution of the images. The scale is related to the region of support and the number of basis functions. Typically, the algorithm is started on a low-resolution image with few basis functions

with large support. The image resolution is then increased and the support of the basis function decreased. This leads to transformations that become more and more local as the algorithm progresses.

## Overall Procedures

The affine registration scheme used in this dissertation, unless otherwise indicated, is devised as follows. First, the images are affinely registered over a global field of view with a downsampling factor of 4 in each dimension to correct for large differences in scale and orientation between the volumes. This is followed by another affine registration at full resolution over the region surrounding the structure(s) of interest. Registering the images first at low resolution, then at high resolution on a region of interest, both speeds up the process and improves registration accuracy in the region of interest. An example of this is shown in Figure 7. On the top row a global registration is applied between the left and right images of different subjects. Blending the resulting registered images produces the center image. As can be seen in the figure, a very nice global fit is achieved. However, the global optimum is not, in general, optimal for smaller regions of interest. Here, for instance, suppose we are interested in the ear. We can see that those structures are not optimally registered in an affine sense as shown in the left and right panels of the bottom row. If, however, we perform another registration in this region of interest, much better local results are obtained, as demonstrated in the middle panel, which is the right image registered to the left. For mutual information computation, the pdfs are estimated with 64-bin histograms. The starting optimization order for Powell's direction set algorithm is translation in x and y, rotation in x, y, z, translation in z, followed by scaling in x, y, z, and skew in x, y, z.

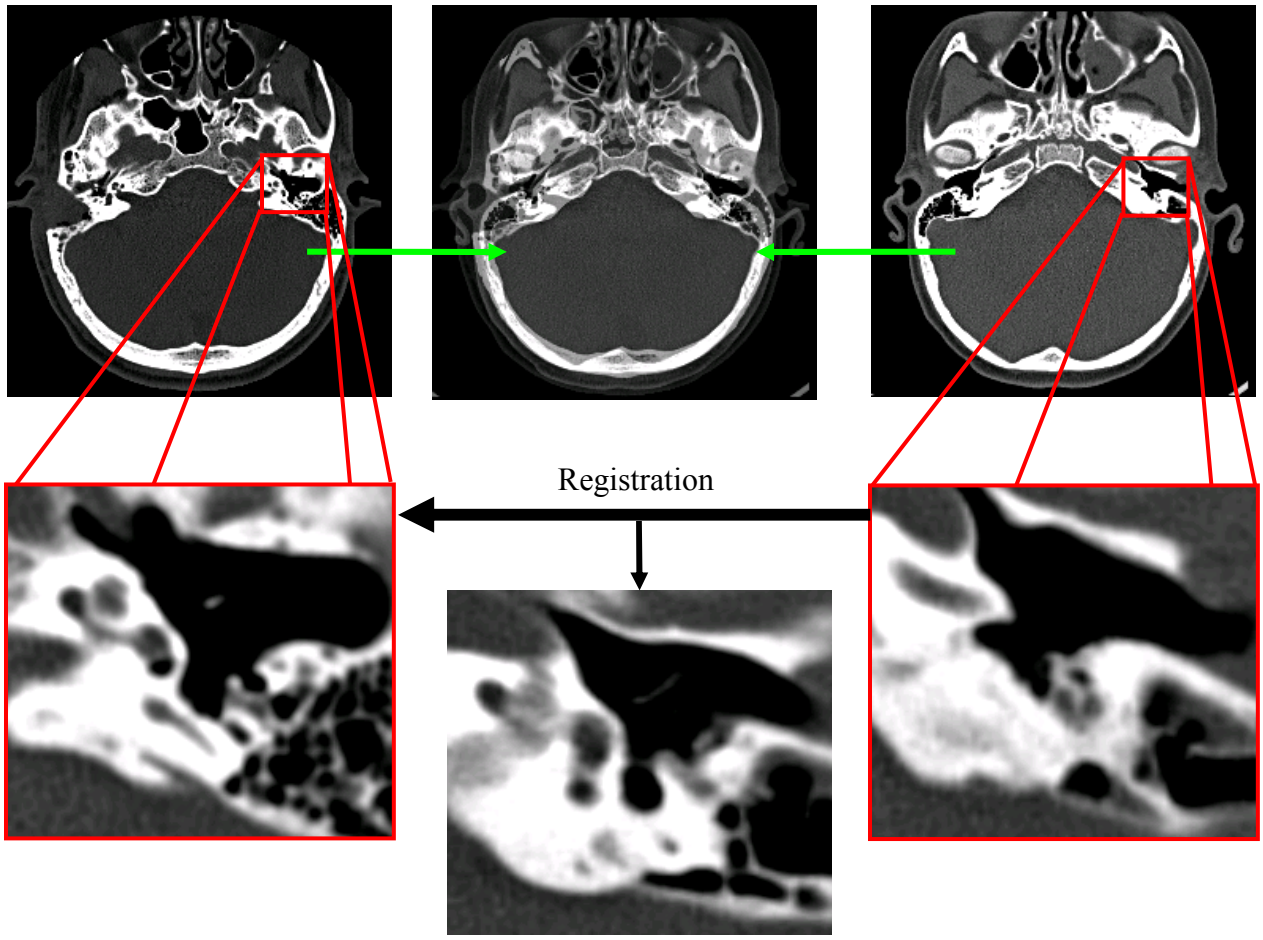


**Figure 7.** Example of global and local affine registration

In some cases, elastic registration will be applied after affine registration. When this is done, ABA is used. pdfs are estimated using 64-bin histograms. Sets of 3, 4, 10, and 14 bases in each dimension are used in a coarse-to-fine optimization approach. The images downsampled by a factor of two for registration. An example result from ABA is shown in Figure 8. In the figure, the top two images are being registered. The bottom row image is the top-left image non-rigidly registered to the top-right. As seen in the figure, ABA produces a non-rigid transformation that attempts to accurately warp the intensity information between two images.

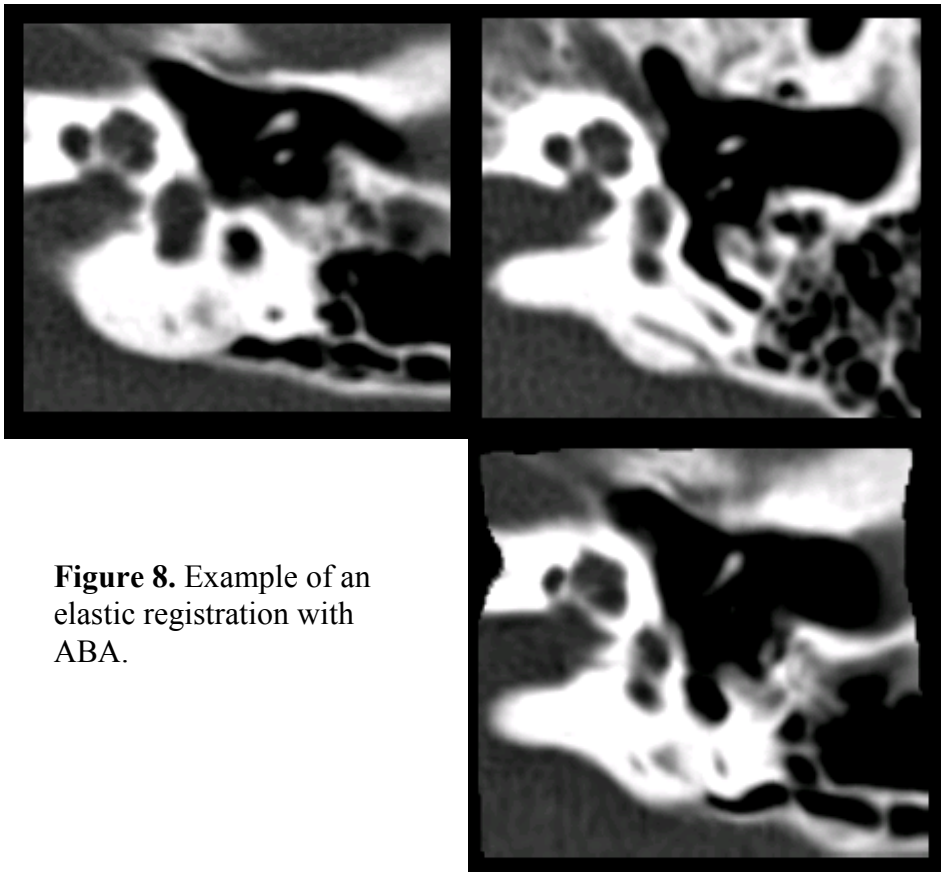

**Figure 8.** Example of an elastic registration with ABA.

## Conclusions

The presented registration methods will be referred to in several instances in this dissertation. Registration will be used for atlas-based registration, segmentation initialization, and as an aid for training segmentation algorithms.

# CHAPTER IV

# VALIDATION METHODS

## Introduction

Validation studies are necessary for analyzing the performance of segmentation algorithms. The standard validation approach is to compare the results of an algorithm, qualitatively and quantitatively, to a gold standard. The most common gold standards are expert drawn segmentations. In this dissertation, the gold standard will be expert drawn/edited segmentations. Quantitative validation approaches will be used, including surface error measures and volumetric overlap errors. These measures will be presented here.

## Surface Errors

Surface distance measures are a common error measure, with the benefit that they are not as sensitive to image resolution as voxel overlap-based measures. Surface errors are computed by measuring mean and maximum distances between automatic and manually segmented surfaces. The distance from a vertex $\vec{p}$ on surface $S_1$ to surface $S_2$ can be computed using the equation

$$\mathrm{d}(\vec{p}, S_2) = \min_{\vec{q} \in S_2}(\|\vec{p} - \vec{q}\|_2). \quad (5)$$

Then, the mean and maximum distances between the two surfaces can be computed as

$$\begin{aligned} \mathrm{d}_{\mathrm{mean}}(S_1, S_2) &= \mathrm{mean}_{\vec{p} \in S_1}\mathrm{d}(\vec{p}, S_2) \\ \mathrm{d}_{\mathrm{max}}(S_1, S_2) &= \max_{\vec{p} \in S_1}\mathrm{d}(\vec{p}, S_2) \end{aligned}, \quad (6)$$

In general, $\mathrm{d}_{\mathrm{mean}}(S_1, S_2) \neq \mathrm{d}_{\mathrm{mean}}(S_2, S_1)$ and $\mathrm{d}_{\mathrm{max}}(S_1, S_2) \neq \mathrm{d}_{\mathrm{max}}(S_2, S_1)$, i.e., the distance measures are not symmetric. Thus, symmetric mean and max errors are computed as

$\text{mean}\left(d_{\text{mean}}(S_1, S_2), d_{\text{mean}}(S_2, S_1)\right)$ and $\max\left(d_{\max}(S_1, S_2), d_{\max}(S_2, S_1)\right)$. A smaller distance between the automatic and manually segmented surfaces indicates higher accuracy.

## Overlap Errors

Volumetric overlap errors are commonly computed using the Dice index [ 10 ], which was introduced as a measure of error in the segmentation of medical images by [ 62 ]. The Dice index measures volumetric overlap between two surfaces $S_1$ and $S_2$ using the equation

$$\text{DSC}(S_1, S_2) = \frac{2N(S_1 \cap S_2)}{N(S_1) + N(S_2)}\ , \qquad (7)$$

where $N(S_1 \cap S_2)$ is the number of overlapping voxels within the surfaces, and $N(S_1) + N(S_2)$ is the sum of the number of voxels within each surface. The dice index ranges from 0, corresponding to no volumetric overlap, to 1, which corresponds to perfect volumetric overlap. A larger dice index between the automatic and manual segmentations indicates higher accuracy. The dice index tends to be unforgiving on smaller thinner structures where the ratio of surface area to volume is relatively large.

## Conclusions

Validation studies are a critical step for analyzing the performance of segmentation algorithms. The error measures presented in this chapter will be used to compare segmentation results to gold standard segmentations.

# CHAPTER V

## AUTOMATIC SEGMENTATION OF THE LABYRINTH, OSSICLES, AND EXTERNAL EAR CANAL

### Introduction

Expert knowledge of temporal bone anatomy requires years of training due to the highly complex shape and spatial arrangement of anatomical structures. Surgeons become very adept at using 2-D CT scans to mentally visualize normal 3-D anatomy. However, novel surgical approaches and pathological variations impede this ability on a case specific basis. While some surgeons become adept at this even more difficult task (mental visualization of atypical anatomy), current technology does not allow sharing or verification of the mentally visualized anatomy. Furthermore, this approach does not allow for precise quantitative measurement of spatial relationships.

One solution to this problem is to create case-specific anatomical models by identifying temporal bone anatomy in CT images. These models could then be presented to the clinician in the form of 3-D visualizations and would allow precise measurement of tolerances between critical structures and pathology, thus providing extremely useful information to the clinician for the purpose of surgical planning. Towards this end, previous work, which will be presented in the following paragraphs, has focused on extracting anatomical surfaces using a combination of thresholding and manual identification techniques. Thresholding consists of identifying structures based on their range of CT intensity values. Thresholding is non-specific, often including neighboring anatomical areas that have the same intensity as the anatomy of interest. Manual identification consists of using trained observers, e.g. radiologists or radiology

technicians, to trace anatomy of interest on each slice of a CT scan. This process is time consuming, often requiring hours to delineate the anatomy of a single patient.

Using a combination of thresholding and manual identification, Jun et. al. [ 23 ] extracted the facial nerve, ossicles, labyrinth, and internal auditory canal in twenty spiral high-resolution CT volumes. The authors utilized this information to report on spatial relationships within the temporal bone. Wang *et al.* [ 54 ] created an extensive library of surfaces by combining serial, computer-assisted manual delineations of 2-D photographic images of histologically-sectioned, frozen temporal bone. Wang's intent was to create a tool for studying anatomical relationships. The main limitation of these and other works [ 53 ] [ 33 ] [ 19 ] is that they require manual delineation of anatomy.

In this chapter, a combination of techniques to automatically identify vital anatomy of the temporal bone in unknown CT scans is presented. The study is focused on three structures of the temporal bone that are relevant to otologic surgery, specifically, the external auditory canal, ossicles, and labyrinth. The methods require approximately 6 minutes to fully process a single scan on a standard computer. The methods discussed in this chapter were published in Noble et. al. [ 36 ].

## Methods

The process used to automatically identify the labyrinth, ossicles, and external auditory canal relies on atlas-based registration. Image registration is introduced in Chapter 2 and the specific algorithms used are presented in Chapter 3. In this application, an atlas was constructed with a CT of a "normal" subject, in which the ossicles, cochlea, and external auditory canal were carefully labeled. To segment a target CT, this atlas is affinely and non-rigidly registered to the

CT producing a compound non-rigid transformation. The structure labels are then transferred from the atlas to the target CT through this transformation, identifying the position of the structures.

To validate our process, we quantified automated identification error as follows. First, the temporal bone structures were manually identified in all CT scans by a student rater, then verified and corrected by an experienced surgeon. Then, surface errors were computed between automatic and manual segmentations.

## Results

The software takes as input clinical CT images. The software runtime on a Windows based machine (we use a 2.4 GHz Intel Xeon processor machine) is approximately 6 minutes for a typical temporal bone CT scan (768 x 768 x 300 voxels). Voxel dimensions of the scans used in this study were approximately 0.35 x 0.35 x 0.4 $mm^3$. The software outputs the identified structures in the form of surface files that can be loaded into generic rendering software.

Identification error was measured individually for the ossicles, external auditory canal, and labyrinth. Table 1 shows error results for the surfaces automatically localized using our methods. As can be seen from this table, overall mean errors for the labyrinth, ossicles, and external auditory canal are around 0.5 mm (less than two voxels), with low median errors for each volume. Over the vast majority of these surfaces, there is excellent agreement between manually and automatically generated surfaces.

Relatively high maximum errors were seen at the labyrinth and auditory canal in three of fourteen ears. For the labyrinth, this error occurs at the most superior portion of the superior semicircular canal. The automatic identification errs laterally in this region and causes a sharp

**Table 1.** This table shows error calculations for external auditory canal, labyrinth, and ossicles. All measurements are in millimeters.

| | Atlas-Based Registration | | |
|---|---|---|---|
| | Auditory Canal | Labyrinth | Ossicles |
| Max | 2.551 | 2.081 | 1.774 |
| Mean | 0.469 | 0.419 | 0.383 |
| Median | 0.430 | 0.388 | 0.358 |

increase in the error. Error in the auditory canal occurs at the anterior wall of the tympanic membrane. In general, the membrane is poorly contrasted in CT. Where the error was observed, the membrane was not visible in the image, making accurate manual or automatic identification very difficult. The maximum errors at these points were on the order of 6 voxels.

Figure 9 shows the qualitative results obtained with the presented localization algorithms. A 3D visualization of the structures and a CT cross section of the structures are shown for selected ears. All surfaces appear qualitatively accurate. While difficult to appreciate in a 2-D presentation, these surface renderings are rotatable, allowing appreciation of 3-D relationships from any perspective angle. Also included in the renderings are surfaces representing the facial nerve and chorda tympani. These structures were segmented using an approach that will be presented in Chapter 8.

## Conclusions

To safely and precisely perform neurotologic surgery, surgeons require intimate understanding of the 3-D spatial relationships of the anatomical structures of the temporal bone. For most common procedures, this understanding becomes second-nature as surgeons easily translate 2-D CT scans into mental 3-D models. However, in cases of altered pathology and/or uncommon surgical

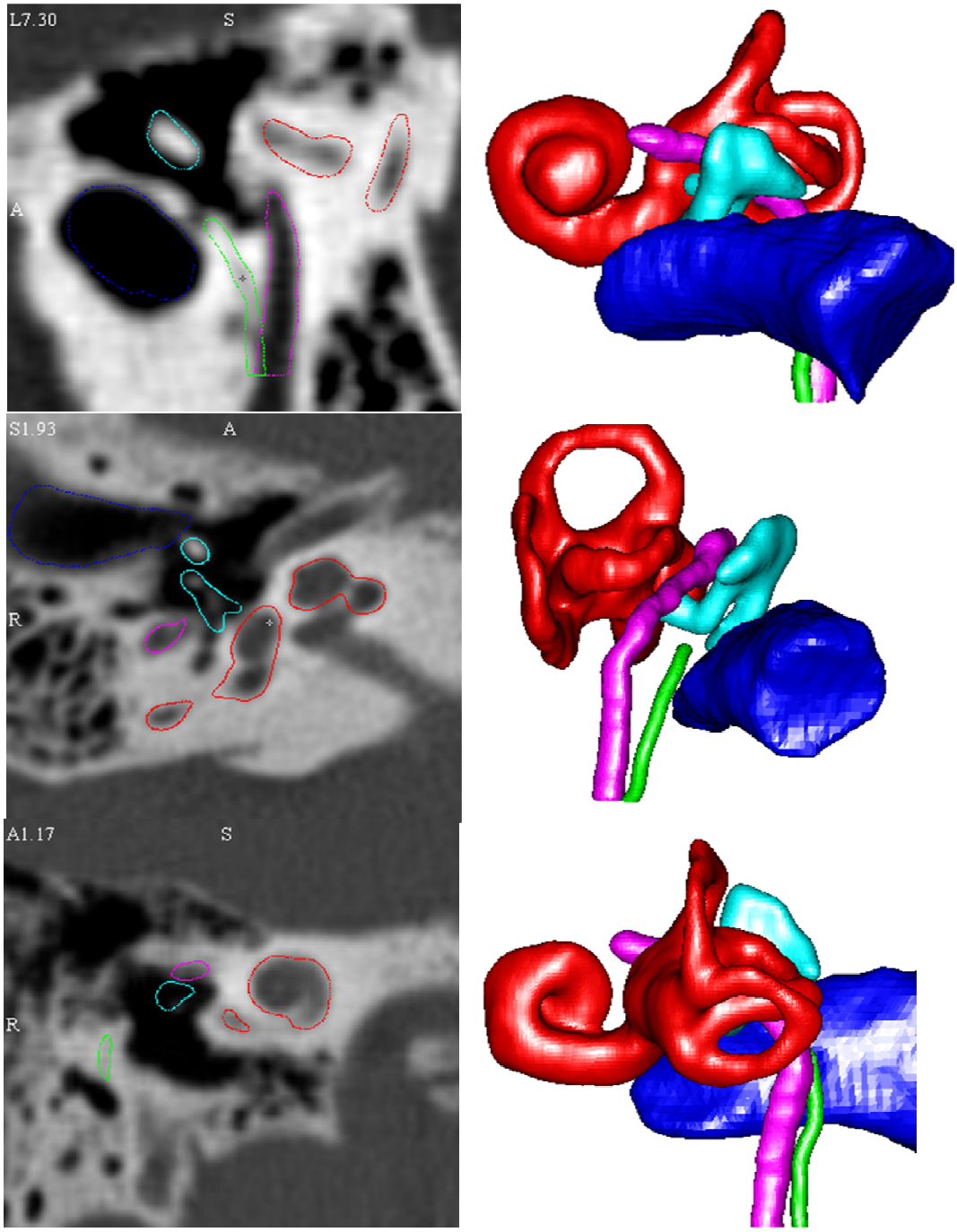


**Figure 9.** 3D renderings and CT cross section automatic segmentation results for three volumes. All segmentation results appear qualitatively accurate.

approaches this task is difficult. In such situations, computer graphics can be used to enhance the surgeon's knowledge by providing virtual 3-D models that can be viewed from multiple angles. To date, such applications have been limited by the amount of time and effort necessary to

manually delineate the anatomy of interest. Manual localization – manually identifying a structure in each and every CT slice – is time consuming, often requiring hours to perform. Thresholding-based, automated algorithms do not provide the submillimetric resolution necessary for neurotologic surgery. For clinical applications, the ideal system would be accurate and automated requiring little to no input from the end-user.

Herein, results of a method are presented, which, for normal anatomy, achieve this ideal – accurate, automated anatomical identification and 3-D presentation of temporal bone anatomy from clinical CT scans. Using atlas-based segmentation (transforming a known specimen to match an unknown specimen), we have shown that the labyrinth, ossicles, and external auditory canal can be repeatedly identified with a mean error of ~0.43 mm.

The error results of this small study suggest that our methods are accurate enough to be used in a variety of applications. However, no identification method is perfect, and it is important that the clinician is made aware of error in automatic structure identification. In order to explicitly let the clinician know where error exists, the software first presents identified structures as shown in the sagittal, coronal, and axial views of Figure 10. In these images, for example, the outlining of the cochlea and facial nerve are quite good, but the outlining of the I-S joint in the coronal scan is slightly inaccurate. Such qualitative assessment of the results must be performed by the user before the software creates the 3-D visualizations, which appear in the lower right panel of the figure, or allows measurement of spatial relationships. It is our intent that this process will alert the end-user to imperfections that could occur on a case-specific basis.

There are several applications in the field of neurotology in which there exists a potential role for such software. (1) From an educational perspective, it would enhance teaching, allowing 3-D relationships to be appreciated from multiple perspective angles. (2) From a clinical

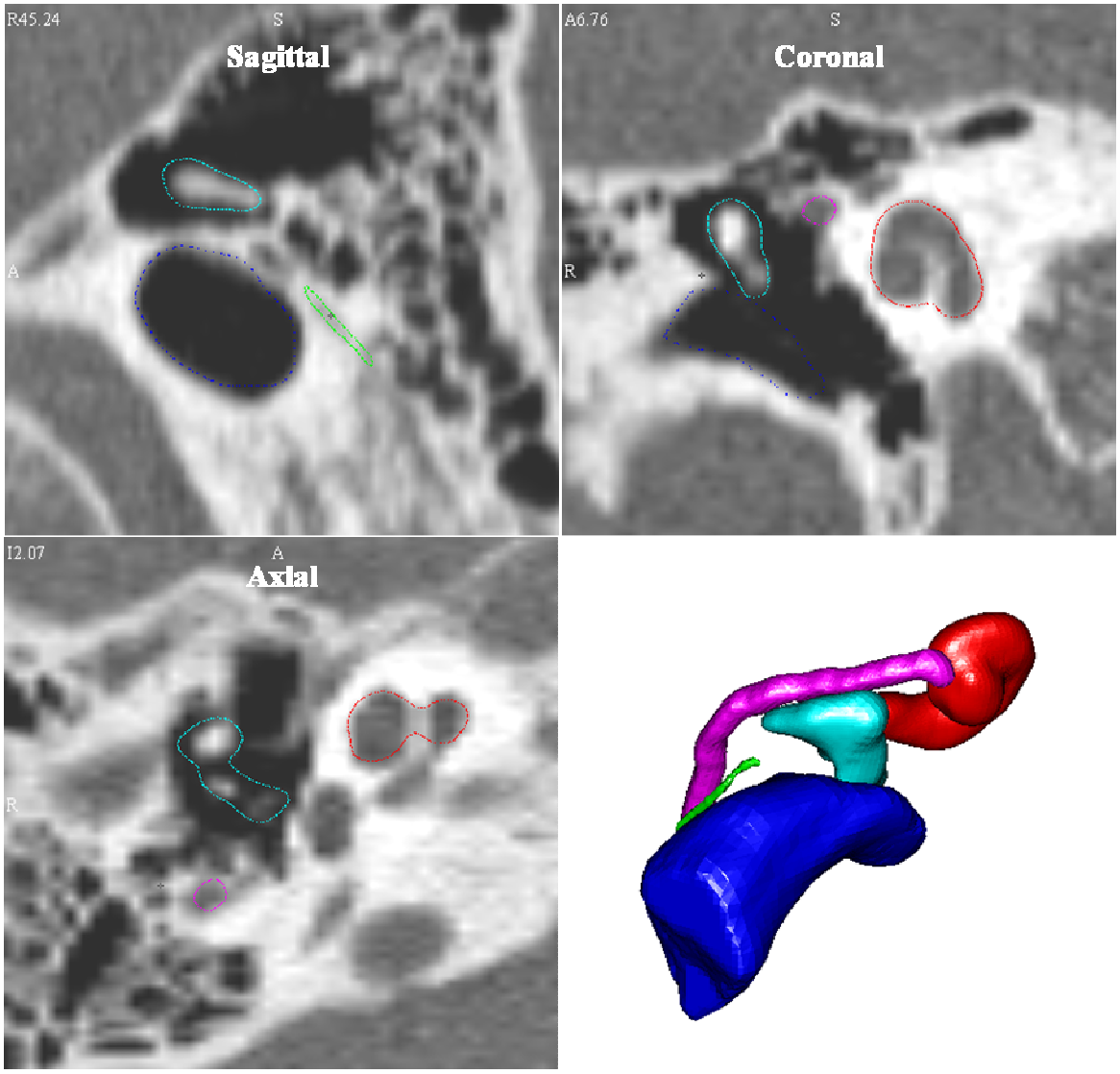


**Figure 10.** Localized structure contours in the CT planes and 3D renderings presented to the end-user.

standpoint, prior to surgery, a patient specific 3-D model of temporal bone anatomy could be used as a final anatomical check-point (e.g. "How wide is the facial recess on this patient?" "How far lateral does the facial nerve come with respect to the lateral aspect of the horizontal semicircular canal?"). Furthermore, with the addition of other anatomical structures, it will be possible to analyze different surgical approaches before cutting skin (e.g. "Would an infracochlear approach to the petrous apex be feasible?").

A clinical application of interest, which motivated the development of this automated localization technique, is a novel surgical approach called percutaneous cochlear implantation. In this previously described technique [ 25 ] [ 26 ], a single drill hole is made from the lateral surface of the skull surface to the cochlea avoiding vital anatomy. The drilling trajectory is planned pre-operatively on CT scans and then implemented with image-guided surgical technology. During early work in this project, it was discovered that trajectory planning was difficult due to poor visualization of the spatial arrangement of anatomy in 2-D CT slices. However, as all of the vital anatomy can be automatically localized prior to planning, the drilling trajectory can be visualized and verified in a 3D environment.

# CHAPTER VI

## AUTOMATIC SEGMENTATION OF INTRA-COCHLEAR ANATOMY

### Introduction

Cochlear implant surgery is a procedure performed to treat profound hearing loss. In a typical surgery, a mastoidectomy is performed, excavating bone posterior to the auditory canal to gain access to the cochlea while avoiding damage to vital structures. The sensitive structures along the traditional surgical approach are shown in Figure 11. The physician drills an approximately 1.0 mm hole through the bone surrounding the cochlea through which the electrode is inserted; this process is referred to as the cochleostomy in the remainder of this chapter. The electrode stimulates the auditory nerve to induce the sensation of hearing.

Research has suggested that implanting the electrode in a particular cavity of the cochlea, the scala tympani, results in better hearing restoration [ 48 ]. But, it has also been reported that at most about 73% of cochlear implant procedures result in full insertion into the scala tympani [ 2 ]. The remaining 27% of implants are either fully inserted into the scala vestibuli, another cavity in the cochlea, or are initially inserted into scala tympani and then cross through the thin membranes (basilar and Reissner's membranes) separating the scala tympani from the scala vestibuli (see Figure 12). For all practical purposes, the third intra-cochlear compartment, the scala media, can be ignored as it occupies a small minority of the intra-cochlear volume. Research has also suggested that the likelihood of scala tympani insertion is maximized using "soft" technique [ 27 ] [ 22 ]. In this approach, the implant is threaded at a tangential angle into

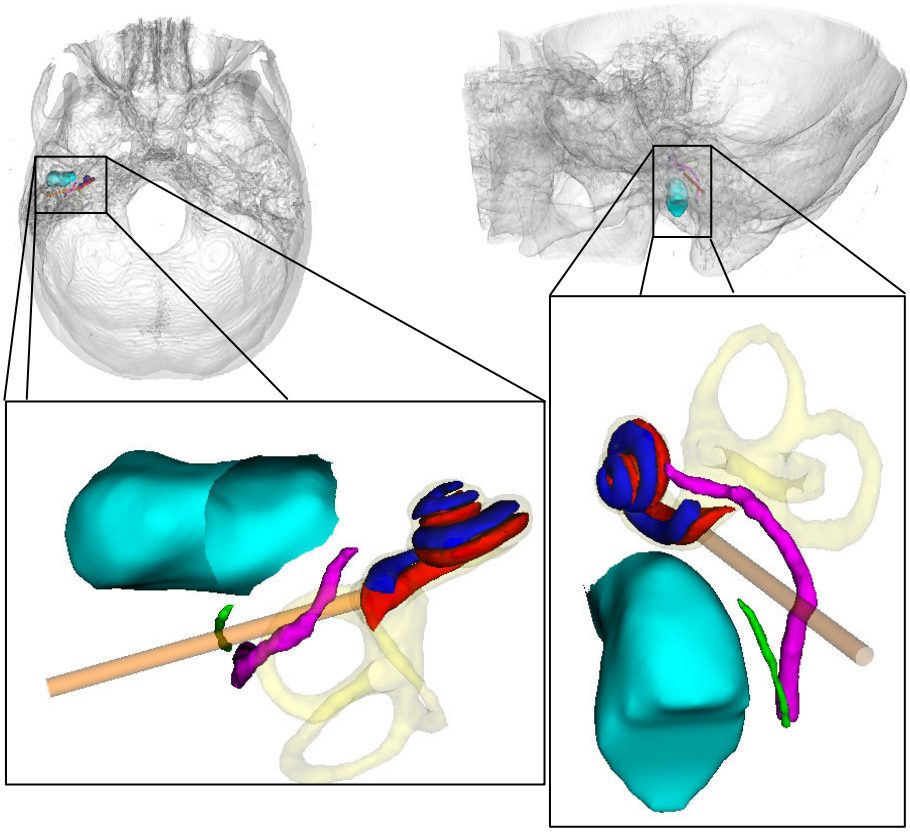

**Figure 11.** Superior-to-inferior view (left) and lateral-to-medial view (right) of ear anatomy. Shown are the cochlear labyrinth (yellow), facial nerve (purple), chorda tympani (green), auditory canal (blue-green), scala tympani (red), scala vestibuli (blue), and the traditional surgical approach (orange tube).

the scala tympani, attempting to inflict as little stress as possible on the soft tissue within the cochlea.

The success of the soft technique is largely dependent on the angle of electrode insertion and the position of the cochleostomy – the angle should be tangential to the scala tympani and the cochleostomy should be centered on the scala tympani in the basal turn region (first turn of the spiral–see Figure 11). Because the scala tympani is not visible in surgery, the surgeon must use other landmarks as a guide to estimate scala tympani position and orientation, a process requiring a great deal of expertise. Errors can occur here due to the imprecise nature of

landmark-based, manual navigation as well as to inter-patient variations in the spatial relationship between the scala tympani and surgical landmarks.

A method that could localize the scala tympani in clinical images would thus help surgeons in determining the proper cochleostomy position and insertion angle, and potentially result in more substantial hearing restoration. For typical cochlear implant procedures, a conventional CT is acquired so that the surgeon can identify abnormalities along the surgical approach. But, the basilar membrane is invisible in these images, which makes automatic identification impossible using only the information available in these scans (see Figure 12). To address this issue, we propose a method that complements the information available in these images. This technique permits the fully automatic and accurate segmentation of both scalae.

The method we propose is based on a deformable model of the cochlea and its components. To create such a model, we have used ex-vivo μCTs of the cochlea. In these scans,

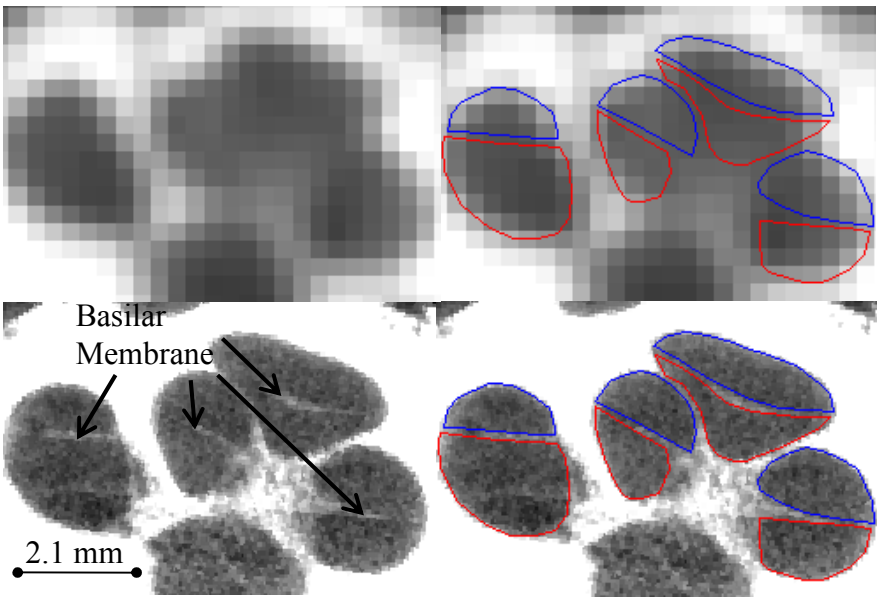


**Figure 12.** CT (top) and μCT (bottom) of a cochlea specimen. Delineated in the right panels are the scala tympani (red) and scala vestibuli (blue).

which have a spatial resolution that is much higher than the spatial resolution of clinical scans, intra-cochlear structures are visible (see Figure 12). Thus, these structures can be segmented manually in a series of scans. The set of segmented structures can be used subsequently to create an active shape model, as proposed by Cootes et al. [ 7 ]. Once the model is created, it can be used to segment the structures of interest, which are only partially visible in the conventional CT images, using the limited information these images provide. In the following sections, the methods used to create the models and how these models are used to segment intra-cochlear structures in clinical CTs are presented. This is followed by the results obtained by these methods, and finally conclusions are presented. The methods presented in this chapter were first presented at the SPIE conference on medical imaging 2010 (Noble et. al. [ 38 ]).

## Model Creation

*Data*. The data set used to build the model consists of image sets of six cadaveric cochlea specimens. For each specimen, one μCT image volume was acquired with a Scanco μCT. The voxel dimensions in these images are 36 μm isotropic. For five of these specimens, one conventional CT image volume was acquired with a Xoran XCAT fpVCT scanner. In these volumes, voxels are 0.3 mm isotropic. In each of the μCT volumes, the scala vestibuli and scala tympani were manually segmented. Figure 12 shows an example of a conventional CT image and its corresponding μCT image.

*Outline.* The approach we use to model intra-cochlear anatomy is summarized in Figure 13. It requires (1) establishing a point correspondence between the structures' surfaces, (2) using these points to register the surfaces to each other with a 7 degrees of freedom similarity transformation

(1) Establish point correspondence
   a. Rigid and non-rigid registration of training volumes to reference volume
   b. Deform training surfaces into reference space
   c. Find closest points from reference to deformed training surfaces
(2) Rigid + scaling register original training surfaces to reference surface with point registration
(3) Compute modes of variation using eigenanalysis

**Figure 13.** Constructing a point distribution model from a set of surfaces.

(rigid plus isotropic scaling), and (3) computing the eigenvectors of the registered points' covariance matrix.

*Establishing Correspondence.* To establish correspondence between surfaces, one μCT volume was arbitrarily chosen as a reference volume, $I_0$, and the remaining training volumes, $\{I_j\}_{j=1:N-1}$, were registered to the reference using the affine and elastic methods discussed in Chapter 3, creating the set of non-rigid transformations $\{\Psi_j\}_{j=1:N-1}$ such that $I_j \approx \Psi_j(I_0)$. The affine registration step required manual initialization, since there was no standard orientation for the samples in the scans. The manually segmented reference surface $S_0$ was then deformed through each set of registration transformations into each target image space, creating the set $\{\Psi_j(S_0)\}$. Residual registration errors were corrected by manually adjusting the surfaces using a tool developed for this purpose. These additional transformations were combined with the transformations computed automatically to produce the set of compound transformations $\{\Psi_j^{\mathrm{M}}\}$. Finally, for each *i*th vertex on each *j*th deformed reference surface $\Psi_j^{\mathrm{M}}(\vec{S}_{0,i})$, the closest point on the respective training surface $\vec{p} \in S_j$ was found. A correspondence between this point $\vec{p}$ and the reference surface vertex $\vec{S}_{0,i}$ was established. Equivalently,

$$\vec{S}_{0,i} \leftrightarrow \mathrm{p}_{S_j}\left(\vec{S}_{0,i}\right) \quad (8)$$

such that $\mathrm{p}_{S_j}\left(\vec{S}_{0,i}\right)$ is the closest point on surface $S_j$ to the point $\Psi_j^{\mathrm{M}}\left(\vec{S}_{0,i}\right)$, equivalently,

$$\mathrm{p}_{S_j}\left(\vec{S}_{0,i}\right) = \mathrm{argmin}_{\vec{p}\in S_j}\left(\left\|\Psi_j^{\mathrm{M}}\left(\vec{S}_{0,i}\right) - \vec{p}\right\|_2\right). \quad (9)$$

This results in one corresponding point on each *j*th training surface for every *i*th reference point. Using image registration to establish a correspondence between point sets was inspired by the work of Frangi et. al. [ 16 ].

Once correspondence is established, each training surface is registered to the reference surface with a similarity transformation (rigid plus isotropic scaling), computed using standard point registration techniques [ 1 ]. Correspondence is established separately for the scala tympani and scala vestibuli, however, point registration is performed taking both structures into account simultaneously. This is done to maintain accurate inter-structure spatial relationships so that the shape model can be constructed with both structures simultaneously.

*Computing the Shape Model.* To build the model, the principal modes of shape variation are extracted from the registered training shapes. This is computed according to the procedure described by Cootes et. al. [ 7 ]: First, the covariance matrix of the point sets' deviation from the mean shape is computed as

$$C = \frac{1}{N}\sum_{j=1}^{N}\left(\vec{w}_j - \bar{w}\right)\left(\vec{w}_j - \bar{w}\right)^{\mathrm{T}}, \quad (10)$$

where the $\vec{w}_j$'s are the individual shape vectors and $\bar{w}$ is the mean shape vector. The shape vectors are constructed by stacking the 3D coordinates of all the points composing each structure

into a vector. The modes of variation in the training set are then computed as the eigenvectors $\{\vec{u}_j\}$ of the covariance matrix such that,

$$\{\lambda_j, \vec{u}_j\}_{j=1}^{N-1} \quad : \quad \lambda_j \vec{u}_j = C\vec{u}_j. \tag{11}$$

There are at most *N*-1 non-zero eigenvalues, $\{\lambda_j\}$, because the rank of a covariance matrix of *N* linearly independent samples is *N*-1. The range of shapes within the training set can then be described by adding a linear combination of the eigenvectors, to $\overline{w}$, the mean shape. The eigenvalue associated with each eigenvector is equal to the variance explained by the *j*th mode of variation in the training set. Typical structures can be closely approximated by using a linear combination of the eigenvectors within two standard deviations of the mean. These modes of variation are extracted for the combined shape of both the scalae for all the samples in the training set.

## Segmentation Using the Active Shape Model

The procedure we use for segmentation with an active shape model is outlined in Figure 14. As can be seen in the figure, the overall process follows the traditional active shape model approach, i.e., (1) the model is placed in the image to initialize the segmentation; (2) better solutions are found while deforming the shape only in ways that are described by the pre-computed modes of variation; and (3) eventually, after iterative shape adjustments, the shape converges, and the segmentation is complete. The following sections detail this approach.

*Initialization.* The first step to perform segmentation is to provide an initial position for the model in the target image. We chose to automate the initialization process by using an image registration-based initialization approach, in which an "atlas" is registered to the volume to be

(1) Initialize model
    a. Affine and non-rigid registration of atlas to target volume
    b. Project corresponding model points from atlas to target
    c. Fit the model to the points given by registration
(2) Search for better solutions
    a. Find an optimal adjustment for each model point
    b. Fit the shape model to the set of adjusted points
(3) Iterate (2) until convergence

**Figure 14.** Performing segmentation with the active shape model.

segmented, the "target." This atlas is a conventional CT volume in which the position of the scalae is known. To create this atlas, a full head clinical CT was selected (in plane voxel size of 0.218 x 0.218 $mm^2$, and slice thickness of 0.8 mm with 0.4 mm overlap). Next, a non-rigid transformation between the atlas volume and $I_0$, the μCT of the cochlea specimen used to create the model, was computed using image registration (see Chapter 3). Because the atlas volume and $I_0$ have very different fields of view, it was necessary to manually initialize this registration. Visually identifiable errors in the resulting transformation were manually corrected. Finally, the scalae model points were projected from $I_0$ to the atlas volume to produce a full head volume in which the position of the scalae and of the basilar membrane are known.

The atlas is used at segmentation time to initialize the model using the steps summarized in Figure 14: (a) First, an automatic non-rigid registration (see Chapter 3) is computed between the atlas and the target CTs. (b) Next, the scalae model points, which have known positions in the atlas, are transferred to the target image using the registration transformation. (c) Finally, to initialize the search, the shape model is simply fit (the fitting procedure is described in the following section) to this set of points. The fitting procedure constrains the shape of the point-set

to be similar to that of typical scalae, removing any unnatural shape changes that may have occurred due to the non-rigid registration transformation.

*Searching Scheme.* Once initialized, the optimal solution is found using an iterative searching procedure. At each search iteration, (a) an adjustment is found for each model point, and (b) the model is fit to this set of candidate adjustment points (see Figure 14). To find the candidate points, two approaches are used, one for "external" and another for "internal" model points. When the model was created, model points corresponding to the exterior of the cochlea were labeled as external points. All other model points, which are positioned inside the cochlea were labeled as internal points. For the internal points, it is impossible to determine the best adjustment using local image features alone, because there are no contrasting features inside the cochlea in CT. Therefore, the original initialization positions for these points, which were provided by image registration, are used as the candidate positions. The registration transformation, as our results will show, is sufficiently accurate to provide this useful information to the segmentation process. External points lie in regions with contrasting features in CT. Therefore, candidates for external points are found using line searches to locate strong edges. At each external point $\vec{y}_i$, a search is performed along the vector normal to the surface at that point. The new candidate point is chosen to be the point with the largest intensity gradient over the range of -1 to 1 mm from $\vec{y}_i$ along this vector.

The algorithm uses the approach just described to find a new candidate position for each point in the model. The next step is to fit the shape model to these candidate points. We do this in the conventional manner described by Cootes. A standard 7 degree of freedom point registration is performed, creating similarity transformation *T*, between the set of candidate points $\{\vec{y}_i'\}$ and

the mean shape $\{\bar{w}_i\}$, where $\bar{w}_i$ are the 3D coordinates of the *i*th point in the mean shape. Then, the residuals

$$\vec{d}_i = T(\vec{y}_i') - \bar{w}_i \quad (12)$$

are computed. The modes of variation that were previously extracted are fitted to the residuals by taking the inner product between each eigenvector with the full residual vector,

$$b_j = \vec{u}_j{}^{\mathrm{T}} \vec{d}, \quad (13)$$

where $\vec{d}$ is composed of $\{\vec{d}_i\}$ stacked into a single vector. Eq. ( 13 ) results in a scalar $b_j$ corresponding to each eigenvector $\vec{u}_j$ that quantifies how much of the deviation of this shape from the mean shape is explained by $\vec{u}_j$. The final approximation to the shape is computed by passing the sum of the scaled eigenvectors plus the mean shape through the inverse transformation, equivalently,

$$\vec{y}_i = T^{-1}\left(\bar{w}_i + \sum_{j=1}^{N-1} b_j \vec{u}_{j,i}\right), \quad (14)$$

where $\vec{u}_{j,i}$ is the *i*th 3D coordinate of the *j*th eigenvector. As suggested by Cootes, the magnitude of the $b_j$'s are constrained such that

$$\sqrt{\sum_{j=1}^{N-1} \frac{b_j^2}{\lambda_j}} \leq 3, \quad (15)$$

which enforces the Mahalanobis distance between the fitted shape and the mean shape to be no greater than 3. This constrains the shape such that its deviation from the mean is reasonable relative to the shape variations described by the training set.

At each iteration, new candidate positions are found and the model is re-fitted to those candidates. The active shape model converges when re-fitting the model results in no change to the surface.

## Validation

Segmentation was performed on CT's of five of these cochlea specimens using a leave-one-out approach, i.e., the volume being segmented is left out of the model. Thus, only four modes of variations were available for each segmentation validation. Because these samples were excised specimens, rather than whole heads, the initial registration with the full head CT used as the atlas required manual initialization—however, in practice the approach is fully automatic. To validate the results, the set of corresponding μCT volumes are exploited. Each μCT was rigidly registered to the corresponding CT of the same specimen. The manually delineated scalae segmentations were then projected from the μCT to CT space. Finally, dice indices and surface errors were computed between automatic segmentations and the registered manual segmentations to validate the accuracy of our results.

## Results

The model was constructed with a total of approximately 15000 unique sets of corresponding surface points. The eigenvalues of the 5 independent modes of variation for the model are given in Table 2. At a range of two standard deviations, the maximum displacement of any point on the scala tympani or scala vestibuli surfaces are approximately 1.0 mm, and mean displacements are approximately 0.15 mm. Points on the basal turn (the first turn of the spiral) vary by approximately 0.35 mm. In the basal turn, the basilar membrane varies by approximately 0.15

**Table 2.** Eigenvalues of the modes of variation for the model.

| Mode of variation | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| eigenvalues | 555.10 | 330.39 | 183.06 | 108.97 | 72.04 |
| % of total variance | 44.60 | 26.54 | 14.71 | 8.76 | 4.36 |

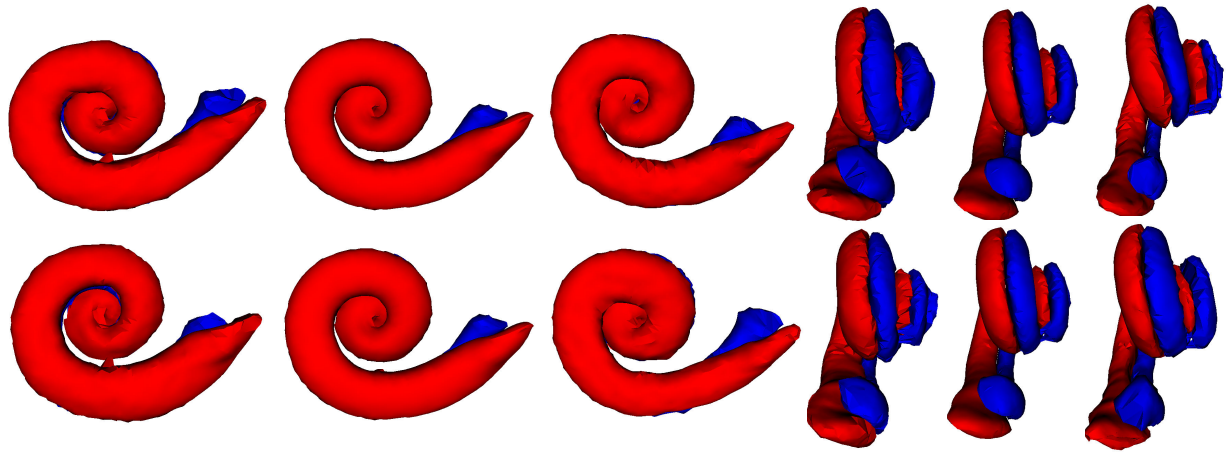

**Figure 15.** First (top row) and second (bottom row) modes of variation of the scala tympani (red) and vestibuli (blue) in the point distribution model. On the left are (from left to right) -2, 0, and +2 standard deviations from the mean in Posterior-to-Anterior view. The same modes are shown on the right in Medial-to-Lateral view.

mm. The first two modes of variation can be visually appreciated in Figure 15. It appears that the first mode of variation is capturing changes in the length of the basal turn, the second mode of variation is capturing changes in the width of the basal turn, and both are capturing changes in the width and angle of the second and third turns of the cochlea.

Quantitative comparisons between the manual segmentations of the scalae, transformed from μCT to the respective conventional CT, and the automatic segmentations are shown in Figure 16. The dice index for measuring volumetric overlap and bi-directional mean/max surface distances are computed between each pair of automatic and manual segmentations. The figure shows the overall distributions of these recorded values. The green bars, red bars, blue

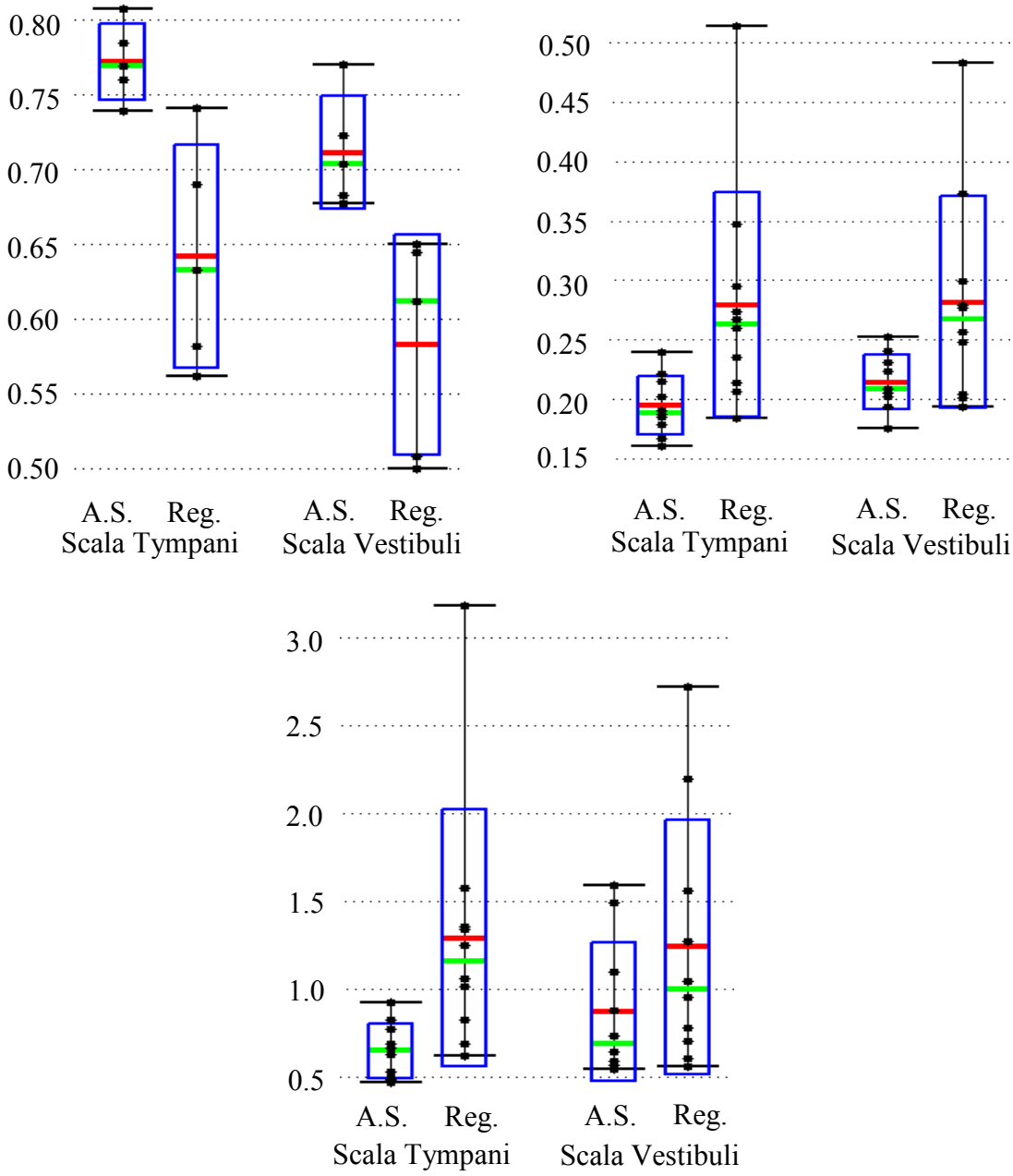


**Figure 16.** Quantitative segmentation results. Shown are the distributions of the dice indices (left), mean surface distance in mm (right), and max surface distance in mm (middle) for the results of the active shape model (A.S.) and registration alone (Reg.).

rectangles, black dots, and black I-bars denote the median, mean, one standard deviation from the mean, individual data points, and the overall range of the data set, respectively. Results for segmentation with the active shape model are shown side by side with results achieved using

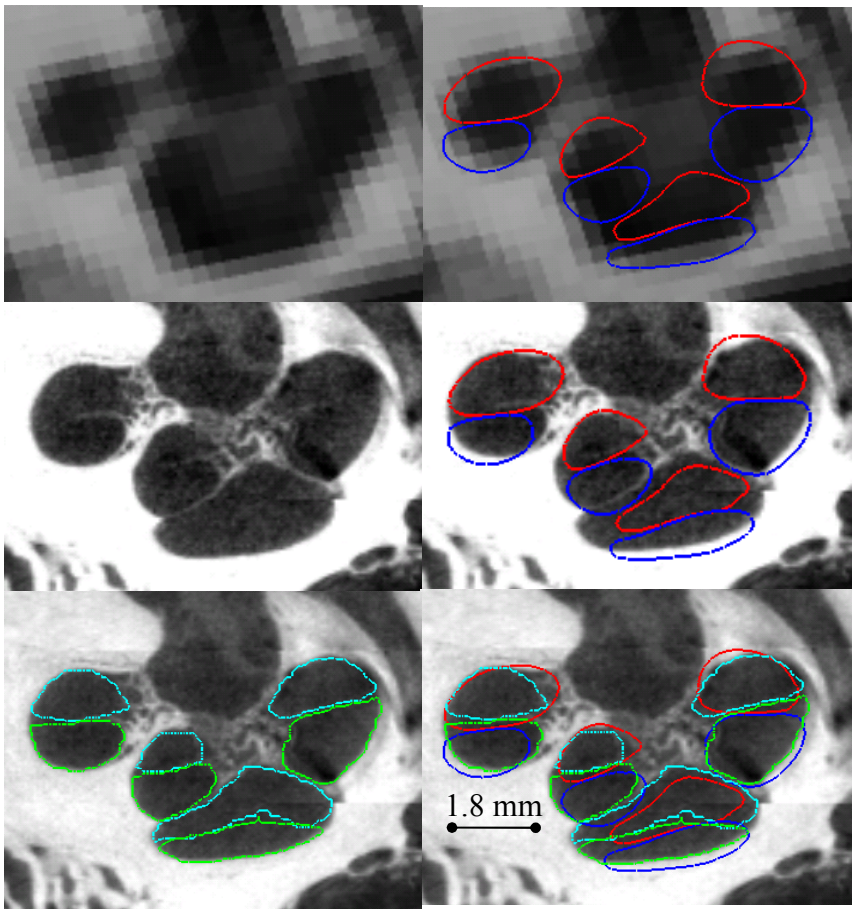


**Figure 17.** Contours of representative segmentation results. Automatic segmentation results for the scala tympani (red) and scala vestibuli (blue) are shown overlaid with the conventional CT (top row), and registered μCT (middle and bottom rows), and are compared to manually delineated contours of the scala tympani (light blue) and scala vestibuli (green).

non-rigid registration alone. As can be seen in the figure, the active shape model achieves mean dice indices of approximately 0.77 for the scala tympani and 0.72 for the scala vestibuli. For typical structures, a dice index of 0.8 is considered good [ 62 ], however, this measure can be unforgiving for very thin, complicated structures, such as the scalae, which are about 1 to 4 voxels wide along their length. For this case, we consider 0.7 to be a high dice index. It can also be seen from the figure that the active shape model effectively improves the volumetric overlap from the results achieved by registration alone in every experiment. A consistent decrease in mean surface errors from an average 0.28 mm with registration to 0.20 mm (< 1 voxel) with the

active shape model approach can also be seen. Maximum surface errors are on average decreased from 1.3 mm using registration to 0.8 mm by using the active shape model approach. In fact, with the active shape model, all surface errors for scala tympani segmentation are sub-millimetric.

Our results can be visually confirmed in Figure 17, Figure 18, and Figure 19. Shown in Figure 17 are contours of a representative automatic segmentation overlaid with the CT (the volume on which segmentation was performed) and the corresponding registered μCT. It can be seen from the figure that the contours achieved by automatic segmentation of the CT are in excellent agreement with contours manually delineated in the high resolution μCT. Localization errors that are apparent in the μCT are less than 2 voxels width in the CT. The automatic segmentations for all 5 experiments are shown in Figure 18 for the scala tympani and Figure 19 for the scala vestibuli, where the surfaces are color encoded with the distance to the manually segmented surface. It can be seen that the active shape model results in mean surface errors under 0.3 mm (< 1 voxel) for the majority of the structure with typical average maximum errors of about 0.8 mm (<3 voxels). The active shape model converges to a surface that looks anatomically correct, whereas registration-based segmentation alone converges to unnatural, irregular surfaces.

## Conclusions

In this chapter, an active shape model based approach for identifying structures that are partially visible in conventional imaging was presented. In this approach, high resolution images of cadaveric specimens are used to serve two functions. First, they provide information necessary to construct a shape model of the structure, permitting segmentation of the structure in conventional

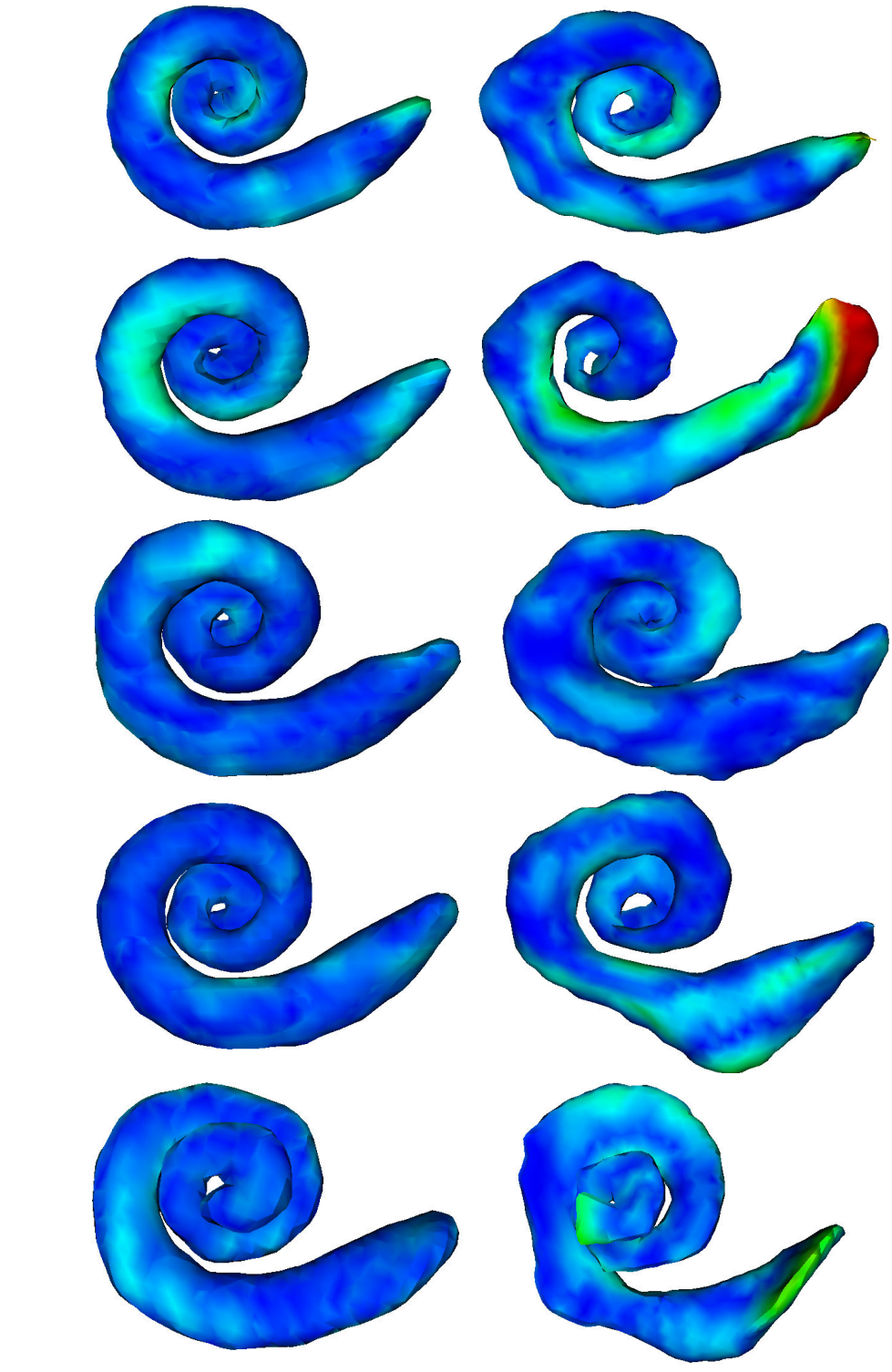


**Figure 18.** Segmentation results for the experiments 1-5 (Up to Down). (Left to Right) Active shape model and atlas-based segmentation of the scala tympani.

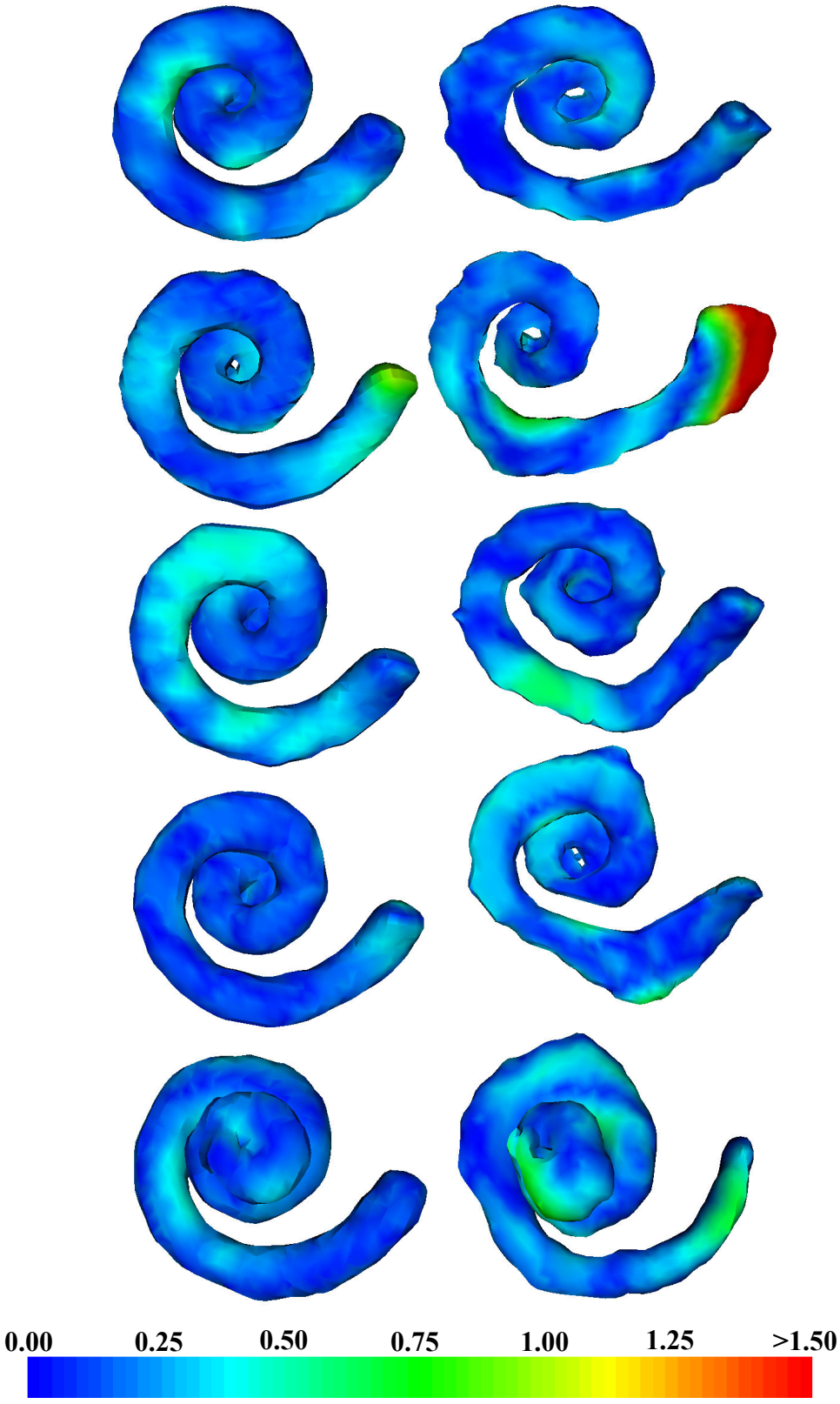


**Figure 19.** Segmentation results for the experiments 1-5 (Up to Down). (Left to Right) Active shape model and atlas-based segmentation of the scala vestibuli.

imaging by coupling information from the image and non-rigid image registration to drive an active shape model to an accurate solution. Second, the high resolution images are used to validate the results. This is performed by projecting manually segmented structures from high resolution to conventional images using image registration, and comparing those structures to automatic segmentations. We have applied this approach to identify intra-cochlear anatomy in CT and validate the results; however, the approach is generic, and could theoretically be applied to many other structures that are fully visible only in high resolution imaging.

This work has shown that it is possible to accurately identify the location of the scala tympani and scala vestibuli using conventional CT. This is possible because the position of the intra-cochlear membrane separating the scalae varies predictably with respect to the rest of the cochlea. The approach we present accurately locates the basilar membrane in conventional CT by attracting the exterior walls of the models of intra-cochlear anatomy towards the edges of the cochlea. This approach achieves dice indices of approximately 0.75, sub-millimetric maximum error distance for the scala tympani, and produces solutions that appear qualitatively reasonable. These results suggest that the segmentation approach can provide the surgeon with explicit knowledge of the location of surgical targets. In future work, experiments will be performed to confirm that providing this information to the surgeon improves implantation results.

# CHAPTER VII

# MINIMAL COST PATH ALGORITHM

## Introduction

Traditional efficient algorithms for finding a path of minimal cost between start and end nodes in a graph were first discovered with the work of Dijkstra [ 11 ]. Dijkstra's algorithm has been commonly used to find efficient data transfer paths through networks. It can also be used to efficiently find a curve *s* that minimizes the energy equation

$$\int_0^L C\big(s(l)\big)dl\,, \qquad (16)$$

where *C* is the energy function, by treating it as a problem of finding a minimal cost path between two end-nodes in a graph $G=\{V,E\}$. *V* is the set of nodes, and *E* is the set of edges connecting neighbor nodes, each of which is associated with a cost. The algorithm is used to find the set of edges connecting two end-nodes that are associated with the minimum total cost. This set of edges represents the curve *s* that minimizes Eqn. ( 16 ).

## Basic Implementation

Pseudocode of the basic implementation of Dijkstra's algorithm is shown in Figure 20. First, the set of candidate seed nodes are initialized and inserted into the priority queue. *m* is assigned to the seed node associated with the minimum cost. The search is conducted inside a while-loop. At each iteration, the neighbors of *m* that have not been visited are found. For each free neighbor, *m* is recorded as its parent node, and its cost is computed as the sum of the cost associated with *m* and the cost of the arc between itself and *m*. The neighbor is then inserted into the queue with

```
O := {seed nodes}
priority_queue.insert({seed nodes}, cost({seed nodes}))
m := priority_queue.pop_head_node()
while m ∉ {end nodes}
        {n}_i = neighbor_nodes(m)
        for ∀ i s.t. n_i ∉ O
                prev_node(n_i) = m;   cost(n_i) = arc_cost(m , n_i) + cost(m)
                priority_queue.insert(n_i , cost(n_i))
        endfor
        m := priority_queue.pop_head_node()
        O := O | m
endwhile
P := {m}
   while ∃ prev_node(m)
    m := prev_node(m)
    P := P | m
   endwhile
```

**Figure 20.** Dijkstra's algorithm for finding a minimal cost path in a graph.

priority equal to its cost. Next, *m* is assigned to be the node popped from the head of the queue, which is the node associated with the minimal cost. *m* is marked as having been visited, and if *m* is not an end node, the next iteration begins. This breadth-first searching scheme ensures that once an end node is popped from the head of the queue, the path of minimal cost, *P*, can be recorded by tracing the marked parent node pointers from the end node back to its associated seed.

The algorithm is guaranteed to produce the path of globally minimal accumulated cost if three conditions are satisfied: (1) A set of edges forming a path from a seed to an end-node must exist in the graph. (2) The costs of edges must be positive semi-definite. (3) The graph and costs must be static with respect to the search algorithm.

Optimized Implementation

The basic implementation of Dijkstra's algorithm can be enhanced with a bi-directional searching scheme, which is generally faster. The bi-directional search is used in this dissertation. Pseudocode for this algorithm is shown in Figure 21.

```
O := {seed nodes} ; E := {end nodes}
priority_queue1.insert({seed nodes}, cost({seed nodes}))
priority_queue2.insert({end nodes}, cost({end nodes}))
m := priority_queue1.pop_head_node()
q := priority_queue2.pop_head_node()
while (m ∉ E) && (q ∉ O)
      {n}_i = neighbor_nodes(m)
      for ∀ i s.t. n_i ∉ O
            prev_node(n_i) = m;   cost(n_i) = arc_cost(m , n_i) + cost(m)
            priority_queue1.insert(n_i , cost(n_i))
      endfor
      {n}_i = neighbor_nodes(q)
      for ∀ i s.t. n_i ∉ E
            prev_node(n_i) = q;   cost(n_i) = arc_cost(q , n_i) + cost(q)
            priority_queue2.insert(n_i , cost(n_i))
      endfor
      m := priority_queue1.pop_head_node()
      q := priority_queue2.pop_head_node()
      O := O | m ; E := E | q
endwhile
if (m∈ E) t = m ; u = E(m)
else  u= q ; t = O(q)
P = {t}
while ∃ prev_node(t)
   t := prev_node(t)
   P := P | t
endwhile
while ∃ prev_node(u)
   u:= prev_node(u)
   P := P | u
endwhile
```

**Figure 21.** Bi-directional Dijkstra's algorithm for finding a minimal cost path in a graph.

As can be seen in the figure, two priority queues are used to conduct the search in two directions from both the seed nodes and the end nodes. When the same node has been cycled from both the queues, the first while loop terminates. The nodes in the minimal cost path are then found by tracing the path backwards from the one overlapping node in both directions to the corresponding seed node and then the corresponding end node.

## Implementation details

The algorithm is implemented in C++. A few optimizations are put in place to benefit speed and memory usage. As will be seen in following chapters, some graphs will be constructed such that it is impossible to have all the nodes in the graph represented in memory. Instead, individual nodes are only realized once the graph search needs them. To efficiently conduct the search, the nodes need to be organized in two different ways. First, they need to be organized in the tree form of the priority queue, so that it is always efficient to access the node associated with the smallest cost. Second, another type of organization is needed to be able to search for particular nodes. As can be seen in Figure 21, the lines $m \notin E$, $q \notin O$, $n_i \notin O$, and $n_i \notin E$ all represent a search in which it is checked if a node exists in, what is usually, a very large group of nodes. Simply scanning through those groups of nodes repeatedly is not speed efficient. Instead, a hash table is implemented such that lists are hashed in the three voxel dimensions of the image being segmented, i.e., there is a list of unique nodes associated with each (x,y,z) voxel location, and those lists are accessible in $O(1)$ time.

Requiring two organizations of the nodes poses a memory issue. The nodes each require 31 bytes of storage, so having two copies of each node is wasteful. Instead, in this implementation, a dynamic list of arrays is used to store the nodes, referred to as the node

container. First, an array is initialized with space for $5x10^6$ nodes. Each time an array fills, another array of $5x10^6$ nodes is initialized and added to the tail of the list of arrays, creating space for new nodes to be stored. Now, the priority queue and the hash table, instead of actually containing nodes, simply contain pointers to nodes in the node container. This saves memory by storing only one copy of each node used. The implemented priority queue method uses an array type container. It is also initialized to an array of size $5x10^6$. This array is reinitialized every time it is filled and extended by another $5x10^6$ node pointers.

## Conclusions

The presented implementation of Dijkstra's algorithm is optimized for speed and memory efficiency for the applications presented in the following chapters.

# CHAPTER VIII

# AUTOMATIC SEGMENTATION OF THE FACIAL NERVE AND CHORDA TYMPANI

## Introduction

Cochlear implantation is a surgical procedure performed on individuals who experience profound to severe sensorineural hearing loss. In cochlear implant (CI) surgery, an electrode array is permanently implanted into the cochlea by threading the array into the basal turn, which is located at the round window shown in Figure 22. Current methods of performing the surgery require wide excavation of the mastoid region of the temporal bone. This excavation procedure is necessary to safely avoid damaging sensitive structures but requires surgical time of at least 2 hours. Recently, another approach has been proposed—percutaneous cochlear access—in which a single hole is drilled on a straight path from the skull surface to the cochlea [ 26 ].

To accomplish percutaneous cochlear implantation, the drilling trajectory must be planned in a pre-operative CT, the selection of which is a major difficulty with the approach. The preferred trajectory passes through the facial recess, a region approximately 1.0-3.5 mm in width bounded posteriorly by the facial nerve and anteriorly by the chorda tympani. Figure 22 shows a view along the preferred drilling trajectory, where the round window is located at the basal turn of the cochlea, and the facial recess is located at the depth of the facial nerve. The facial nerve, a tubular structure approximately 1.0-1.5 mm in diameter (~3 voxels), is a highly sensitive structure that controls all movement of the ipsilateral face. If damaged, the patient may experience temporary or permanent facial paralysis. The chorda is a tubular structure approximately 0.3-0.5 mm in diameter (~1 voxel). If the chorda is damaged, the patient may experience loss in the ability to taste. In previous work studying the safety of CI drilling

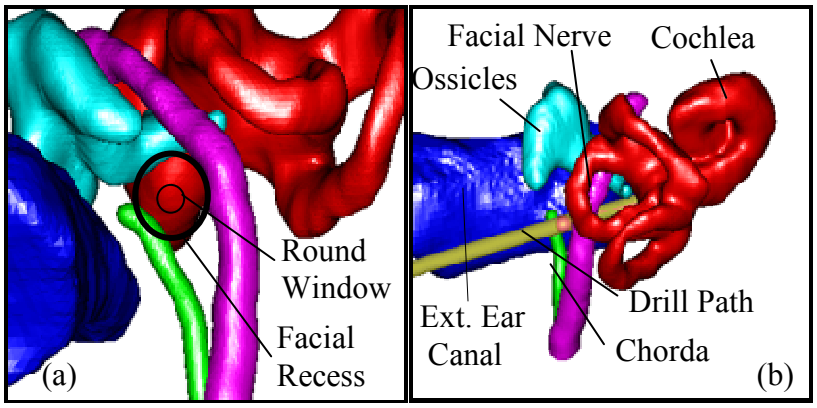


**Figure 22**. 3D rendering of the structures of the left ear. (a). Left-right, Posterior-Anterior view. (b) Posterior-Anterior view.

trajectories, in which error of the drill guide system was taken into account [ 37 ], it was determined that safe trajectories planned for 1 mm diameter drill bits generally need to lie at least 1 mm away from both the facial nerve and the chorda. Fitting a trajectory in the available space can thus be difficult, and any planning error can have serious consequences.

During the planning process, the physician selects the drilling trajectory in the patient CT by examining the position of the trajectory with respect to the sensitive structures in 2D CT slices. This is difficult, even for experienced surgeons, because the size of the facial nerve and chorda and their curved shape makes them difficult to follow from slice to slice. The planning process would be greatly facilitated if these structures could be visualized in 3D, which requires segmentation. Segmentation of these structures is also necessary to implement techniques for automatically planning safe drilling trajectories for CI surgery [ 37 ].

Atlas-based segmentation is a common technique, which relies on image registration to perform automatic segmentation of general structures in medical images. But there are some anatomical regions or structures for which registration is unable to produce sufficiently accurate results, and the facial nerve and chorda are representative of this kind of anatomy. These structures also present problems for generic tubular structure segmentation techniques (see

Chapter 2 for detailed background).

In this chapter, a novel method is proposed that combines an atlas-based approach with a minimum cost path finding algorithm. The atlas is used to create a spatially varying cost function, which includes geometric information. Once the cost function is evaluated, a 3D minimum cost path is computed. This process is used to extract the centerline of the facial nerve and of the chorda. The centerlines are then expanded into the full structures using a level-set algorithm with a spatially-varying speed function.

The remainder of this chapter is organized as follows. First, a summary of the general approach is presented. This is followed by a description of how the structure models are created and used to segment images. Then, results that were obtained are presented. Following this are the discussion and conclusions. The methods in this chapter were published in Noble et. al. [ 39 ].

## General approach

The method requires a model of the structures to be segmented. Because the target structures are tubular in shape, these models typically include two components. The first captures information related to the properties of the structure's centerline such as the intensity of a voxel located at a specific location along the centerline, or the orientation of the centerline along its length. The second captures information related to the structure's diameter. Once a model is built as discussed in the following sub-section, it is used to segment the structure in other image volumes. This is done as follows. First the image volume in which the model has been defined is registered to a new volume. Using the registration transformation, the model's centerline is projected onto the volume. This is illustrated in Figure 23 for a very simple example: the

segmentation of a dark structure on a bright background. The model centerline is shown in (a). In (b), the image to be segmented is shown. The dotted curve represents a structure in the image that is close in shape and intensity to the model, but it is affected by noise. The straight line that closes the dotted curve represents a structure in the image that we do not wish to segment, but since it is of approximately the correct intensity and connects the starting and ending points, it would represent an issue to most tubular structure segmentation algorithms. The result of the registration is displayed in (c). Note that the registration is inaccurate. Registration error is modeled here because, in general, the registration is imperfect. If the registration was perfect, it would permit the segmentation of the structure using an atlas-based method alone. Once the model is projected onto the image to be segmented, a cost matrix is created by computing, at each voxel, the difference between the voxel's characteristics and the characteristics of the closest model voxel. Finding correspondence between the image voxels and the model is shown in (d). For this simple problem, the resulting intensity-based cost image is shown in (e). The model is also used to provide *a priori* information about the orientation of the centerline as a type of shape cost. This information is useful for branching structures or structures that lack contrast with surrounding structures. This curve orientation information is represented in vector form across the image as shown in (f). Once the costs are defined, starting and ending points are chosen. This is also done via the registration transformation, i.e., a set of starting and ending points are defined in the atlas, and projected onto the other images. Next, a minimum cost path algorithm finds the optimal path between starting and ending points. For this simple example, the path finding algorithm would give us a structure centerline result as shown in (g). Finally, once the centerline is localized, the complete structure is segmented using a level-set approach. The speed function used in this step also makes use of a-priori information provided by the model.

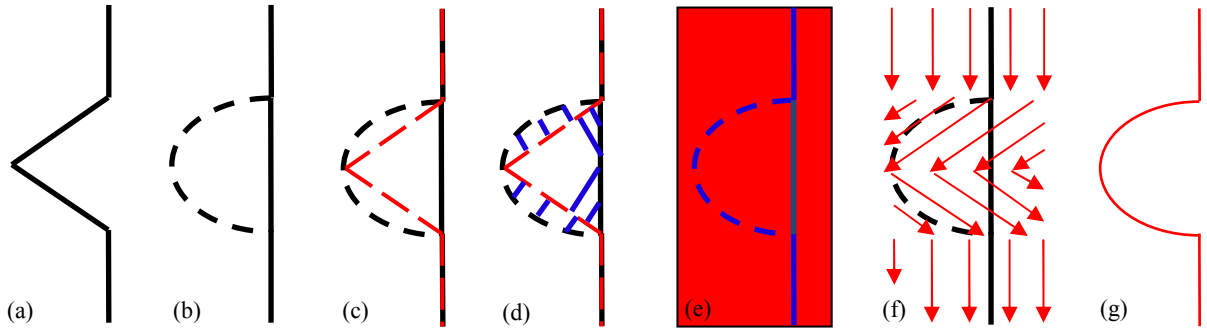


**Figure 23**. 2D example of the segmentation approach. (a) The model centerline. (b) A target image to be segmented. (c) The atlas centerline (red) registered to the target image. (d) Blue lines show correspondences found from points in the target image to the atlas centerline. (e) Intensity based cost image with blue=low and red=high cost. (f) Illustration of the vectors that drive the shape cost. (g) Centerline segmentation based on the intensity and shape cost terms.

## Model Construction

*Data.* Two sets of CT volumes acquired with IRB approval were used in this study. The first is the training set, and it consists of a total of 12 CT volumes, which include 15 ears unaltered by previous surgery. These images were all acquired on a Philips Mx8000 IDT 16 scanner. The second set, which is used as the testing set, consists of another 7 CT volumes, which include 10 ears unaltered by previous surgery. These images were acquired from several scanners, including a Philips Mx8000 IDT 16, a Siemens Sensation Cardiac 64, and a Philips Brilliance 64. The scans were acquired at 120-140 kVp and exposure times of 265-1000 mAs. Typical resolutions were 768 x 768 x 300 voxels with voxel size of approximately 0.3 x 0.3 x 0.4 $mm^3$.

*Outline.* The model used in this chapter consists of (1) the centerline of the left facial nerve segmented manually in one reference CT volume, (2) the centerline of the left chorda segmented manually in the same CT volume and (3) expected values for three characteristics at each voxel along these structures. A separate model for the right anatomy is not necessary since it can be approximated by reflecting the left model across the mid-sagittal plane without bias due to the

symmetry of the left and right anatomy of the human head. The three expected characteristics, which have been used are (a) the width of the structure $E_W$, (b) the intensity of the centerline $E_I$, and (c) the orientation of the centerline curve $E_O$. The expected values for each of the three structure characteristics have been computed from all 15 ears in the training set as follows.

*Alignment of Training Data:* First, one volume among the 12 training CT volumes was chosen as the reference volume, which will be referred to as $A$ for atlas. Next, all the other training images are globally and locally affinely registered to $A$ to correct for large differences in scale and orientation between the volumes (see Chapter 3).

*Extracting Image Data.* To extract feature values from each individual CT volume in the training set, the structures of interest are first segmented by hand (note that manual segmentation is used only for the creation of the model). This produces binary structure masks from which a centerline is extracted using a topology preserving voxel thinning algorithm [ 5 ]. At each voxel along the centerlines, the intensity, structure width, and structure orientation are estimated. To estimate the width, the intersection of the mask with a plane perpendicular to the centerline is computed. Structure width is then estimated by averaging the lengths of radii cast in that plane to the edge of the structure mask at rotation intervals of $\pi/8$ radians. The orientation of the centerline is estimated using a backward difference of the voxel locations with the exception of the first voxel, where a forward difference is used.

*Determining Correspondence across Model Images.* To establish a correspondence between points on the atlas centerline and points on the centerlines of each other volume, the atlas is

registered non-rigidly to each of the volumes (see Chapter 3). Once the registration step is completed, each of the points on the atlas centerline is projected onto each of the volumes in the training set using the corresponding non-rigid transformation. The point on a training volume's centerline that is the closest to the projected point is found and a correspondence is established between these two points. For each point on the atlas centerline, this results in 14 corresponding points on the training volumes' centerlines. Note that localizing the closest point on the training volumes' centerline is required because registration is not perfect. As a consequence, the atlas centerline points projected onto the training volumes do not always fall on their centerlines. The expected feature values for each point on the atlas centerline are then computed as the average of the feature values of 15 points (one point in the atlas and 14 corresponding points in the other volumes). The facial nerve and chorda models generated with this method are shown in Figure 24. In the plots, the solid lines are mean intensities along the model centerline curve for the facial nerve (a) and chorda (b). The dashed lines are intensities along the curve in the image chosen as the atlas. A clear change in intensities along the structures is visible.

## Structure segmentation

Once the models are built, they are used to segment new volumes. In Figure 25, a general overview of the method developed for this purpose is presented. In this figure, circles indicate an image (upper case roman letters), values associated with sets of voxels (lower case roman letters) or a transformation (greek letters). Rectangles are operations performed on the images. $T$ is the target image, i.e., the image that needs to be segmented, $A$ is the entire atlas image, and $A^{E}$ is an atlas sub-image that only contains the ear region. The target image is affinely registered first to the full atlas volume and then to the ear sub-volume. This results in image $T'$ and an affine transformation $\tau^{a}$. A non rigid registration is then computed between $T'$and $A^{E}$ to produce

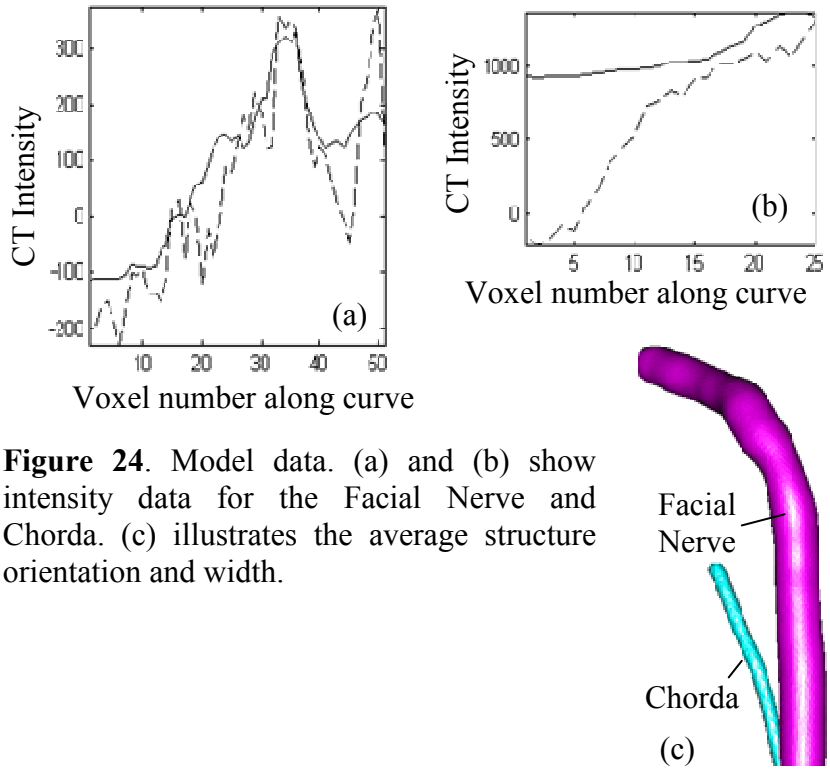


**Figure 24**. Model data. (a) and (b) show intensity data for the Facial Nerve and Chorda. (c) illustrates the average structure orientation and width.

transformation $\tau^n$. Note that segmentation of the structures of interest is performed on the images *T*'. This rescales and reorients the target image to match the atlas and permits using the a priori structure orientation information stored in the model. An alternative would be to register the atlas to the target but this would necessitate also reorienting the structures' orientation information stored in the model. The transformation $\tau^n$ is used to associate each voxel in *T*' with expected feature values. Using *T*', these expected feature values, and the cost function, costs are computed for all voxel connections in the range of *T*' and stored in a cost matrix. The starting and ending points for each structure are localized also using transformation $\tau^n$ and masks *M*. Once the cost matrix is computed and the structure extremities localized, the structures' centerlines are computed with a minimum cost path algorithm. Once the centerlines are found, the complete structures are extracted. A speed function is first generated using expected feature values in the model, which are projected onto the image *T*'. Next, a level set-based algorithm is utilized to

grow the centerlines into full structures. When the structures have been found in *T'*, they are projected back to *T,* which completes the process.

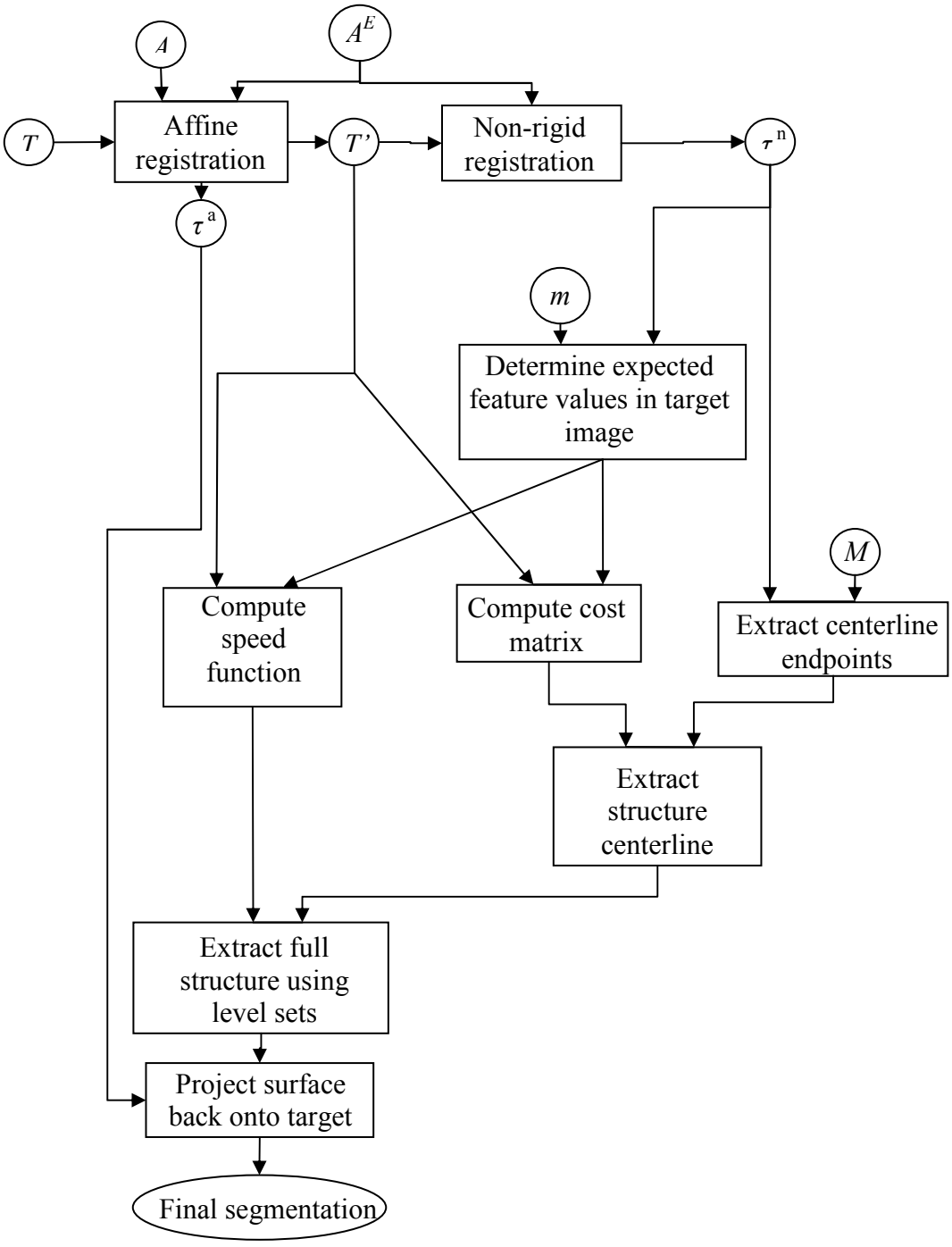


**Figure 25**. Flow-chart of the segmentation process.

*Cost matrix for centerline segmentation:* Minimum cost path algorithms require two pieces of information: (1) starting and ending points and (2) a cost associated with the transitions from each point to its neighbors. To compute costs based on a-priori information, i.e., intensity and orientation, stored for each point along the path in the atlas image, the following approach is used. First, the atlas centerline points are projected from the atlas to image *T*' using $\tau^n$. Next, for each voxel in *T*', the closest projected point is found. The expected feature values for each voxel in *T'* are then set to be the feature values of the closest projected point. The value of the cost matrix at each voxel *v* is then simply computed as the difference between the expected feature values and the feature values computed from the image at *v*. This approach permits to compensate for small registration errors.

The terms included in the cost matrix are based on (1) the expected intensity, (2) the expected structure orientation, and (3) local intensity minima in the image. These features were chosen after observation of structure characteristics in a number of CT images. Table 3 shows the exact analytical expression for each term used in the cost function and lists the value of the parameters used for the facial nerve and the chorda tympani, respectively. It also shows sensitivity values that will be explained later. The constant terms in these expressions are scaling factors, which have been introduced to map the range of the expressions before multiplication by parameter $\beta$ and exponentiation by parameter $\alpha$ approximately onto the [0,1] interval. The first term penalizes voxels whose intensity value departs from the expected intensity values. The second term favors making a transition in the direction predicted by the model. In this expression, $\vec{t}$ is the direction of the transition (i.e., up, left, right, etc.), and $\vec{E}_o(x)$ is the structure direction predicted by the model. A transition in the direction predicted by the model has an associated cost of zero while a transition in the opposite direction has an associated cost

**Table 3.** Terms used to compute the costs at every voxel in the image. For each term, the values and sensitivities are shown.

| Cost Function | Purpose | Facial Nerve | | Chorda | |
|---|---|---|---|---|---|
| | | α | β | α | β |
| $\beta\left(\left\|T'(x)-E_I(x)\right\|/2000\right)^{\alpha}$ | Penalize Deviation from Expected Intensity | 0.5 \| 30% | 3.0 \| 70% | 2.5 \| 20% | 10.0 \| 40% |
| $\beta\left(\left(1-\frac{\vec{t}\cdot E_O(x)}{\|\vec{t}\|\|E_O(x)\|}\right)/2\right)^{\alpha}$ | Penalize Deviation from Expected Curve Orientation | 4.0 \| 80% | 1.0 \| 80% | 4.0 \| 80% | 1.0 \| 80% |
| $\beta\left(1-\left(\frac{\#\{y \text{ in Nbhd}(x): T'(y)>T'(x)\}}{\#\{y \text{ in Nbhd}(x)\}}\right)^{\alpha}\right)$ | Penalize Deviation from Local Intensity Minima | 2.0 \| 70% | 1.0 \| 50% | 2.0 \| 70% | 1.0 \| 40% |

of one. The third term favors voxels that are local intensity minima. The expression $\#\{y \text{ in } Nhbd(x): T'(y) > T'(x)\}$ is the number of voxels in the 26 connected neighborhood of $x$ that have an intensity value that is larger than the intensity value of $x$. Note that this term is not model-dependent. The overall cost to move from one voxel to the other is the sum of terms (1) and (3) at the new voxel plus the transition cost in that direction. All values are pre-computed before finding the path of minimum cost. The method used to select the parameter values for each structure is discussed in a later section.

*Identifying Starting and Ending Points:* We rely on an atlas-based approach to localize the starting and ending points in image $T'$. First, masks for anatomical structures are manually segmented in the atlas. These masks are then projected onto the image to be segmented, and the extremity points are computed from these projected masks.

To localize the ending and starting points for the facial nerve, a facial nerve mask is delineated in the atlas such that it extends 2mm outside the facial nerve's segment of interest at each extremity. The starting and ending points for the facial nerve centerline are then chosen as the center of mass of the most inferior 5% of the projected facial nerve mask and the center of mass of the most anterior 5% of the projected facial nerve mask. In all the training images used in this study, this results in starting and ending points that are outside the segment of interest but within 2 mm of the nerve centerline as assessed by manual identification of the facial nerve in the CT scans. The paths thus start and end outside the segment of interest, and the localization of the starting and ending points is relatively inaccurate. However, the path finding algorithm corrects for this inaccuracy within the segment of interest. This makes our approach robust to unavoidable registration errors associated with the localization of the ending and starting points.

The chorda is a structure that branches laterally from the facial nerve at a location that exhibits high inter-patient variability. Thus the chorda mask was created such that the structure extends into the facial nerve 4 mm inferior to its branch from the facial nerve. The starting point for the chorda is then chosen as the most inferior 5% of the projected chorda mask. This point falls outside the segment of interest for the chorda. The path will thus start outside of the ROI but the branching point at which the chorda separates from the facial nerve is, with a high probability, located between the starting and ending points. The other end of the chorda in the segment of interest is placed on the undersurface of the tympanic membrane. But, because the exact location where it joins the membrane is ambiguous in CT images and also exhibits large inter-patient variability, the end point for the chorda segmentation is chosen to be any point within the projected tympanic membrane mask. This accounts for the variability in the final position by assuming that the true final position is the one that corresponds to the path with the

smallest total cost.

*Centerline Segmentation:* The starting and ending points defined using the approach described in the previous subsection and the cost matix are passed directly to the minimal cost path finding algorithm (see Chapter 7), which returns the final structure centerline segmentation.

*Full Structure Segmentation:* Because the chorda is extremely narrow, the chorda segmentation is completed simply by assigning a fixed radius of 0.25 mm along the length of its centerline. For the facial nerve, the centerline is used to initialize a full structure segmentation scheme. A standard geometric deformable model, level-set based method is used [ 47 ]. In this approach, the contour is the zero level set of an embedding function, $\phi$, which obeys the following evolution equation:

$$\phi_t + F|\nabla\phi| = 0, \qquad (17)$$

in which $F$ is the speed function, i.e., the function that specifies the speed at which the contour evolves along its normal direction. The speed function for the facial nerve has two components, the analytical expressions of which are shown in Table 4. In these expressions the terms $E_w(x)$ and $E_I(x)$ are the expected width and intensity value at $x$ predicted by the model. The first term is larger when the expected width is greater, and the second term is larger when the intensity in the image is similar to the predicted intensity. The total speed defined at each voxel is the product of the terms in the table. Traditionally, Eqn. ( 17 ) is solved iteratively until the front stabilizes and stops. This is not possible here because the contrast between the facial nerve and its surrounding structures is weak. To address this issue we solve Eqn. ( 17 ) a fixed number of times (here three times) and we choose the values of $\alpha$ and $\beta$ such that the propagating front

**Table 4**. Terms used to compute the speed function at every voxel in the image. For each term, the values and sensitivities are shown.

| Speed Function | Purpose | Facial Nerve | |
|---|---|---|---|
| | | α | β |
| $e^{\left(\beta E_W(x)^{\alpha}\right)}$ | Slow propagation where structure is thinner | 1.0 \| 80% | 0.7 \| 80% |
| $e^{\left(-\beta\left(\left\|T'(x)-E_I(x)\right\|/2000\right)^{\alpha}\right)}$ | Slow propagration where intensities deviate from expectation | 0.5 \| 80% | 1.3 \| 60% |

reaches the vessel boundary in three iterations. The method used to select values for $\alpha$ and $\beta$ is discussed in a later section.

## Creation of the Gold Standard and Evaluation Method

Gold standards for both the training and testing sets were obtained by manual delineation of these structures. The contours were drawn by a student rater, then corrected by an experienced physician. Contours were drawn only over regions of the structures that were judged to be of clinical significance for planning CI surgery, i.e., only along the length of the structure in the region that is directly in danger of being damaged by the drilling procedure. As a consequence, the most extreme inferior and posterior processes of the facial nerve, and the inferior and superior processes of the chorda are typically not included in the ROI. To quantitatively compare automatic and manual segmentations, surface voxels from structures are first identified. Because portions of the structure models extend beyond the ROI, some voxels are present in the automatic segmentation that fall outside this region. To avoid biasing the quantitative analysis, the portions of the automatic segmentations that lie beyond the manually chosen extrema are

eliminated prior to calculations. Thus, the extreme processes of the automatic tubular structures are cropped to match the manual segmentation.

Two distance measures to compare similarity between automatic and manual segmentations. For each point on the automatic surface, the distance to the closest point on the manual surface is computed, which will be called the automatic-manual (AM) error. Similarly, for each point on the manual surface, the distance to the closest point on the automatic surface is computed, which will be called the manual-automatic (MA) error. These two types of error are discussed in Chapter 4. Once the parameters are chosen, as discussed in the following section, the algorithm was applied, without modification, to the 10 ears included in the testing set, and these errors were measured.

## Parameter Selection

Using the 15 ears in the training set, parameter values were chosen, and the sensitivity of the algorithm to these values was studied. Parameter selection was done manually and heuristically. For each parameter, an acceptable value was chosen first by visually observing the behavior of the algorithm as a function of parameter values. Once an acceptable value was found, it was modified in the direction that reduced the maximum AM or MA error over all volumes in the training set until a value was found at which the error clearly increased. The final value was chosen away from this point in the generally flat error region, which preceded it. Once values were determined for each parameter, the sensitivity of the resulting parameters was analyzed. This was done by sequentially modifying the parameter values in 10% increments around each selected value until the maximum AM or MA error over all volumes in the training set increased

above 1mm. The deviation (in percent) at which this happens is reported in Table 3 and Table 4 for each parameter.

## Implementation and timing

The presented methods were all coded in C++. Segmentation of the facial nerve and chorda tympani requires approximately 5 minutes on an Intel quad-core 2.4 GHz Xeon processor, Windows Server 2003 based machine.

## Results

Mean, median, and maximum AM and MA errors for the testing set are shown in Table 5. Segmentation results for 3 cases are shown in Figure 26. This figure includes the case exhibiting the worst facial nerve error (4-R – 4$^{th}$ patient, right ear), the case exhibiting the worst chorda error (1-L), and lastly a case with very little all around error (3-L). In this figure, the first two columns visualize the facial nerve (first column) and the chorda tympani (second column) in gray level images. To generate these figures, the 3D medial axis of the automatically segmented structures was mapped onto a straight line using thin-plate splines. The images show the intensity values in a plane passing through the straightened center lines. This permits to show the entire length of the structures in a single 2D image. The solid green and purple dashed lines are the intersections of the automatic and manual surfaces with the image plane, respectively. Small differences between the manual and automatic centerlines, as is the case for the second chorda, will cause the manual surface to only partially intersect the image plane. The thin dashed cyan lines are the contours of the projection of the manual surface on the image plane. The third column shows the silhouette of the facial nerve and chorda tympani on a plane perpendicular to a

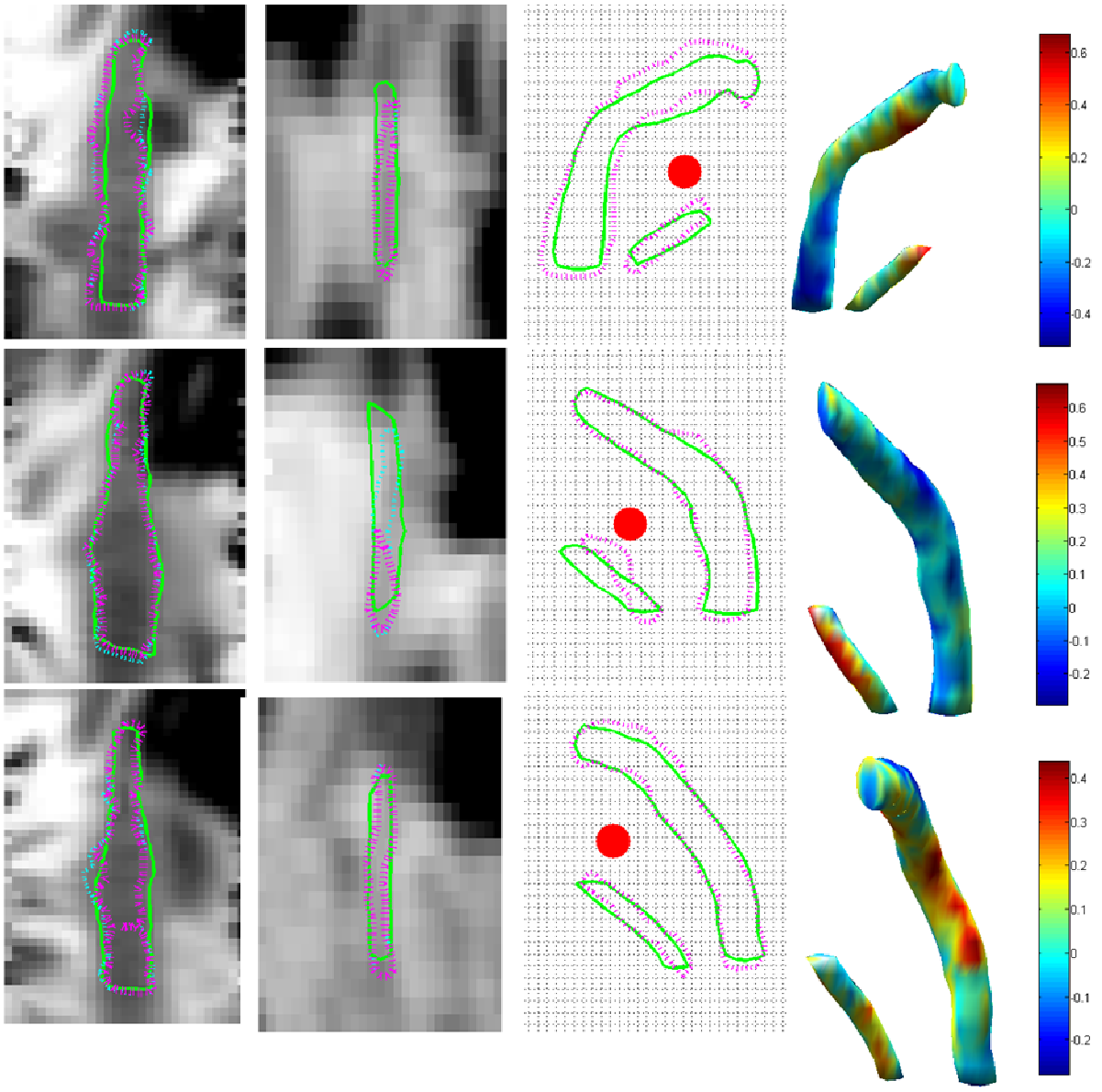

**Figure 26**. Segmentation results for CT volumes 4-R (top row), 1-L (middle row), and 3-L (bottom row).

typical trajectory, i.e., the contour of the structures' projection on this plane. The purple and green contours correspond to the manually and automatically segmented structures, respectively. The red dots in these images are the cross sections of the drilling trajectory. The grid lines are spaced according to the size of one voxel (~0.35 x 0.4 $mm^2$). The fourth column shows the

**Table 5.** Facial nerve and chorda segmentation results. All measurements are in millimeters and were performed on a voxel by voxel basis. Shown are the mean, maximum and median distances of the automatically generated structure voxels to the manually (AM) segmented voxels and vice versa (MA). The last column contains mean, max, and median values for voxels from all ears.

| | **Volume** | 1 | 2 | 3 | 4 | | 5 | | 6 | | 7 | **Overall** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Ear** | L | R | L | L | R | L | R | L | R | L | **Statistics** |
| **Facial Nerve** | Mean Dist. AM | 0.103 | 0.093 | 0.140 | 0.093 | 0.146 | 0.175 | 0.161 | 0.126 | 0.159 | 0.122 | **0.132** |
| | Max Dist. AM | 0.487 | 0.438 | 0.492 | 0.521 | 0.800 | 0.524 | 0.557 | 0.618 | 0.651 | 0.400 | **0.800** |
| | Median Dist. AM | 0.000 | 0.000 | 0.246 | 0.000 | 0.000 | 0.234 | 0.234 | 0.206 | 0.206 | 0.000 | **0.103** |
| | Mean Dist. MA | 0.085 | 0.128 | 0.121 | 0.104 | 0.215 | 0.138 | 0.124 | 0.147 | 0.149 | 0.195 | **0.141** |
| | Max Dist. MA | 0.468 | 0.400 | 0.470 | 0.477 | 0.838 | 0.469 | 0.469 | 0.509 | 0.509 | 0.738 | **0.838** |
| | Median Dist. MA | 0.000 | 0.000 | 0.000 | 0.000 | 0.260 | 0.234 | 0.000 | 0.206 | 0.206 | 0.000 | **0.000** |
| **Chorda** | Mean Dist. AM | 0.349 | 0.207 | 0.156 | 0.179 | 0.212 | 0.082 | 0.095 | 0.153 | 0.120 | 0.055 | **0.161** |
| | Max Dist. AM | 0.689 | 0.438 | 0.348 | 0.582 | 0.582 | 0.381 | 0.234 | 0.634 | 0.364 | 0.400 | **0.689** |
| | Median Dist. AM | 0.344 | 0.310 | 0.246 | 0.260 | 0.260 | 0.000 | 0.000 | 0.206 | 0.206 | 0.000 | **0.226** |
| | Mean Dist. MA | 0.224 | 0.199 | 0.043 | 0.114 | 0.026 | 0.079 | 0.054 | 0.111 | 0.100 | 0.053 | **0.100** |
| | Max Dist. MA | 0.630 | 0.438 | 0.246 | 0.544 | 0.260 | 0.469 | 0.332 | 0.600 | 0.600 | 0.369 | **0.630** |
| | Median Dist. MA | 0.122 | 0.310 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | **0.000** |

surface of the automatically segmented structures. The color map encodes the distance between this surface and the closest point on the manually segmented surface in millimeters. In this figure, the distance is negative when the automatic surface is inside the manual surface. It is positive when it is outside.

## Discussion and Conclusions

The method requires selecting a number of parameters (10 for the facial nerve and 6 for the chorda). Although, at the present time, this requires initial manual adjustment, this study shows that the method is not very sensitive to the selection of these parameters. Any of the parameters can be changed by 20% before the segmentation error reaches 1mm. Several of these parameters can be changed by 80% or more. The results also show that the method is robust. The algorithm has been trained on 12 CT volumes (15 ears) acquired at our institution on the same scanner. It

was tested without modification on 7 additional CT volumes (10 ears) acquired both at multiple institutions on different scanners. Results also show that the proposed method is accurate. The mean AM and MA errors are on the order of 0.1 mm for the facial nerve and slightly higher for the chorda. Maximum errors are on the order of 0.8 mm for the facial nerve and 0.7 mm for the chorda. Note that these errors are differences between manual and automatic segmentation. Figure 26 shows that the largest errors in the chorda segmentation occur at the superior end of the structure, where it exits the tympanic bone. Precise and consistent localization of this point is very difficult even manually. Manual contouring of the facial nerve is also difficult, leading to contour discontinuities between slices. This is especially acute when the nerve is surrounded by structures with similar intensities. This is apparent in Figure 26 where the automatic contours are generally smoother than the manual ones. It is thus possible that the automatic segmentation is, in fact, more accurate than the manual one. The figure also shows that the segmentation errors are generally not larger in areas that matter most (i.e, along the drilling trajectory) than they are in other regions of the structure, thus suggesting that the method can be used to generate reliable surfaces to be used for surgical planning.

This chapter presents the first method for the automatic segmentation of two critical structures for cochlear implant surgery: the facial nerve and the chorda tympani. Because of the shape and size of these structures, and because of the lack of contrast between these and surrounding structures, accurate segmentation is challenging. Purely atlas-based methods are ineffectual for this problem. Computing transformations that are elastic enough to compensate for anatomic variations leads to transformations that are not constrained enough to deform the structures of interest in a physically plausible way. More constrained transformations lead to results that are not accurate enough. An optimal path finding algorithm based purely on intensity

was also unsuccessful. Even if starting and ending points are chosen manually, changes in intensity along the structures, lack of intensity gradients between the structures of interest and background structures, and incorrect paths of similar intensity from starting to ending points led to inaccurate segmentations.

The solution lies in providing the optimal path finding algorithm with expected feature values. These can be provided with an atlas-based approach because a perfect registration is not required. An approximate registration is sufficient to provide the algorithm with the expected feature values for centerline voxels in a region. The optimal path algorithm then chooses the path that contains voxels associated with feature values that are similar to the expected features.

Another novel element of the approach is a spatially varying speed function also defined using a model. This speed function is then used to drive a geometric deformable model and segment the complete facial nerve. Traditional geometric deformable model approaches involve the definition of a speed function that stops the front at the edge of the structures of interest. A common problem is the issue of leakage when edges are poorly defined. The solution we have developed is to design a speed function that evolves the front from the centerline to the structure edges in a fixed number of steps. The results show that this approach is robust and accurate.

# CHAPTER IX

## AUTOMATIC SEGMENTATION OF THE OPTIC NERVES AND CHIASM

### Introduction

Segmentation is an important step in the radiation therapy process. Modern IMRT (Inverse Modulated Radiation Therapy) treatment techniques require both the delineation of structures to be irradiated and of structures to be spared, so that an optimal radiation delivery plan can be computed unique to each patient. This segmentation process is time consuming and error-prone. The optic nerves and optic chiasm are sensitive structures in the head that are especially difficult to segment. Figure 27 shows these structures in MR and CT images. The arrows in these images show regions where contrast is lacking in either the CT images (orange, dashed lines), the MR images (yellow, thin arrowhead), or both (blue, solid line). A number of methods have been proposed in recent years to segment these structures automatically. Bekes et al. [ 4 ] have proposed a geometric model-based method for semi-automatic segmentation of the eye balls, lenses, optic nerves and optic chiasm in CT images. They report quantitative sensitivity and specificity results from STAPLE [ 56 ] of approximately 77% and 95% for the optic nerves and approximately 65% and 94% for the chiasm. Qualitatively, they report a general lack of consistency with the results they obtain for the nerves and chiasm. Atlas-based methods are commonly used for image segmentation (see for instance [ 45 ] [ 14 ] [ 9 ] [ 44 ] [ 12 ]). Several atlas-based approaches have also been proposed to segment the optic nerves and chiasm for radiation therapy [ 8 ] [ 17 ] [ 21 ]. In this type of approach, one reference volume, commonly referred to as the atlas, is first segmented using interactive methods. This volume is then

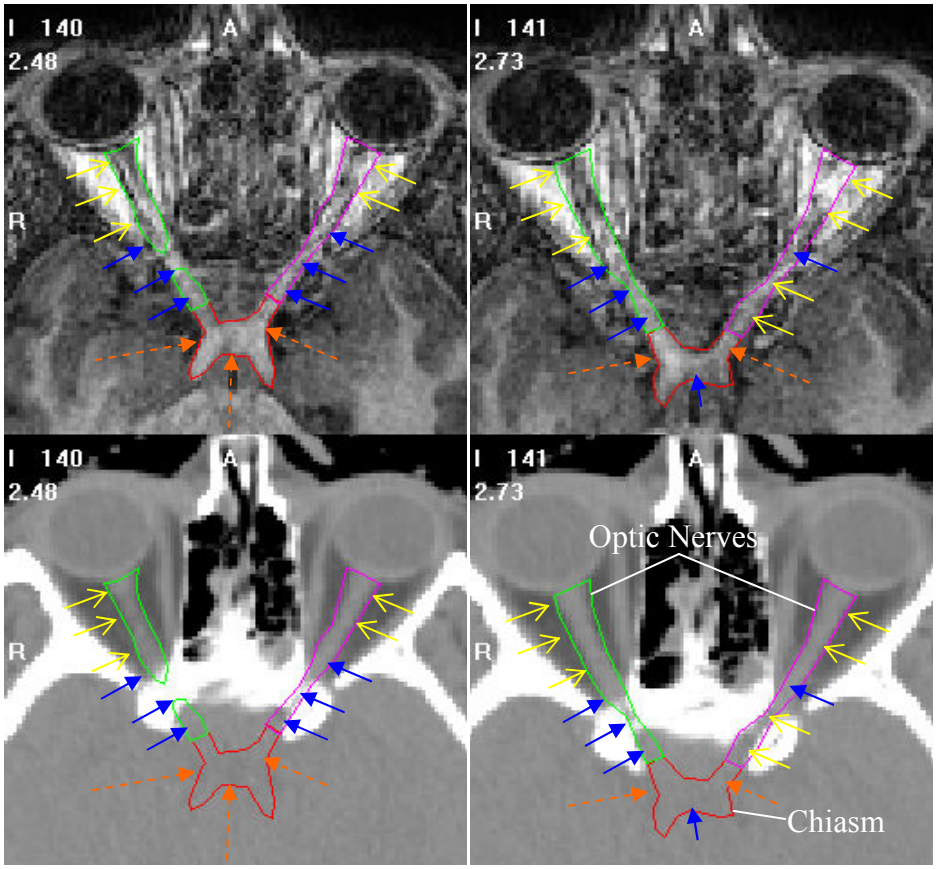


**Figure 27**. Transverse slices containing the optic nerves (anterior contours) and chiasm (posterior contour) in MR (top row) and CT (bottom row). The arrows indicate areas where the structures are not well contrasted in MR (yellow, thin arrowhead), CT (orange, dashed line), or both (blue, solid line).

registered to other volumes to be segmented. Structure labels in the reference volume are subsequently projected onto the other volumes using the computed transformation. Although this approach has been reported to produce good results for most head structures, the optic nerves and chiasm were found to remain challenging with dice indices ranging from 0.39 to 0.78 for the optic nerves and 0.41 to 0.74 for the chiasm.

In this chapter, a new and different method to attack the problem is presented. First, because it is more and more common in clinical practice to acquire both MR and CT volumes for treating patients with brain tumors, both modalities are used. Second, as opposed to previous

methods, which localize the optic nerves and chiasm independently, the optic nerve and the contra-lateral optic tract are considered as a single structure. Third, the optic chiasm is not explicitly segmented. Instead, the chiasm is found as the intersection of the two optic nerve/optic tract structures, which is its true anatomic definition. This is illustrated in Figure 28. The first structure of interest (SOI) is made of the left optic nerve and right optic tract (left SOI), the second structure is made of the right optic nerve and left optic tract (right SOI). The intersection between these two structures is the optic chiasm. This approach has two main advantages. First, it is not necessary to find the optic chiasm directly, which is the most difficult structure to segment. Second, the segmentation of the optic nerves and tracts can be achieved by segmenting two independent tubular structures.

A review of tubular structure segmentation methods is presented in Chapter 2. Typical types of approaches are unlikely to work for the optic nerves and chiasm due to the variations in intensity along the structure's length and the lack of contrast with the background. In this chapter, the method presented in the previous chapter to segment the facial nerve and chorda tympani in CT images is adapted to segment the optic nerves and chiasm. The facial nerve and chorda are other examples of tubular structures that are difficult to segment because of their size (diameter on the order of 2-3 voxels for the facial nerve and smaller for the chorda) and lack of contrast with surrounding structures. This approach combines an optimal path algorithm with a-priori information provided by the registration of a model to the image to be segmented.

In the next section, the data sets used are discussed. This is followed by a presentation of the general approach, evaluation methods, and parameter selection. Finally results and conclusions are presented. The methods presented in this chapter were first presented at the SPIE conference on medical imaging 2009 (Noble and Dawant [ 35 ]).

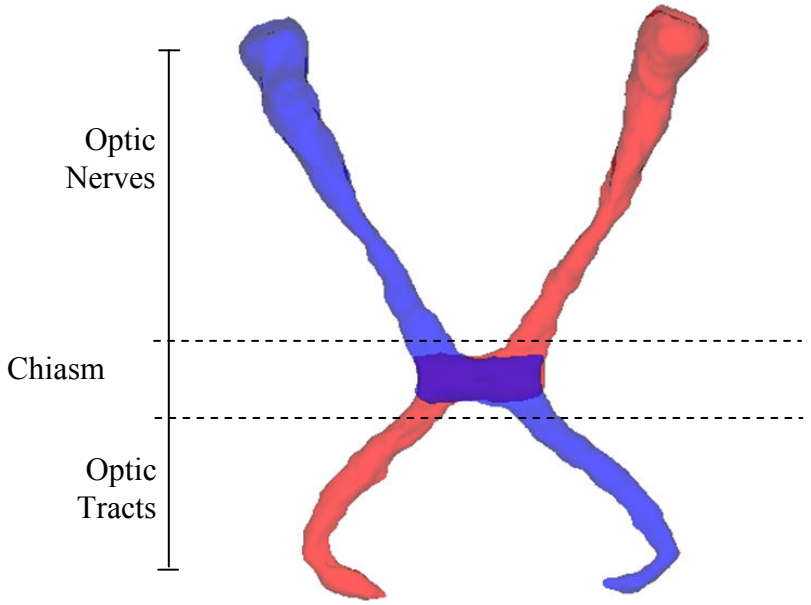


**Figure 28**. The optic nerves, chiasm, and tracts represented as two intersecting tubular structures.

## Data

A model training set of four, a parameter training set of ten, and a testing set of ten CT/MR pairs were acquired with IRB approval and used in this study. The CTs were acquired on four different scanners at 120-130 kVp and exposure time of 350-1750 mAs with voxel sizes ranging from 0.5 $mm^3$ isotropic to 1.0 x 1.0 x 3.0 $mm^3$. Six contrast enhanced T1 weighted MRs were acquired on 1.5 T scanners with voxel size of 1.0 x 1.0 x 1.2 $mm^3$. The remaining eighteen MRs, ten of which compose the testing set, were non-contrast T1 weighted acquired on 1.5-3 T scanners with voxel size 1.0 $mm^3$ isotropic. Each CT/MR pair was rigidly registered using standard techniques.

## General approach

The basic approach used for identifying the left and right SOIs is identical to those presented in the last chapter. A model of the SOI is built using training images. At segmentation time, an atlas

is used to align the model to the target image, and a minimal cost path algorithm is used to identify the centerline of the SOI with a cost function that is dependent on priors from the model.

Because the left and right structures to be segmented are mirror images of each other, a model needs to be built only for one of them. The model is created for the left SOI, and a right SOI can be added to the model by reflecting the image across the mid-sagittal plane. The model consists of three quantities for each point along the centerline: (1) the orientation of the centerline curve $E_O$, (2) the width of the structure $E_W$ and (3) an intensity vector $\vec{E}_I$ that captures the intensity of voxels surrounding each centerline point. Intensity models are built for the MR and CT images separately but the MR model is used on some sections of the structure while the CT model is used on the others as discussed later.

The models are built as follows. All corresponding MR and CT image volumes are registered to each other. The structures of interest are segmented manually in each of these MR/CT pairs and the centerline of these structures is extracted using a thinning method. One MR/CT set is selected as the reference set, which will be referred to as the atlas; the other volumes will be referred to as training volumes. Correspondence among training volume centerlines is found using the procedure presented in the previous chapter.

Once correspondence is established $E_O$ and $E_W$ are computed as the average curve orientation and average structure width over the corresponding centerline points, where curve orientation is estimated using central differences of the curve points. The intensity vector is computed with the method illustrated in Figure 29 (this figure shows a 2D case but the procedure can readily be expanded to 3D). Panel (a) in the figure shows the pixel of interest, marked with a blue dot, and the surrounding patch of intensities. In panel (b), a red circle is drawn around the point of interest, defining a neighborhood. Here we use a circle of radius $R_O$ (in 3D we define a

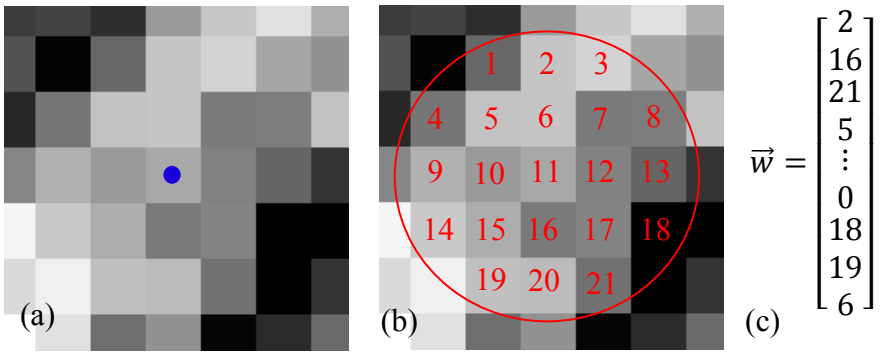


**Figure 29**. Construction of an intensity feature vector. (a) shows the pixel of interest surrounding patch of intensities. The sequence of neighboring pixels is shown in (b). In (c), a vector of the remapped intensities of the neighborhood is shown.

sphere). Then, every point within that neighborhood is visited following a pre-determined sequence (as shown in (b)), and an intensity vector that contains the pixel (voxel) intensities is created. This process is applied to the atlas centerline point as well as its set of corresponding training centerline points, resulting in $N$ intensity vectors for each atlas centerline point, which are averaged. One common problem with MR images is the variability in intensity characteristics between volumes. To address this issue we use the following procedure. The values of the average intensity vector are remapped into sequential integer values with the lowest value remapped to zero, the second to 1, etc., to produce the vector $\vec{w}$. An example of this intensity remapping is shown in panel (c). This representation retains the information about the relative intensities of the voxels surrounding the point of interest while discarding exact intensity information. It will thus not be impacted by intensity differences between volumes as long as these intensities can be related by a monotonic function. The feature vector $\hat{w}$ is then obtained by normalizing $\vec{w}$ to length 1. The similarity between two points can be compared simply by computing the Euclidean distance between their respective feature vectors, where a smaller distance indicates greater similarity. The parameter $R_O$ controls the locality of the measurement.

Once the models are built, new images can be segmented. The segmentation procedures for each SOI are identical to those presented in the previous chapter. The atlas is first registered to the target image. Using the registration transformation, the model is projected onto the target image space. Using information from the model, the structure centerline is localized by using a graph search algorithm to minimize the cost function based on the expected intensity feature vectors and the expected curve orientation. For a given location, the feature vector is either based on MR or CT data. Mimicking a typical surgeon, the optic nerve is localized using CT data while the chiasm is localized using MR data. This approach is advantageous due to the structure contrast observed in the two modalities for these structures (see Figure 27). To automatically decide which image should be used, a point at the optic nerve/chiasm junction along the atlas centerline was chosen so that all image voxels corresponding to atlas points towards the optic nerve from this point use the CT features while voxels that lie towards the chiasm use MR features. The cost for feature vectors is defined by

$$\beta\left(\frac{\left\|\vec{E}_I - \hat{w}\right\|}{\sqrt{2}}\right)^{\gamma}, \qquad (18)$$

where $\hat{w}$ is the feature vector of the point of interest and $\beta$ and $\gamma$ are weighting and sensitivity parameters. The maximum distance between any two normalized ranking vectors composed of all positive values can be shown to be $\sqrt{2}$. Thus, the constant factor of $\sqrt{2}$ in Eqn. ( 18 ) is used to scale the range of the function to the interval [0,1] prior to scaling by $\beta$ and exponentiation by $\gamma$. The curve orientation cost is computed by

$$\left(\left(1 - \frac{\vec{t} \cdot \vec{E}_O}{\left\|\vec{t}\right\|\left\|\vec{E}_O\right\|}\right)/2\right)^{\delta}, \qquad (19)$$

where $\vec{t}$ is the direction of the voxel transition (i.e., up, left, right, etc.). This term favors

transitions in the direction predicted by the model. The inner fraction in Eq. ( 19 ) corresponds to the cosine of the angle between $\vec{t}$ and $\vec{E}_O$, which ranges from -1 to 1. Thus, the constant factor of 2 in the equation scales the function into the interval [0,1] prior to exponentiation by $\delta$. The total cost for transitioning to any particular voxel in the image is the sum of the intensity and orientation terms in Eqns. ( 18 ) and ( 19 ).

Finally, endpoints are needed. Instead of using a single starting and ending point in our optimization, we include many and simply use the starting point and ending point that result in the optimal path. Specifically, we include all voxels on a disc centered at the endpoint given by non-rigid registration, perpendicular to the curve, and with the radius equal to the maximum distance from the endpoint to its corresponding points in the training images. This is the radius necessary to ensure that registration error in the endpoint regions can be handled by the segmentation algorithm. These starting and ending points and the cost function are used by the graph search algorithm to identify the centerline of the SOI by treating it as a problem of finding a minimal cost path between two end-nodes in a graph $G=\{V,E\}$. $V$ is the set of nodes, and $E$ is the set of edges connecting neighbor nodes, each of which is associated with a cost. The algorithm is used to find the set of edges connecting two end-nodes that are associated with the minimum total cost. This set of edges represents the curve that minimizes the cost function. In this application, $V$ is composed of the set of voxels, $E$ connects neighboring voxels in a 26-connected neighborhood, and the cost of each edge is the sum of Eqns. ( 18 ) and ( 19 ).

The segmented structure centerline is used to initialize a standard level-set based geometric deformable model. The speed function designed for this application is computed as

$$E_W\left(1-\frac{\vec{E}_I-\hat{w}}{\sqrt{2}}\right), \qquad (20)$$

where $E_W$, $\vec{E}_I$, and $\hat{w}$ are the expected width, expected feature vector, and actual feature vector at the point of interest. The expected width term ensures this function is larger where the structure is expected to be wider. The intensity feature vector term increases function values where the intensity features are more similar to the expected centerline intensity features. Thus, function values tend to be relatively large at the structure centerline and slowly decrease inside the structure away from the centerline. The intensity feature terms ensure that function values drop more drastically at the edge of the structure. To initialize the level-set algorithm, an initial level-set distance map is computed, treating the voxels corresponding to the extracted structure centerline as interior points and all other voxels as exterior points. Traditionally with level-set approaches, the evolution equation is solved iteratively until the level-set front stabilizes and stops. As for the structures of the middle ear treated in the preceding chapter, this is not possible here because structure contrast is very weak. To address this issue, the speed function is a strictly expanding function, and Eqn. ( 17 ) is solved only once over time $\Delta t$, where $\Delta t$ is used as a parameter to control the total expansion of the zero-level set.

After the left and right SOIs are localized using this approach, the final segmentation of the optic pathways is achieved by computing the union of the two.

## Evaluation

Gold standards for the training and testing sets were obtained from manual delineations of the structures of interest. For the training sets, the contours were first drawn by a student rater then corrected by two experienced radiation oncologists concurrently. Two independent sets of delineations were obtained for the ten MR/CT pairs in the testing set. These delineations were created first by two separate student raters. Then, one set was corrected by an experienced

radiation oncologist, and the other was corrected by an experienced radiologist. To evaluate accuracy, the automatic segmentations are compared to these manual delineations using the dice similarity coefficient. Mean and maximum symmetric surface errors are also reported. The point of transition of the optic nerves into the chiasm is somewhat subjective. In fact, the optic nerves and chiasm are simply particular regions of the single optic pathway structure, and they are treated with equal priority in radiotherapy planning. Because of this, and because there are no visual indications of this transition between nerves and chiasm in MR/CT, the manual delineations and the automatic segmentations represent the optic nerves and chiasm as the single optic pathway structure. To evaluate accuracy separately for the optic nerves and chiasm, we use an approach similar to Gensheimer et. al. [ 17 ]. After segmentation, the coronal plane that appropriately separates the optic nerves and chiasm was identified manually for each volume tested. Then, for each volume, error was measured for the optic nerves by comparing the portions of the automatic and manual segmentations that lie anterior to this plane. Similarly for the chiasm, error is measured by comparing the portions of the segmentations posterior to this plane.

## Parameter Selection

For each parameter, an acceptable value was first chosen by visually observing the behavior of the algorithm as a function of parameter values. Once an acceptable value was found, it was modified in the direction that increased the mean dice coefficient over the testing volumes until a value was found at which the error clearly increased. The final value was chosen away from this point in the generally flat error region preceding it. Once all final values were chosen, the dice coefficients for each image set were recorded. Next, the parameter sensitivity was analyzed. This

| Parameter | $R_O$ (mm) | $\beta$ | $\gamma$ | $\delta$ | $\Delta t$ |
|---|---|---|---|---|---|
| Value | 3 | 2 | 0.05 | 4 | 0.35 |
| Sensitivity (%) | 40 | 90 | 100 | 60 | 50 |

**Table 6.** Parameter values and sensitivies.

was done by sequentially modifying parameter values in 10% increments around their selected value until the dice coefficient dropped below 0.5 on any test set.

## Results

The presented methods were implemented in C++ and tested on an Intel Core 2 Duo 2.4 GHz, Windows XP based computer. The final values of the various parameters used and the results of their sensitivity analysis are listed in Table 6. As can be seen from the table, the most sensitive parameter $R_O$ can be changed by 40% before causing the dice coefficient in any volume to drop below 0.5 in our testing set.

Segmentation of the optic nerves in a test volume requires approximately 20 minutes. Dice coefficients between the automatic segmentations and the manually delineated training and testing sets are shown in the left graph of Figure 30. This figure also shows the inter-rater results and the mean results reported by other articles. The distribution range, median, mean, and one standard deviation from the mean are indicated by the black I-bar, green line, red line, and blue box, respectively. The purple circles, red x's, and blue square's indicate approximate results reported by Gensheimer et. al. [ 17 ], D'haese et. al. [ 8 ], and Isambert et. al. [ 21 ], respectively. Bekes et. al. [ 4 ] do not report dice indices. As can be seen in the figure, mean dice coefficients achieved for the testing set are just below 0.8 for both the optic nerves and at 0.8 for the chiasm. The distributions of the testing set results are similar to those of the training set. All experiments

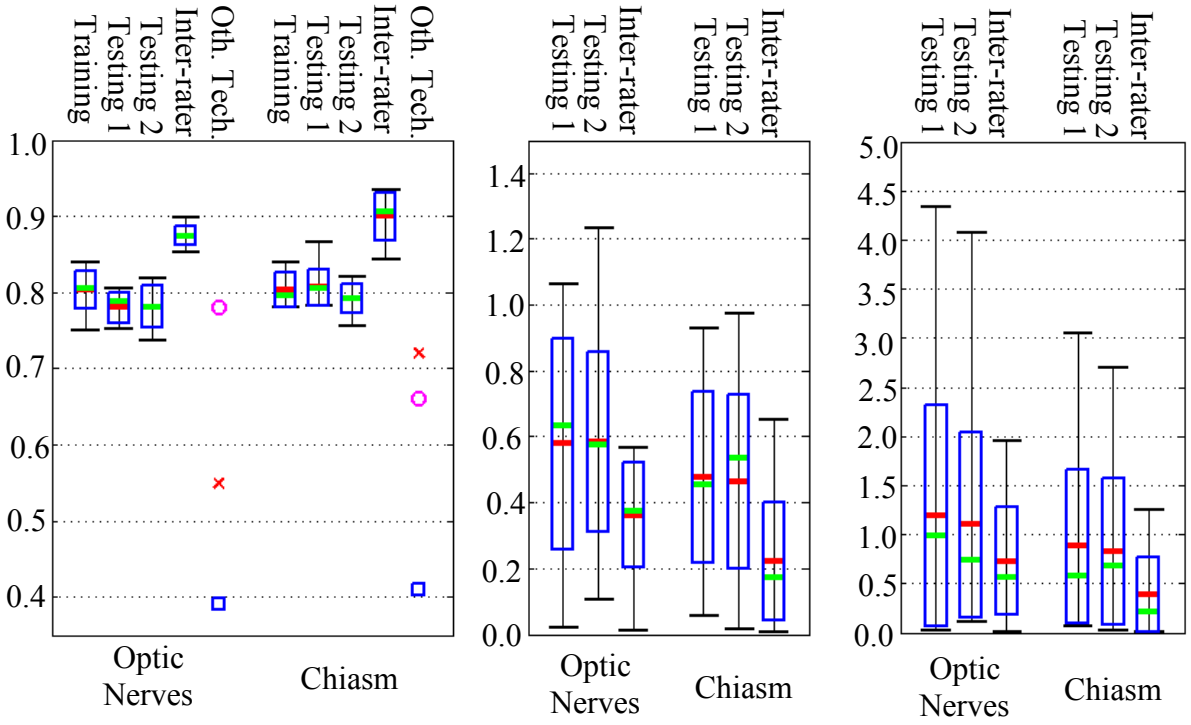


**Figure 30**. Dice coefficients and mean/max surface errors of the automatic segmentation results

result in dice indices that are larger than mean results reported by other methods, with the exception of the optic nerve results achieved by Gensheimer et. al., where comparable results are achieved. Shown in the middle and right graphs of the figure are plots of the mean and maximum surface errors in mm. Error distributions are indicated similarly to the dice coefficient plot. As can be seen from the figures, the automatic segmentations result in mean errors of approximately 0.5 mm (about half a voxel), maximum overall errors of approximately 4.0 mm (about 4 voxels), and typical maximum errors of less than 1.0 mm (less than a voxel). As can also be seen in the plots, automatic segmentations, when compared to gold standards segmentations that are manually delineated carefully in a laboratory setting, are slightly less accurate than inter-rater variability.

Renderings of several cases are shown in Figure 31. From top to bottom, the figure

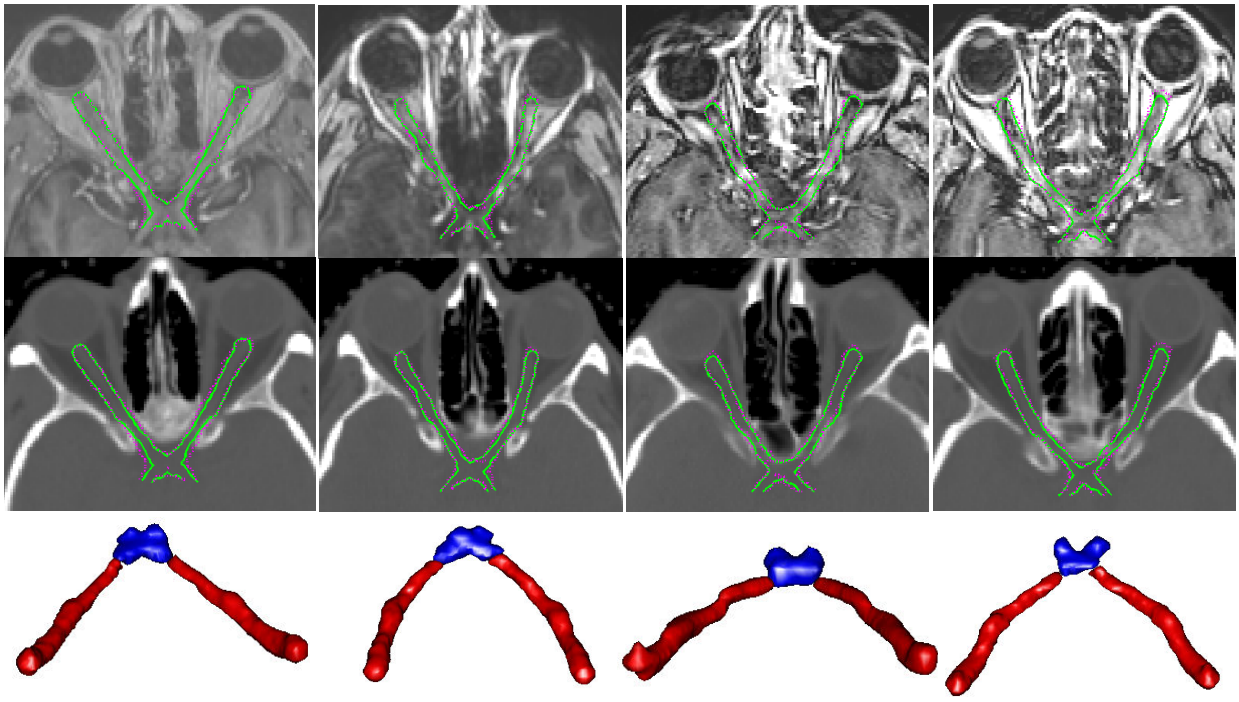

**Figure 31**. Renderings of the automatic segmentation results

contains MR data (non-contrast datasets on left, contrast enhanced datasets on right), CT data, and a 3D rendering of the structures of interest. In the 2D panes, the automatic segmentation is shown as the green, solid contour and the manual segmentation is shown with the purple dotted contour. To generate the 2D views, a thin-plate spline transformation was computed that warps the medial axes of the structures to lie in a plane, and then the image data and contours were passed through this transformation. This was done so that a cross section of the entire structure could be viewed in one 2D plane. In the 3D views, the automatically segmented optic nerves (red) and chiasm (blue) are shown. All results appear to be qualitatively accurate and reasonable.

## Discussion and Conclusions

The proposed method requires selecting a total of 5 parameters. Although this required manual adjustment, our study shows that our method is not extremely sensitive to the selection of these

parameters. Any of these parameters can be changed by more than 40% before the dice coefficient with respect to manual segmentations drops below 0.5. The model has been constructed using 4 image sets that were acquired with a single protocol (non-contrast 1.5 T), parameters trained using 10 other image sets acquired on various scanners according to varying protocols (four non-contrast 1.5 T and six contrast-enhanced 1.5 T), and tested on 10 other image sets acquired on various scanners according to varying protocols (contrast-enhanced 1.5-3 T). This shows that the method is robust to changes in various intensity characteristics, including variations not seen in the training set. Our results also show that our method is accurate. Typically, a dice coefficient of 0.8 is considered good [ 62 ], although the dice coefficient is typically unforgiving on very thin structures such as the optic nerves and chiasm. With the presented approach, mean dice coefficients of approximately 0.8 are achieved with an overall minimum dice coefficient of 0.74.

This chapter presents the most accurate method to date accomplishing automatic segmentation of the optic nerves and chiasm. The segmentation error is comparable to the inter-rater difference observed when contours are delineated without time constraint in a laboratory setting. Error in manual delineation obtained under these conditions can be considered a lower bound. The dice similarity measure is sensitive to factors such as image resolution, and the only truly fair approach for comparing the performance of two algorithms is to test them on the same dataset. Although it was not possible in this study to use the same datasets as those used in previous studies, the consistently higher dice indices that this method achieved, as indicated by the distributions in Figure 30, suggest that the methods presented in this chapter outperform previously presented approaches for localizing these structures. This claim is supported by the fact that Bekes et. al. [ 4 ], D'haese et. al. [ 8 ], and Isambert et. al. [ 21 ] conclude that their

automatic segmentations of the optic nerves and chiasm are not sufficiently accurate for radiotherapy planning. Similarly, Gensheimer et. al. [ 17 ] report dissatisfaction with their chiasm segmentation results. Their qualitative assessments suggest that the proposed method is more accurate.

This chapter also presents the first attempt to segment the optic nerves and chiasm as a union of two tubular structures. Purely atlas based methods have proven to be ineffectual for this problem. Treating the optic nerves and chiasm as two tubular structures allows the use of many general tubular structure localization algorithms, such as those based on optimal path finding approaches. However, due to lack of contrast and changing characteristics along the structures length, typical optimal-path based approaches would likely be ineffectual. The solution lies in providing the optimal path finding algorithm with a priori local intensity and shape feature information, which is the approach used by the presented algorithm.

# CHAPTER X

## AUTOMATIC SEGMENTATION OF THE OPTIC TRACTS

### Introduction

Movement disorders such as Parkinson's disease (PD), essential tremor (ET) and dystonia affect approximately 1 million, 1.5 million, and 250,000 people respectively in the United States. A viable and cutting edge treatment technique for movement disorders is deep brain stimulation (DBS) [ 49 ]. Deep brain surgery involves implanting electrodes in specific regions of the brain. This stimulation, while not completely effective, helps regulate the signals sent between different nuclei and reduces the severity of the symptoms. Because of the small size of the nuclei and the potential shift of the deep brain structures due to air invasion, this neurosurgical procedure is traditionally performed in two stages. A target location is first selected pre-operatively by a neurosurgeon and then refined intra-operatively using multiple exploratory electrodes to map the electrophysiology of the brain around the planned target. The nucleus that is targeted for stimulation depends on the disorder of the patient: the treatment of symptoms will generally be addressed by stimulating the subthalamic nucleus (STN) for PD, the ventral intermediate nucleus (Vim) for ET and the globus pallindus internal (Gpi) for dystonia.

If possible, the surgeon will stimulate Vim or STN targets in the operating room and directly observe the patient's response. This is used to refine the optimal position for the final implant. The surgeon typically is not able to see any clinical benefit when the Gpi is stimulated. This causes the surgery to be more difficult for dystonia. Programming the implant after the surgery can take from several months to, according to some groups, a year to see full clinical

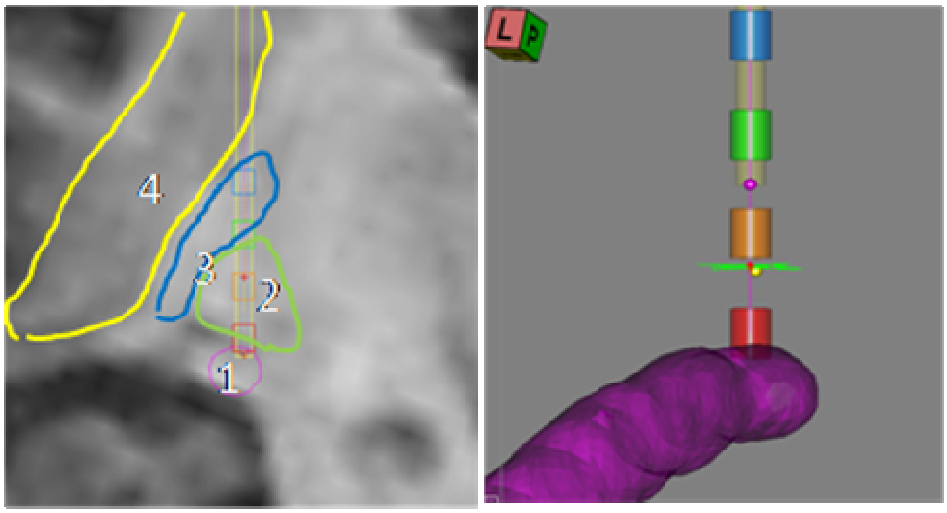


**Figure 32**. Segmentation of the optic tract (1), GPi(2), globus pallindus external (3) and putamen (4) overlaid with a typical 4-contact DBS implant. The bottom contact is placed at the superior boundary of the optic tract in a way that the two middle contacts will be in the Gpi nucleus (lower medial part of the GP).

benefit [ 50 ]. Therefore, surgeons and neurologists use other landmarks to assist during the procedure. One landmark that is commonly used is the optic tract because of its proximity to the Gpi as seen in Figure 32.

Monitoring for visual evoked potential (VEP), the electrical response of the nervous system that is directly produced by a visual stimulus, has been described by several groups as helpful in assessing the location of the optic tract during the surgery, which is the traditional approach for reliably localizing relevant functional deep brain targets [ 52 ]. However, the position of the target anatomy relative to a VEP is not obvious. MR could be used to map the target structures and the optic tracts to provide explicit information in the O.R.; however, the optic tracts are not clearly visible on the clinical imaging sequence used for DBS procedure due to poor contrast. An automated approach to segment and localize the optic tracts in MR could be beneficial as a part of a computer assistance system, in both intra-operative guidance and post-operative programming.

In this chapter, the only existing method for automatic segmentation of the optic tracts will be presented. Despite the fact that many techniques have been developed for the segmentation of tubular structures (see Chapter 2), it is unlikely that typical techniques alone could produce robust and accurate segmentation results for the optic tracts because of the unique challenges presented by the characteristics of this problem, i.e., the lack of contrast and changing intensity characteristics along the structures' length.

The optic tracts are a part of a larger anatomical structure, the optic pathways, which includes the optic nerves, chiasm, and tracts. In the previous chapter, an approach for finding the optic nerves and chiasm was presented. In this chapter, the same approach will be used to identify the optic tracts in MR. This technique overcomes the issues of lack of local intensity contrast or the changing intensity characteristics along the structures' length, making this method robust and accurate. The methods presented in this chapter were first presented at the IEEE International Symposium on Biomedical Imaging 2010 (Noble et. al. [ 34 ]).

## Methods

*Data*. With IRB approval each patient had pre-operative MRI and CT, and a post-operative CT acquired on the day of the surgery. Typical CT images were acquired at kVp = 120 V, exposure = 350 mAs and 512x512 pixels. In-plane resolution and slice thickness were respectively 0.5 mm and 0.75 mm. MRI (TR 12.2 ms, TE 2.4 ms, 256x256x170 voxels, with typical voxel resolution of 1x1x1 $mm^3$) were acquired using the SENSE parallel imaging technique (T1W/3D/TFE) from Philips on a 3T scanner.

*Segmentation*. In this chapter, the algorithm presented in the previous chapter is used to segment the optic tracts. The algorithm first computes the medial axis of the structure as the optimal path with respect to a cost function based on image and shape features. The medial axis is then expanded into the full structure using a level-set algorithm. The algorithm uses a statistical model and image registration to provide the segmentation framework with a priori, spatially varying intensity and shape information, thus accounting for unique local structure features. The statistical models were based on the manual segmentations of the SOI's in 6 volumes, which were not part of the testing set.

Once the models are built as described in the previous chapter, new sets of images can be segmented. Briefly stated, the algorithm works in three steps: 1) the starting and ending points of the structure centerline are found using a basic atlas-based technique using non-rigid registration, 2) a minimum cost path algorithm finds the centerline of the structure, and 3) the structure centerline is used to initialize a level-set based geometric deformable model to segment the full structure.

*Evaluation.* The goal of the method is to segment the optic tracts to be used as an internal landmark for guidance in deep brain surgeries. Therefore the validation is focused on the section of the optic tracts next to the Gpi. The segmentation is first visually evaluated in 8 patients, and the related Dice index and symmetric mean surface error distances are computed between the automatically segmented tracts and the manual segmentation. The manual segmentation was drawn by a rater and reviewed by an experienced neurosurgeon.

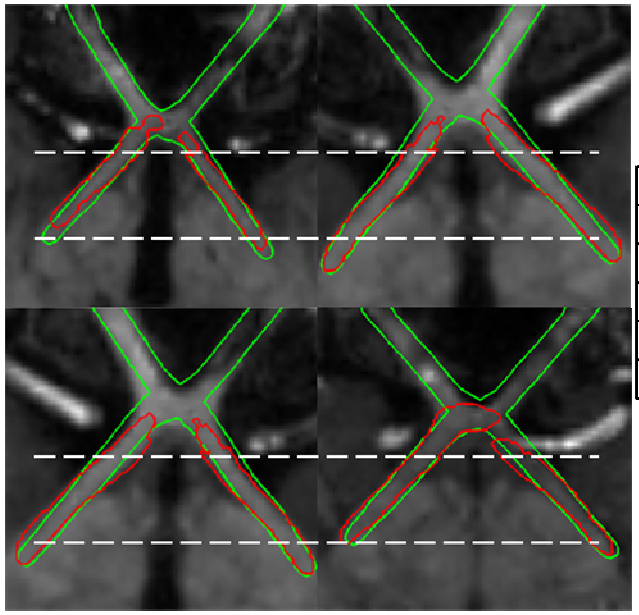

**Figure 33**. Visual evaluation of the results: automatic (green/light gray) are compared to the manual segmentation of the tract segment around the Gpi, defined by the white dotted lines.

| Testing image | 1 | 2 | 3 | 4 | Mean |
|---|---|---|---|---|---|
| Dice index | 0.74 | 0.73 | 0.83 | 0.82 | 0.76 |
| Mean Dist. (mm) | 0.25 | 0.24 | 0.15 | 0.24 | 0.31 |
| Testing image | 5 | 6 | 7 | 8 | Stdev |
| Dice index | 0.74 | 0.7 | 0.75 | 0.77 | 0.04 |
| Mean Dist. (mm) | 0.28 | 0.56 | 0.3 | 0.5 | 0.14 |

**Table 7.** Dice coefficients and surface errors of the automatic segmentation results

## Results

Results for several cases are shown in Figure 33. The automatic segmentation is shown as the green, solid contour and manual segmentation is shown with the red contour. To generate the 2D views, a thin-plate spline transformation was computed that warps the medial axes of the structures to lie in a plane, and then the image data and contours were passed through this transformation. This was done so that a cross section of the entire structure could be viewed in one 2D plane.

Table 7 shows the Dice index and mean surface errors between the manual and automatic segmentation of the optic tracts in 8 patients. A rendering of the results, color-encoded with error distance, is shown in Figure 34.

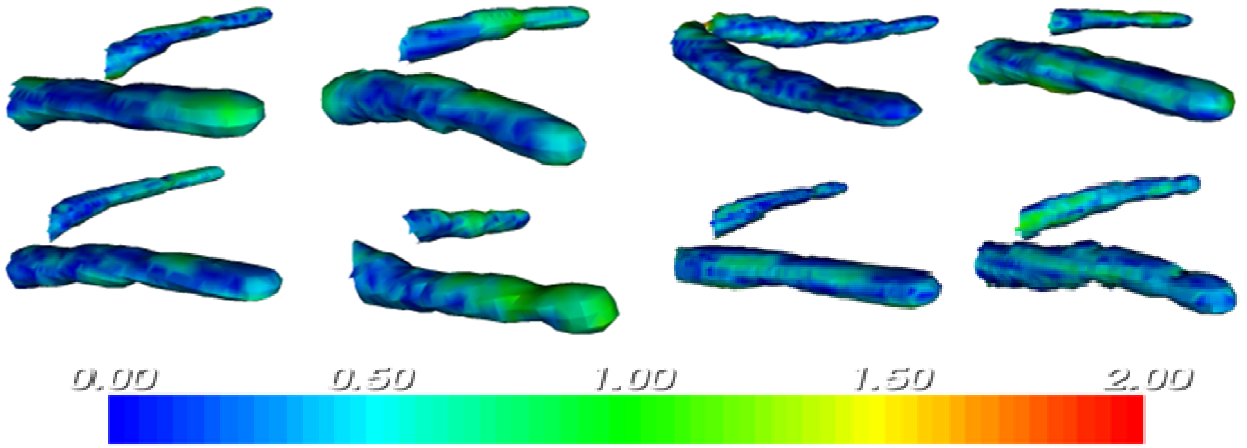


**Figure 34**. Segmentation results for images 1-8 (left-right, top-bottom), color-encoded with distance (mm) to the manual segmentation.

## Discussion and Conclusions

Results presented in this chapter corroborate the findings of the previous two chapters. The automatic segmentations of the optic tracts appear to be qualitatively and quantitatively accurate. Typically, a dice coefficient of 0.8 is considered as an accurate segmentation for a regular sized structure, although the dice coefficient is typically more unforgiving on very thin structures such as the optic tracts, so 0.7 is considered a good result [ 62 ]. With the approach presented in this chapter, mean dice coefficients of 0.76±0.04 are achieved with an overall minimum of 0.70.

For this application, an accuracy of 2 mm is acceptable, while 1 mm (1 voxel wide in MR) is desirable. Error in the results on average is only 0.31 mm, while the worst error seen was 1.6 mm. Thus, the results from every experiment in this study were considered acceptable. In addition to being useful for pre-operative planning, the automatic segmentations could be used for intra-operative guidance. A difference between the location of the segmented optic tracts and the location at which the optic tracts are detected by monitoring the visual evoked potential could prove to be a robust measurement of intra-operative brain shift. This concept is illustrated on two cases with and without substantial intra-operative brain shift. The brain shift was evaluated by

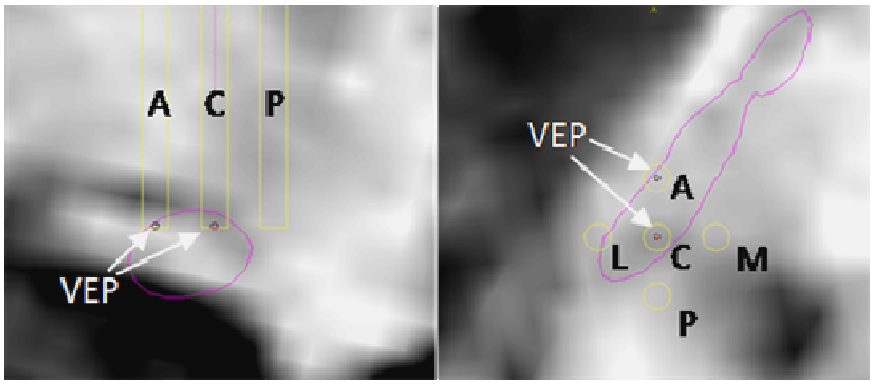


**Figure 35**. Coronal and axial view of the patient MR T1 image overlaid with the cross section of the automatic segmentation of the optic tract (purple) and the 5 test tubes inserted in the brain during the surgery (yellow). The VEPs are shown as small pink tube sections

analyzing the size of the air pocket in the post-operative CT of the patient acquired immediately after the procedure, as described in detail in [ 41 ]. Figure 35 presents the automatically segmented optic tract overlaid with the recording electrodes used intra-operatively for the patient with no air invasion. Five electrodes were used to record VEPs, labeled A (anterior), C (center), P (posterior), L (lateral), and M (medial). The evidences of VEP correlate well with the location of the automatic segmentation of the tracts.

In Figure 36 the concept of brain shift measurement is illustrated. The left panel shows a coronal section including a 4-contact implant overlaid with the pre-operative MRI. Based on the pre-operative plan, the position of the optic tract should overlap with the two bottom contacts, as demonstrated on the left in the figure. However, the location of the VEPs recorded during surgery are not consistent with the predicted optic tract position, thus suggesting brain shift.

Albeit preliminary, the presented results hold promise for intra-operative guidance. A combination of automatic segmentation of the optic tracts and intra-operative recordings could provide accurate measurement of intra-operative brain shift in the target area (in this case estimated to be 1.2mm). This could, in turn, be used to update information based on pre-

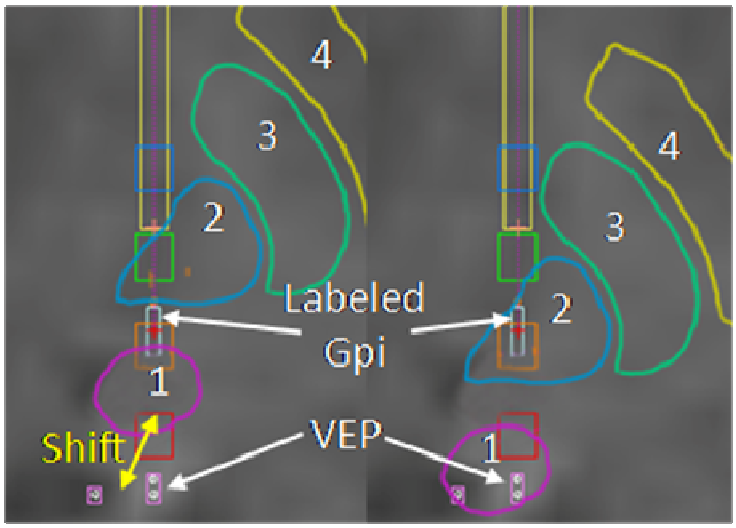


**Figure 36**. The left panel shows a manually segmented (1. Optic tract, 2. Gpi, 3. Gpe, 4. Putamen) coronal section of the pre-operative MRI overlaid with intra-operative data (VEP and labeled Gpi) and the implant. The right panel shows the segmented structures manually moved to match the intra-operative recordings.

operative images, thus providing intra-operative guidance. This is illustrated in the right panel of Figure 36. Here, anatomical structures segmented in the pre-operative images have been displaced by the shift estimated from the difference between the segmented optic tracts and the intraoperative measurements. The same could be done for electrophysiological maps indicating likely regions of efficacy and/or side effects [ 41 ] [ 31 ].

# CHAPTER XI

## A NEW APPROACH FOR TUBULAR STRUCTURE MODELING AND SEGMENTATION USING GRAPH-BASED TECHNIQUES

### Introduction

Tubular structure localization has been studied extensively over the past few decades. Algorithms capable of localizing tubular structures have been extremely useful for identifying vessels, arteries, nerves, and other tubular shaped structures in medical images. A brief review of the different classes of algorithms is presented in Chapter 2. A more extensive review was presented by Lesage et. al [ 28 ]. One common segmentation approach is to use anatomy-specific models constructed from a training set. For instance, de Bruijne et. al. [ 6 ] use specialized active shape models to localize abdominal aortic aneurysms. In the work presented herein, a new algorithm is presented that can be used to segment a wide variety of tubular structures using a novel approach to build anatomy-specific models, which requires only a small number of sample segmentations for training. The segmentation problem is solved by finding an optimal path in a high-dimensional graph. The graph is designed with novel structures that permit the incorporation of prior information from the model into the optimization and account for several weaknesses of traditional graph-based approaches.

One common approach for tubular structure extraction is to treat it as the problem of finding an optimal path in a graph, $G=\{V,E\}$, using, for instance, the optimal path finding techniques presented in Chapter 7. Typically, when used for tubular structure segmentation, $V$, the set of nodes, corresponds to the image voxels; $E$, the set of edges, connects neighboring voxels; and the costs associated with the edges are defined by some heuristically derived

function. This basic approach is widely used because it tends to be easy to implement, straightforward to apply simple cost functions, easy to use in automated or semi-automated application, is computationally efficient, and most importantly, it finds a global optimum. The algorithm also has four identifiable limitations: (1) It is difficult to incorporate local and/or training-based prior knowledge into the segmentation framework. Thus, the SOI (Structure Of Interest) must be enhanced prior to segmentation. This is not always possible as many tubular structures have intensity characteristics that vary along their length and borders that lack contrast. (2) The algorithm identifies a curve, rather than the 3D structure. Further processing must be applied to identify the full structure. Some researchers have addressed this limitation by finding a 4D optimal path [ 29 ], where the first 3 dimensions correspond to the voxels as usual, and the $4^{th}$ dimension corresponds to structure width. Their cost function has terms that favor smoothly varying width along the structure. When the algorithm is completed, the result is the structure centerline, and a width value at every centerline point. Thus, using an optimal path finding algorithm, they are able to compute a representation of a 3D tubular structure that is constrained to circular type cross-sections. (3) The algorithm is biased with respect to length, i.e., because the algorithm finds the path composed of edges with the minimum cumulative costs, it tends to favor shorter, straighter paths, which can result either in segmentation failure or in erroneous shortcutting at structure corners. (4) The method requires manual definition and tuning of cost functions.

The approach presented in this chapter is to use an unconventional optimal path-based extraction scheme. Segmentation is performed on a unique high-dimensional graph, which is designed to incorporate information from a novel tubular structure model. This approach addresses limitations of the traditional 3D formulation.

## Overview

The segmentation approach requires a model of the SOI that is constructed using a training set. After being supplied with a set of images in which the SOI is segmented, the training approach proceeds as follows: (1) A point correspondence between the centerlines of the exemplar segmentations is extracted using a graph-based optimization technique. (2) The structure segments corresponding to the same anatomical sub-regions across the structures, referred to as "localities," are extracted using the centerline correspondence and a gradient descent optimization method. (3) Statistics on a set of local intensity and shape features are sampled in each locality of each training sample. (4) Weighting values are computed that assign a higher importance to more discriminant features. The training process results in a unique tubular structure model that contains statistics and weighting values for a set of features associated with each subsection of each exemplar in the training set.

Once the structure model is constructed, it can be used to segment new images using an optimal path finding algorithm. This algorithm requires three pieces of information: (1) the curve's starting and ending voxels, (2) a defined graph structure, and (3) a defined cost function. In this work, the starting and ending voxels were chosen manually (this could be automated in future work using, for instance, atlas-based methods as done in Chapters 8-10). A novel graph structure has been implemented that is unique in two ways: (1) Nodes are associated with specific training samples/localities. This allows costs to be computed using information from the model associated with each sample/locality. Also, this permits reconstruction of the full 3D structure from the optimal path by assuming, since the structure features that will be used are width dependent, that the width of the structure at a node in the optimal path should match the width of the sample in the locality indicated by the node. (2) Novel graph structures are

implemented such that the only paths that exist between the seed and end-node obey hard constraints on the minimum/maximum length of the structure. The length of a path in a graph can be constrained by creating an abstract dimension $L$ in the graph. The seed lies at $l$=0 in $L$, and a set of candidate endnodes are chosen at $l$=[$l_{\min}$, $l_{\min+1}$ , … , $l_{\max}$] in $L$. The graph is constructed such that all neighbors of the seed are at $l$=1, all of their neighbors are at $l$=2, and in general all neighbors of node $v_i$ at $l_i$ lie at $l_i$+1. It follows from this formulation that $v_i$ is only reachable from the seed in $l_i$ steps, i.e., by paths that contain $l_i$ edges. An optimal path algorithm will find the path from the seed to the candidate endnode that results in the path of minimal cost, and because the graph is structured so that the only paths that exist between the seed and endnodes contain a bounded number of edges, it is guaranteed that the optimal path will have length on the interval [$l_{\min}$ , $l_{\max}$]. This concept can be extended to apply length constraints on each individual locality. This ensures that the lengths of localities in the optimal path, and the length of the path itself, are within specified bounds. The bounds on the length of each locality are chosen as the range of the lengths of each locality in the training set. Using this graph and the cost function described later, an optimal path algorithm can be used to segment new images.

In the following sections, these methods are described in detail. First, the structure model that stores the local a priori information is introduced. Following this will be details about the construction of the structure model. This is followed by a presentation of the ideal graph structure for segmentation. It will be shown that performing a search on such a complex graph is computationally inefficient. Then, a pruned graph will be presented that drastically improves computational performance without introducing significant compromise. Finally, the strategy for validating the proposed methods is presented.

## Structure Feature Model

The structure model will contain statistics and weighting values for a set of features associated with each subsection of each exemplar in the training set. Before feature values can be sampled, the SOI must be separated into a set of distinct localities $\{p\}$, which is performed automatically as part of the training algorithm. A locality $p$ of a particular training sample $s$ can range from one to any number of voxels in length. Figure 37 shows a synthetic 2D example of a SOI with 2 training samples divided into 5 localities. As shown, the localities correspond to similar regions across each training structure. After the localities are determined, feature statistics and weights can be computed.

Although other features could be used, in this work intensity and shape-based features that are used for training and segmentation include: (1) intensity values along radii normal to the curve at distances of 0, 0.2, 0.9, and 1.1 times the width of the structure at every $\pi/4$ radians; (2) intensity derivatives on those radii at 0.1 and 1.0 times the structure width; and (3) the curve orientation, which provides a local prior on the curve's shape. To use these affine-variant features, all of the images must be affinely aligned to an image chosen as the reference (see Chapter 3). Later, when segmenting a target image, the target image should be affinely registered to the reference image as a preprocessing step.

## Building the Structure Model

Point correspondence across centerlines is necessary to determine the localities that correspond to the same distinct regions across the sample SOIs, and is extracted using a graph-based optimization. One training volume is chosen as the reference. The search determines one voxel along each training centerline to be associated with each reference centerline voxel. The cost function used in the search is optimized when local features between the corresponding voxels

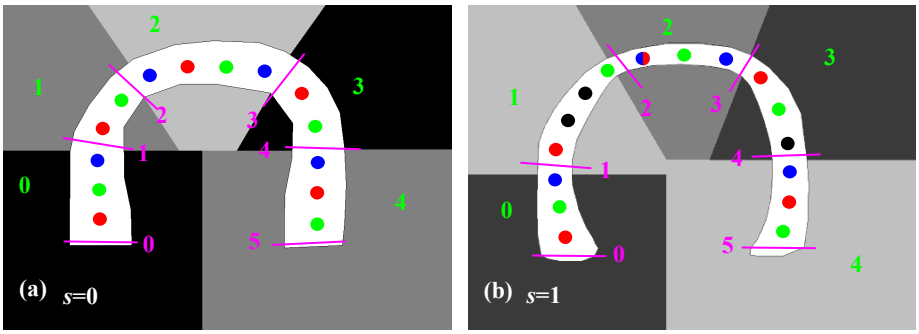


**Figure 37.** Synthetic 2D example of a structure model. (a) and (b) are the reference and training SOIs. RGB values of centerline points indicate correspondence across volumes. Localities and their boundaries are indicated in green and purple

are best matched. The color-coded voxels in the centerlines in Figure 37 represent example correspondences. In (b), black voxels indicate no association with a reference voxel, and multiple colors indicate multiple associations. The main purpose of this step is to provide a way to establish a correspondence between locality boundaries across training samples. Thus, a one-to-one correspondence between centerline points is not necessary.

The problem of finding correspondence between the reference and a training centerline is treated as the problem of finding a minimal cost path in a 2D graph, $G = \{V, E\}$. The $i$th node in $V$ can be written as $v_i = (r_i, q_i)$, where $r_i$ and $q_i$ are indices to points along the centerlines of the reference and training volumes, respectively. If $v_i$ is a node in the segmented path, then a correspondence will be established between the $r_i$th point in the reference centerline and the $q_i$th point in the training volume centerline. The edge structure is such that $e_{i,j}$ exists if $r_j = r_i + 1$ and $q_i \leq q_j \leq q_i + 1 + \delta$. The seed and end-node are set to $v_s = \{0,0\}$ and $v_e = \{N-1, M-1\}$, where $N$ and $M$ are the length of the reference and training centerlines. This graph structure is chosen so that the resulting path will describe point correspondence with desirable properties: (1) because $e_{i,j}$ implies $r_j = r_i + 1$, this enforces that every reference centerline point will be in the minimal cost path once, thus, every reference centerline point will correspond to one and only

one point in the training volume. (2) $q_i \leq q_j$ enforces that lines of correspondence cannot cross in a reverse direction as we move along the reference structure. (3) $q_j \leq q_i + 1 + \delta$ enforces that lines of correspondence cannot skip over substantial lengths of the training structure. (4) The choice of $v_s$ and $v_e$ ensure that the entire reference and training structures are included. The parameter $\delta$ is set to 3 for all structures tested in this chapter.

All that is needed now is to define a cost function. The cost function used here is:

$$C(e_{i,j}) = \sum_{k=0}^{K-1} \frac{\left(f_r^k(r_j, r_i) - f_q^k(q_j, q_i)\right)^2}{\sigma_r^k(r_j, r_i)\sigma_q^k(q_j, q_i)}, \quad (21)$$

where the numerator compares $f_r^k(r_j, r_i)$, the value of the $k$th training feature at the transition between points $r_j$ and $r_i$ in the reference centerline, to $f_q^k(q_j, q_i)$, the value of the $k$th training feature between $q_j$ and $q_i$ in the training centerline, and there are $K$ total features. That difference is then scaled by estimations of the standard deviation of each feature, which can be estimated in a local region around that particular centerline point. Finally, all those scaled feature differences are summed into a total cost for the edge. A path through this graph that optimizes this feature based cost function will provide a point correspondence within the constraints of the graph that matches local feature values of the centerlines. Applying Dijkstra's algorithm to this graph once for each training volume in the training set will result in one corresponding point in each training volume centerline for every reference centerline point.

Although this could be automated, the number of localities for each SOI is chosen manually. However, the positions of the locality boundaries are determined automatically using a gradient descent technique. The cost function used favors smaller intra-locality variance and larger inter-locality variance of model features across all the training samples, and it includes a regularization term that favors boundaries spread evenly across the SOI. An example of optimal

locality selection is shown in Figure 37. The points on the centerline reference structure can be numbered 0 to *N*-1 going from the start to end point. The *M* localities can be determined by *M*+1 boundaries $\{b\}$ that bracket the *M* localities. Each boundary corresponds to a numbered point along the SOI. $b_0$ and $b_M$ are fixed at the start and endpoint of the structure. A particular set of boundaries implies a particular set of localities in the reference structure; and because a correspondence is established from the reference to the training centerlines, this in turn implies the localities in the training centerlines. So, the boundaries will be the parameters in the optimization algorithm to find a desirable set of localities.

The boundaries are initialized such that they are evenly divided across the reference centerline. From there, each boundary is visited sequentially, and adjusted in increments of 1 centerline point in the direction that reduces the value of a cost function. This process is done iteratively until the boundaries can no longer be adjusted to decrease the overall cost. The cost function used is defined as

$$C(\{b_i\}_{i=1}^{M-1}) = \sum_{i=1}^{M-1} C(b_i), \quad (22)$$

and

$$C(b_i) = (|b_{i-1} - b_i| + |b_{i+1} - b_i| - 2b_{\mathrm{E}})^2 + \sum_{k=0}^{K-1} \left( \left(\sigma_{p_{i,i}}^k\right)^2 - 2\left(\sigma_{p_{i,i-1}}^k\right)^2 + \left(\sigma_{p_{i-1,i-1}}^k\right)^2 \right), \quad (23)$$

where the first term is a regularization term on the distances between the boundaries, and the summation punishes high intra-locality variance and low inter-locality variance on the *K* feature values. $b_i$ is the boundary between localities $p_{i-1}$ and $p_i$ (see Figure 37). $b_{\mathrm{E}}$ is the distance between boundaries when they are evenly spaced over the SOI. Thus, the first term is minimized when the boundaries are not too far or too close to each other. $\left(\sigma_{p_{i,i}}^k\right)^2$ and $\left(\sigma_{p_{i-1,i-1}}^k\right)^2$ are the

intra-locality variance of the *k*th feature value in the localities adjacent to $b_i$. Thus, this term is minimized when the variance of the features in each of these localities is low. $\left(\sigma^k_{p_{i,i-1}}\right)^2$ is the inter-locality variance of the *k*th feature value across the localities adjacent to $b_i$. The negative of this term is minimized when the feature variance between neighboring localities is large. Using this cost function, the gradient descent optimization of the boundaries converges to a solution and determines the distinct localities of the SOI. Once the localities are found, the mean and variance of each feature for each locality/sample are computed.

The next step involves determining the features that are the most discriminant in each locality. A desirable feature is one that produces a low cost inside the SOI relative to the background. To identify these, features are first computed across the training images, and the resulting feature images are convolved with a valley filter with scale equal to the structure width. A high positive response indicates a desirable feature. A low or negative response indicates an undesirable one.

To train the features of sample $s_i$, which indicates a particular exemplar in the training set, the effectiveness of each feature associated with $s_i$ is tested against the other training samples. Each voxel in the image associated with each $s_{j\neq i}$ is treated as a feature testing point. For each *t*th voxel to be visited, $\vec{v}_t$, the closest point on the structure centerline in $s_j$, $\vec{c}_t$, is found. The values of the features in $s_i$ to be tested will be those associated with $p_t$, the locality to which $\vec{c}_t$ belongs. Then, the un-weighted cost value for each individual feature associated with locality $p_t$ in $s_i$ is computed at $\vec{v}_t$. The weighting values will be chosen using a valley detector on the cost values, with the valley kernel at $\vec{v}_t$ given by

$$\rho_t = (\|\vec{c}_t - \vec{v}_t\|^2 - {w_t}^2)e^{-\frac{\|\vec{c}_t-\vec{v}_t\|^2}{2{w_t}^2}}. \quad (24)$$

This is a valley detecting kernel based on the 2nd derivative of a Gaussian with scale $w_t$, the expected width of the structure at $\vec{c}_t$ given by the model. The weights can then be computed as

$$\omega_{s_i,p_t}^k = \max\left(0, \sum_{t=0}^{T-1} \rho_t \, \tilde{C}_{s_i,p_t}^k(\vec{v}_t)\right), \quad (25)$$

where $\tilde{C}_{s_i,p_t}^k(\vec{v}_t)$ is the un-weighted cost value of the *k*th feature associated with sample $s_i$ in locality $p_t$ evaluated at $\vec{v}_t$. By summing the un-weighted feature costs multiplied by the kernel values over the *T* voxels that are tested, a convolution between the valley detecting kernel and the un-weighted cost values of the features is being performed. Weights are set to zero anytime the filter produces a negative result. This process for choosing weighting values on the features should give higher weight to features that are more discriminant. After this step, the creation of the models is completed. These consist of a sequence of structure samples, each sectioned into localities representing similar anatomical regions; and statistical values and importance weightings for a set of features within each sample/locality.

## Ideal Graph

Once the structure model is constructed, it can be used to segment new images using an optimal path finding algorithm to optimize a high-dimensional graph $G = \{V, E\}$. In the ideal solution, *V* will be structured as a 7D set of vertices. Three dimensions still correspond to individual voxel locations. Let the *i*th node in *V* be denoted $v_i = (x_i,y_i,z_i,s_i,p_i,l_i,w_i)$. $x_i$, $y_i$, and $z_i$ describe the corresponding voxel position. $s_i$, $p_i$, $l_i$, and $w_i$ denote the sample, locality, length, and width quantities associated with that node. As explained above, sample indicates a particular exemplar in the training set, and locality indicates one of the discretely labeled segments of the SOI. These values of a node will dictate which features from the model will be used to assess the feature

based cost of transitioning to that node. $w_i$, similarly to Li and Yezzi's [ 29 ] fourth dimension, corresponds to the width of the structure at that point. Length quantifies the number of nodes in the current locality of the path traced from this node back to the seed. At first glance, this length value may seem to violate the condition for finding the global optimum using Dijkstra's algorithm that the graph must be static with respect to the search, but as will be shown, the graph can be statically structured while allowing knowledge of the length of the path to the node in its locality.

The graph will be constructed such that the $s$,$p$,$l$, and $w$ values of a node can be used to achieve the following desirable properties: (1) With tubular structure detection, the best discrimination functions should be regionally effective. For example, it is unlikely that a single feature-based cost function is adequate to describe local intensity and shape features for the entire SOI. On the other hand, using many cost functions that are too local can also be negative. Discrimination functions that have effect over the local range of a single voxel of the SOI centerline could lead to overfitting, producing inaccurate segmentation results. Thus, cost functions with regional effect are desirable, because they balance the tradeoffs between features that are too global or too local. (2) The locality, or region, labels on the centerline points from the start to the end of the SOI in the model are monotonically increasing, and a path in the graph should share this property. (3) Anatomical structures that are tubular in shape should vary smoothly in width. (4) The length of each locality in the final solution should be similar to the lengths of the localities of the sample structures in the model. The set of edges $E$ and their costs will be used to achieve these properties.

$E$ will connect only, but not all, vertices that correspond to voxel neighbors. A directed edge $e_{i,j}$ from vertex $v_i$ to vertex $v_j$ exists only if $(x_j, y_j, z_j) \in \text{nbhd}(x_i, y_i, z_i)$, which is the neighborhood criterion. Given that criterion is met, an edge exists in two possible situations:

$$p_j = p_i \,,\; l_j = l_i + 1 \,, \tag{26}$$

or

$$p_j > p_i \,, l_j = 1 \,. \tag{27}$$

These conditions enforce two constraints: (1) the constraints on $p$ enforce that the locality values are monotonically increasing from the seed to end node. (2) The constraints on $l$ are used so that a cost function can be designed that favors paths between the seed and endnodes that contain a specified number of edges in each locality, i.e., a cost function that targets paths with localities of a given *a priori* length. A more general example of this is a graph that targets paths of a specified overall length. As introduced earlier, this can be accomplished by creating a graph with an abstract dimension $L$. The seed lies at $v_\text{s} = (x_\text{s}, y_\text{s}, z_\text{s}, l_\text{s} = 0)$, and a set of candidate endnodes are chosen at $\{v_\text{e}\} = \{x_\text{e}, y_\text{e}, z_\text{e}, l_\text{e} = [1,2, \dots, \infty]\}$. The graph is constructed such that all neighbors of the seed are at $l$=1, all of their neighbors are at $l$=2, and in general, all neighbors of node $v_i$ at $l_i$ lie at $l_i$+1. It follows from this formulation that $v_i$ is only *reachable* from the seed in $l_i$ steps, i.e., by paths that contain $l_i$ edges. Thus, all possible paths from the seed to the endnode $(x_\text{e}, y_\text{e}, z_\text{e}, l)$ must contain $l$ edges. The graph's cost function can be used to target paths of a given *a priori* desired structure length $l_\text{d}$, by for instance, placing a cost on the edges to an endnode $(x_\text{e}, y_\text{e}, z_\text{e}, l)$ of the form $\alpha|l_\text{d} - l|$, which is minimized when $l = l_\text{d}$, i.e., when the optimal path from the seed to an endnode contains the desired number of edges. This concept is extended to apply length constraints on each individual locality. The two edge conditions on the

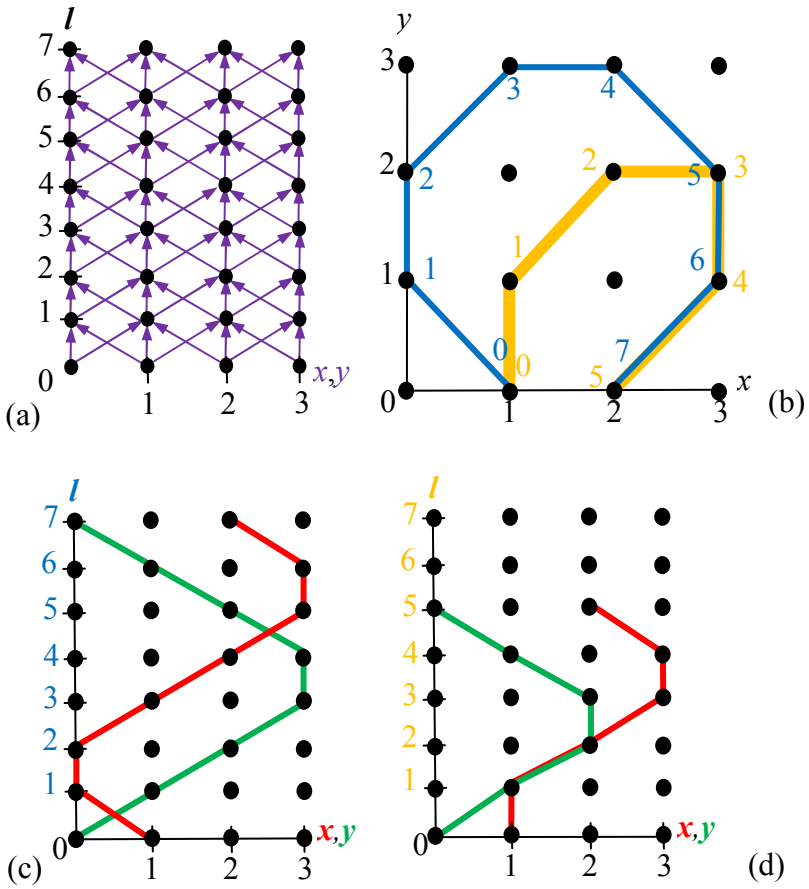


**Figure 38.** Graphs demonstrating the behavior of length values. The reachability map of the graph is shown in (a). Two allowable paths are also plotted in $x$ vs. $y$ (b), $x$ and $y$ vs. $l$ for the blue path (c), and $x$ and $y$ vs. $l$ for the orange path (d).

ideal graph in Eqns. ( 28 ) and ( 29 ) enforce the condition that a node $v = (x, y, z, s, p, l, w)$ belongs to a path that contains $l$ edges in locality $p$. Thus, the $l$ value of a node in the abstract dimension $L$ describes the length in locality $p$ of any path that can reach $v$. The cost function presented below is designed to favor localities in the optimal path that are of a specified length. Figure 38 shows a 3D example of the behavior of the graph in the length dimension. In the example, nodes are labeled ($x$,$y$,$l$). For the sake of simplicity, the locality dimension is ignored, but one can assume that this is an example of a one locality structure without loss of generality. (a) shows the reachability plot of the graph. Notice that directed edges exist only in the forward

direction in the length dimension. (b) shows an ($x$,$y$) plot of the graph. Examples of two allowable paths in the graph are shown in blue and orange. The numbers along the paths indicate the length values associated with the nodes in those paths. The seed and end-nodes would be at ($x$,$y$,$l$) = (1,0,0) and (2,0,.). (c) and (d) show the $l$ values for the blue and orange curves vs. $x$ and $y$ (red and green). As illustrated by the figure, an allowable path in the graph traverses nodes associated with incrementally increasing length values; and several nodes correspond to one ($x$,$y$) position, because each ($x$,$y$) position may be reachable by paths of different lengths.

The cost of a directed edge from $v_i$ to $v_j$ can be written as

$$C(e_{i,j}) = F(e_{i,j}) + N(e_{i,j}) \ , \tag{28}$$

where $F$ and $N$ are the feature-based and graph-based cost terms. The graph-based term can be decomposed as

$$N(e_{i,j}) = S(e_{i,j}) + P(e_{i,j}) + W(e_{i,j}) \ , \tag{29}$$

where

$$S(e_{i,j}) = \begin{Bmatrix} \beta_1 & s_j \neq s_i \\ 0 & s_j = s_i \end{Bmatrix} \ , \tag{30}$$

$$P(e_{i,j}) = \begin{Bmatrix} \beta_2\left(-l_i + \sum_{k=i}^{j-1} E_l(p_k)\right) & p_j \neq p_i, E_l(p_i) > l_i \\ \beta_2(l_i - E_l(p_i)) & p_j = p_i, E_l(p_i) < l_i \\ 0 & \text{Otherwise} \end{Bmatrix} \ , \tag{31}$$

$$W(e_{i,j}) = \beta_3 |w_j - w_i| \ , \tag{32}$$

are the sample, locality/length, and width-based cost terms. Eqn. ( 30 ) penalizes transitions between nodes corresponding to different training exemplars, to enforce the property that the discrimination functions should be regional in effect. Eqn. ( 31 ) favors transitions between

nodes of different localities when the current length $l_i$ is greater than $E_l(p_i)$, the length locality $p_i$ is expected to be. Similarly, it favors transitions between nodes of the same locality when the current length $l_i$ is less than $E_l(p_i)$. By computing $E_l(p_i)$ using training data, Eqn. ( 31 ) ensures that localities of the minimum cost path are similar in length to the training data, and that the overall structure length is similar to the structure length of the training data. Eqn. ( 32 ) favors transitions between nodes with similar width values. This penalizes segmentations that do not vary smoothly in diameter.

There are many choices of feature based cost function that could be successful in this application. The general form of the cost function can be written as

$$F(e_{i,j}) = A(e_{i,j}) + B(e_{i,j}) \ , \qquad (33)$$

where $A$ is a function that includes all the model-based feature costs that are dependent on the training data, and $B$ is a function that includes any model-independent, image-based cost functions, e.g., vesselness filters.

Applying Dijkstra's algorithm to this graph would be guaranteed to find the 3D tubular structure in the range of the graph, which is not biased toward straight line paths, and is optimal with respect to the cost function and graph structure. However, the graph itself will be overly-expansive for any practical application. Because of the length dimension associated with each node, the graph is theoretically infinite in the number of nodes and edges. However, even if the maximum length is constrained to 20 nodes, to find a structure in a 26-connected neighborhood, in a 50x50x50 voxel image with 10 training samples, 10 localities, and 5 possible widths, the graph would consist of over a billion nodes and over a trillion edges. For practical purposes, this is not an efficient solution. However, this ideal graph structure can be pruned into an efficient solution.

Pruned Graph

For the practical solution, a pruned graph is constructed. The ideal graph presented in the previous sub-section used the cost function as a means of achieving the desired node transition properties. In contrast, the pruned graph will have an altered edge structure that places hard constraints on the solution to account for the limitations of a traditional graph. $V$ will be structured as a 6D set of vertices. Let the $i$th node in $V$ be denoted $v_i = (x_i,y_i,z_i,s_i,p_i,l_i)$, where values are defined similarly as in the previous sub-section. Note that the width dimension is removed. As will be shown, the structure width will still be implied in the optimal path, thus a 3D tubular segmentation will still be achieved.

$E$ will connect only, but not all, vertices that correspond to voxel neighbors. Directed edges exist in two situations. In both situations, for there to be an edge connecting vertex $v_i$ to vertex $v_j$, then $(x_j, y_j, z_j) \in \text{nbhd}(x_i, y_i, z_i)$ must be satisfied. If this voxel neighborhood criterion is satisfied, the edge exists if either: (1) $l_j = l_i + 1$, $p_j = p_i$, $s_j = s_i$, and $l_i < \gamma_1(p_i)$, which is a constraint that allows $v_i$ to transition to nodes with an incremented length value and corresponding to the same training sample and structure locality only if the length of the structure locality at $v_i$ is less than $\gamma_1(p_i)$, the bound on the maximum length of locality $p_i$; or (2) $l_j = 1$, $p_j = p_i + 1$, and $l_i \geq \gamma_2(p_i)$, which allows transitions between any sample, but only from locality $p_i$ to $p_i + 1$, only when the length value of $v_j$ is restarted at 1, and only when the length of the locality at $v_i$ is greater than or equal to the bound on the minimum length of locality $p_i$, $\gamma_2(p_i)$. These transition properties are illustrated in Figure 39. (a) shows an example of an $l$ vs. $p$ reachability plot, where the red boxes are used to indicate the min/max bounds placed on the length of each locality by $\gamma_1$ and $\gamma_2$. $l$ must incrementally increase in value in a locality until it is long enough to transition to the next locality, where the $l$ value is restarted at one. (b)

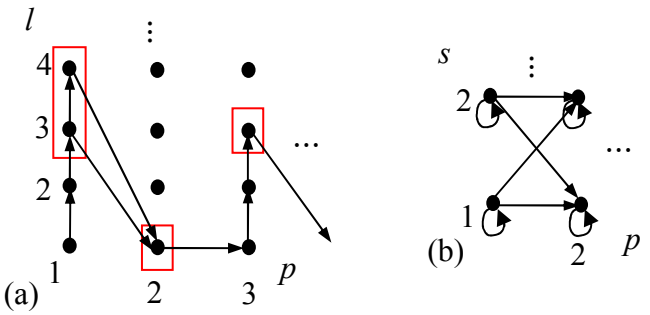


**Figure 39.** Transition properties of the length, locality, and sample dimensions in the pruned graph.

illustrates the property that $s$ can only change with a transition of locality, at which point, it is allowed to transition to any sample.

By defining the edge structure in this way, there are hard minimum and maximum constraints on the length of each locality as well as the entire curve, and transitions to different training samples are only allowable upon transition from one locality to the next. Because the graph is constructed in this way, the structure of the graph is unique to each SOI. In this sense, the graph is trained using a priori information. This training allows the edge structure to be substantially pruned, which reduces the density of the graph. In practical application, the graph is not fully constructed prior to the optimization, but rather, when a node is reached in the search, a function is used to determine the node's neighbors according to the criteria above. Pseudocode of the neighbor function is shown in Figure 40.

The cost of a directed edge from $v_i$ to $v_j$ for the pruned graph can be written as

$$C(e_{i,j}) = F(e_{i,j}) \ . \qquad (34)$$

Thus, the pruned graph does not require a graph-based cost term, but needs only a cost function based on intensity and shape features. $F$ includes model-based terms but can also include other model-independent, image-based terms, such as vesselness filters (see e.g. Frangi et. al. [ 15 ]).

**function** neighbor_nodes$\left(v_i = (x_i, y_i, z_i, s_i, p_i, l_i)\right)$

$N = \emptyset$

if $(l_i < \gamma_1(p_i))$

$N = N | v_j \ \forall \ v_j = \left(x_j, y_j, z_j, s_j, p_j, l_j\right)$ s.t.

$\left(x_j, y_j, z_j\right) \in \text{nbhd}(x_i, y_i, z_i), \ \ l_j = l_i + 1, \ s_j = s_i, \ \text{and} \ p_j = p_i$

endif

if $(l_i \geq \gamma_2(p_i))$

$N = N | v_j \ \forall \ v_j = \left(x_j, y_j, z_j, s_j, p_j, l_j\right)$ s.t.

$\left(x_j, y_j, z_j\right) \in \text{nbhd}(x_i, y_i, z_i), \ \ l_j = 1, \ \text{and} \ p_j = p_i + 1$

endif

return $N$

**Figure 40.** Pseudocode of the neighbor_nodes() function for the pruned graph

In this work, a directed vesselness filter is used, where the filter is aligned with the expected curve orientation given by the model. The un-weighted vesselness feature cost function for edge $e$ connecting node $v_i$ to $v_j$ is given by:

$$g_{s_j,p_j}(e_{i,j}) = e^{-\left|\left\{\nabla^2 G_{s_j,p_j} * I\right\}(x_j,y_j,z_j)\right|}, \qquad (35)$$

where $I$ is the image, $G_{s_j,p_j}$ is a cylindrical Gaussian with its central axis aligned with the expected curve orientation given by the model and with scale equal to the expected width of the structure given by the model. The term $\left|\left\{\nabla^2 G_{s_j,p_j} * I\right\}\left(x_j, y_j, z_j\right)\right|$ is large when there is a ridge or a valley oriented in the direction expected by the model, which makes the cost value of the whole equation smaller. Thus, this equation can be used as a directed cylinder detector, which should help to locate tubular structures of a given size and orientation.

The analytical form of the segmentation cost function associated with edge $e$ connecting node $v_i$ to $v_j$ is:

$$F(e_{i,j}) = \sum_{m=0}^{M-1} \omega_{s_j,p_j}^{m} \frac{g_{s_j,p_j}^{m}(e_{i,j})}{\sigma_{s_j,p_j}^{m}} + \sum_{k=0}^{K-1} \omega_{s_j,p_j}^{k+M} \frac{\left(x^k - f_{s_j,p_j}^{k}\right)^2}{\left(\sigma_{s_j,p_j}^{k}\right)^2}, \quad (36)$$

where $M$ and $K$ are the number of image-based and intensity-based features, respectively. The second term is the cost associated with the model-based features, while the first is associated with the image-based features. $f_{s_j,p_j}^{k}$ and $\sigma_{s_j,p_j}^{k}$ are the mean and standard deviation of the $k$th model-based feature associated with sample $s_j$ and locality $p_j$. $x^k$ is the value of that feature evaluated in the target image. $\sigma_{s_j,p_j}^{m}$ is the standard deviation of the $m$th image-based feature, and $g_{s_j,p_j}^{m}(e_{i,j})$ is the value of that feature in the target image. The $\omega$'s are the weighting values associated with each of those features, computed using Eqn ( 25 ). The standard deviations are used as normalizing factors to bring all the features into roughly the same scale. The standard deviations and weights for the image-based features are computed with the same approach used for the model-based features. With the graph and cost function defined, the algorithm can be used to segment new images.

## Implementation Details

$\gamma_1(p_i)$, which controls the maximum length of each locality, is set to the maximum length of $p_i$ seen in the training set. $\gamma_2(p_i)$, the minimum bound on the length locality, is sensitive to the length of the structure. For shorter SOIs, such as the chorda tympani and the carotid, which are about 20-30 voxels in length, $\gamma_2(p_i)$ can be set to the minimum length in the training set. For the longer SOIs, which are 45-65 voxels in length, the risk of short-cutting in the optimal path is greater so $\gamma_2(p_i)$ is set to the minimum length in the training set + 2. In future work, it may be

possible to automate the selection of this parameter, or to simply define $\gamma_2$ with a heuristic function that is dependent on the length of the SOI.

It will be assumed that if a point in the centerline of the minimal cost path is associated with exemplar *s* and locality *p*, then the SOI in the target image at that point should have the same width as *s* in locality *p*. This assumption is justified by the choice of width dependent feature measurements, i.e., when comparing features in a testing image to a particular exemplar, the lengths of the radii on which features are measured are dependent on the width of the structure in the particular locality of that exemplar; thus, a similar width is implied when features are best matched to that exemplar.

The voxel locations of the seed and end-nodes for each experiment were chosen manually. The sets of seed and end-nodes were then chosen to be $\{x_s, y_s, z_s, s_i, 0,0\}_i$, and $\{x_e, y_e, z_e, s_i, p_{\max}, l_j\}_{i,j}$. Thus, there is one seed node for each exemplar, and all the seeds are initialized to locality and length zero. There are end nodes for each exemplar and for all length values within the length bounds such that $\gamma_2(p_{\max}) \leq l_j \leq \gamma_1(p_{\max})$, and every end node is associated with the last locality. The path of minimal cost will connect one of these seeds to one of these end nodes.

## Validation

To validate the approach, the algorithm is tested using the pruned graph on the facial nerve (FN) and chorda (CH) in CT; the optic pathways (OP) in CT/MR; the carotid artery (CA) in CT; and the superior (SC), horizontal (HC), and posterior (PC) semicircular canals in CT. All of these structures pose a significant problem for segmentation algorithms, because they have intensity characteristics that vary along their length and borders that lack contrast. The FN and CH are

tiny nerves in the ear that range from 5 to 1 voxel in width. The structures lie in pneumatized bone, and surrounding structures have highly variable topology. The CA is large in comparison, and has partially contrasted structure borders in CT in the temporal bone region. The OPs are composed of the optic nerves, chiasm, and tracts. This structure suffers from lack of contrast in both MR and CT. In this study, one OP is defined as one optic nerve connecting to the contra-lateral tract through the chiasm. Previous studies have shown that localizing both OPs in this fashion and computing the union of the two can lead to an accurate segmentation of the whole optic system (see Chapter 9). The SC, HC, and PC, are tiny bone encapsulated canals in the ear that are horseshoe shaped. Thus, they are challenging for typical graph-based techniques because shortcutting may occur. A training set of 7 binary segmentations was acquired for the first four structures, and a set of 6 segmentations was acquired for the last three. The FN, CH, OP, CA, SC, HC, and PC were divided into 6, 4, 7, 5, 8, 4, and 6 localities. Voxel resolution for the FN, CH, SC, HC, and PC images was ~0.3x0.3x0.4 $mm^3$ and was ~1.0x1.0x1.5 $mm^3$ for the CA and OP images. Experiments were performed using a leave-one-out approach, i.e., the image being segmented is left out of the model. The results are validated by computing surface distance errors (see Chapter 4) between the manual and algorithmic segmentations.

## Results

Quantitative error between manual and algorithmic segmentations was measured using symmetric mean and max surface distances, which are shown in mm in Figure 41. The green bar, red bar, blue rectangle, and black I-bar indicate the median, mean, one standard deviation from the mean, and the range of the errors, respectively. Each dotted-line division in the graph indicates an error of approximately 1 voxel. Mean errors are seen to be sub-voxel, and the mean of the maximum errors are under 3 voxels for all SOIs.

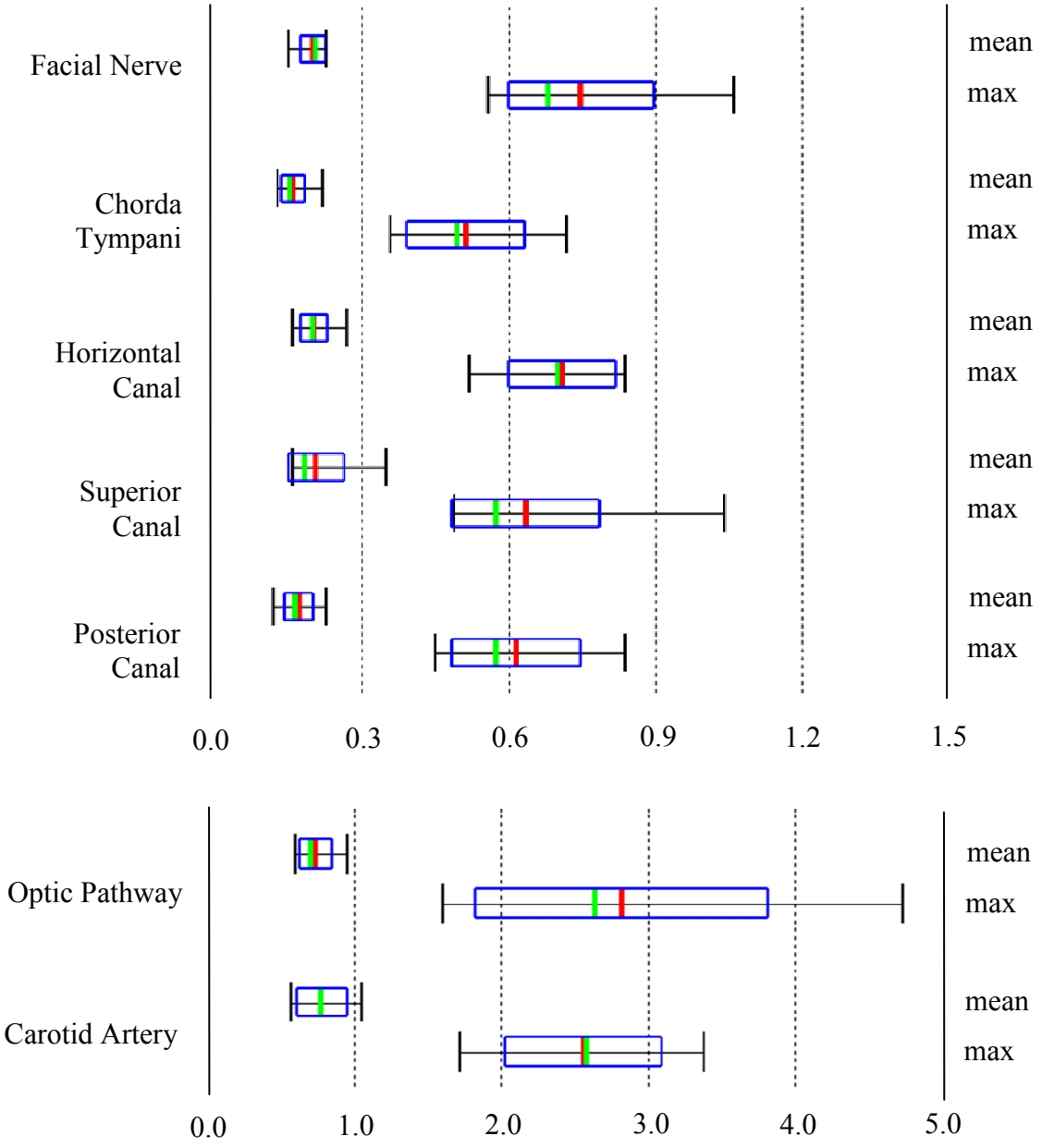


**Figure 41**. Quantitative segmentation results.

Qualitative results are shown in Figure 42 and Figure 43. For each SOI in Figure 42 (CH, FN, OP, and CA from top-to-bottom) and Figure 43 (HC, SC and LC from top-to-bottom), renderings of the results of the test case that results in the worst quantitative error are shown. The left and middle columns are 3D renderings of the manual and algorithmic segmentations, color-encoded with surface error distance in mm. The right column shows contours of the manual (red)

and algorithmic (green) segmentations in their respective images. To generate the 2D views, a thin-plate spline transformation was computed that warps the medial axes of the structures to lie in a plane, and then the image data and contours were passed through this transformation. This was done so that a cross section of the entire structure could be viewed in one 2D plane. As the example demonstrates, the primary location of error maxima for the FN occurs at the endpoints of the structure. This is due to errors in endpoint selection. Error maxima for the CA occur at the sharp corners of the structure, where, even though the path length is bounded, the path still may tend to try to shortcut in these regions within the allowable length bounds. Error maxima for the OP occurs at the corners of the chiasm. In this region some shortcutting occurs, and the anatomy is more rectangular in shape, which makes the tubular segmentation model less accurate. The worst errors of about 4 voxels, as demonstrated by the quantitative results and by the examples in Figure 42, are rare and highly localized. Error maxima for the semicircular canals tends to be at the structure endpoints due to errors in endpoint selection. The one exception is the superior canal shown in Figure 43. The SC of the individual in this image was significantly longer than the other 5 SCs in the training set, which were of similar length. Without a sample in the model that is similar to this structure, shortcutting occurred and the algorithm converged to an inaccurate solution with maximum surface errors of 1.04 mm. The second worse SC segmentation resulted in maximum surface error of only 0.75 mm. On average, results appear to be qualitatively accurate, and excellent agreement is seen between the manual and automatically generated contours.

## Discussion and Conclusions

In this work, a novel approach for tubular structure segmentation was presented. An optimal path algorithm is used to identify the SOI in a unique high-dimensional graph, which incorporates

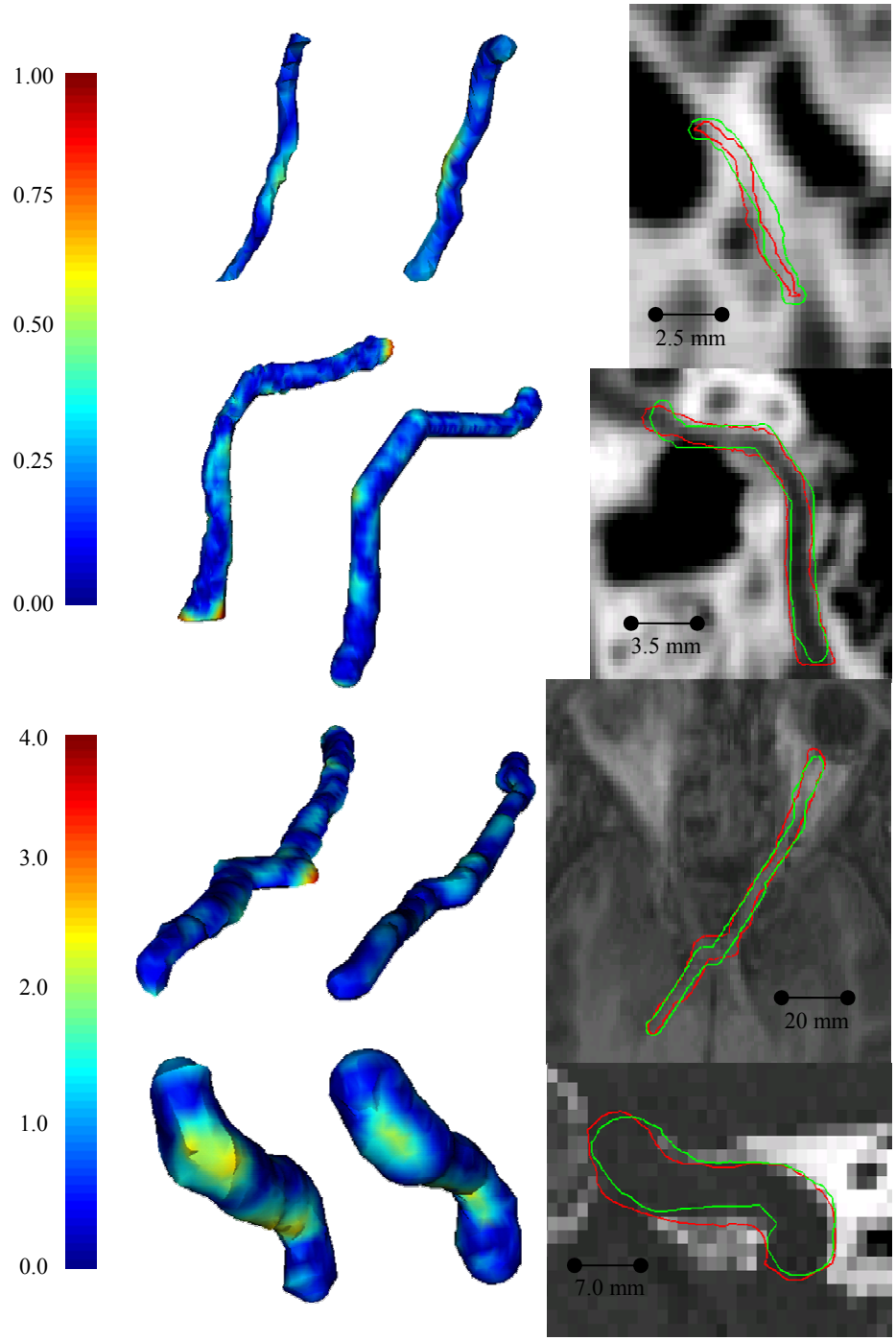


**Figure 42.** Qualitative segmentation results for the CH, FN, OP, and CA

information from a novel anatomy specific tubular structure model. The graph and structure model are used to overcome some limitations inherent to approaches typically used when

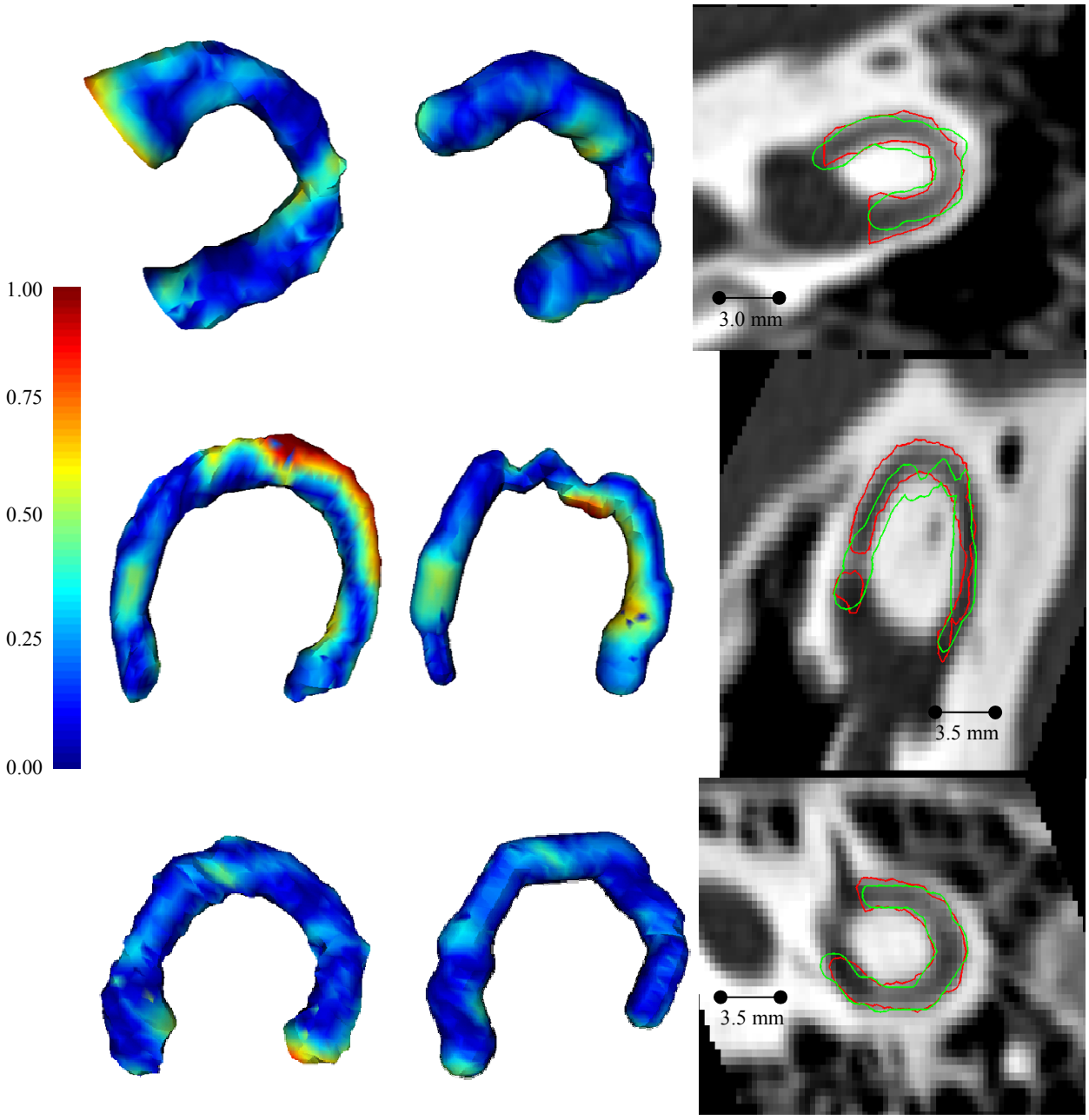


**Figure 43.** Qualitative segmentation results for the HC, SC, and PC

applying an optimal path finding algorithm. Similarly to the work of Li and Yezzi [ 29 ], the graphs presented herein allow a 3D representation of the surface to be recovered from the optimal curve. These graphs also allow local, a priori, intensity and shape information from SOI

exemplars to be included and optimized by the graph search. Finally, using *a priori* knowledge of the approximate length of the SOI, the graphs bound the allowable structure length to remove the bias towards straight curves inherent to traditional methods. While mild shortcutting may still occur, total failure is avoided.

An approach for training the novel structure model was also presented. Using a procedure to find correspondence across exemplars, detect unique structure subsections, and extract and train discriminant features, a model of the SOI is constructed that contains statistical values and importance weightings for a set of intensity and shape features for each subsection of each training exemplar. The information in this model is then incorporated into the segmentation framework. Combining the advantages achieved by the graph structure with the weighting of the features from the model used in the cost function accounts for all four limitations of traditional graph-based approaches listed in the introduction.

For validation, the approach was tested on seven challenging anatomical structures. As can be seen in the images of Figure 42 and Figure 43, there is poor image contrast at the borders of these structures and changing intensity profiles along the lengths of the structures. There are also substantial inter-patient variations in the intensity and shape features for each of these structures. All of these factors make the segmentation problem very difficult. However, the results of the presented approach are on average sub-voxel accurate, suggesting that it is generally applicable for fixed topology tubular structures. There are a few existing methods for the segmentation of the FN, CH, and OP. The segmentation results presented in this chapter are comparable to those of the most accurate of those existing approaches, which were specifically designed to localize these structures (see Chapters 8-10). This also indicates that the approach presented in this chapter is effective.

The method does require selecting two parameters for different SOIs: the number of localities of the SOI and the minimum length bound on localities. However, the selection of these parameters is intuitive, i.e., the number of localities should roughly correspond to the number of visually distinct regions of the SOI, and the minimum bound should be high enough to prevent shortcutting of the optimal path. In future work, these parameters will be chosen automatically. One possibility for choosing the number of localities is brute force optimization, i.e., try every possible number of localities between 1 and $N$ on the training set, and use the one that produces the most accurate results. Similarly for the locality minimum length bound, an optimization could start at the minimum length seen in the training set and be increased until sufficiently accurate results are achieved. Overall, excellent results are achieved with parameters that are very similar, suggesting that the choice of these parameters may not be highly dependent on the SOI.

One weakness of the technique is that the computational complexity drastically increases with the size of the structure. Because of this, relatively short SOIs were studied in this chapter. Thus, future work will also include exploring the application of this algorithm to longer structures.

# CHAPTER XII

# CONCLUSIONS

This dissertation presents studies on the development of algorithms for the automatic segmentation of the structures of the ear, as well as a new approach for modeling and segmentation of general tubular structures. Many otological procedures would benefit from a system that automatically identifies anatomical structures of the ear in CT. Conventional (registration-based) segmentation techniques are suitable for identifying the ossicles, external auditory canal, and cochlea, because these structures have high intensity contrast in CT. For other ear structures, conventional techniques are inadequate, and other segmentation techniques were developed. The scala tympani and scala vestibuli were segmented using an active shape model based technique. In Chapter 8, a minimal cost path-based algorithm was developed to segment the facial nerve and chorda tympani. These structures were also segmented using the general approach presented in Chapter 11, which can also be used to identify the carotid artery and the three semicircular canals. All segmentation approaches were validated using common quantitative and qualitative evaluation techniques, which suggest that the presented methods are clinically applicable. Future work towards the automatic segmentation of ear anatomy will include methods for identifying the sigmoid sinus and the tegmen, which were not included in this study. Automatic segmentation of these structures would be useful for assistance or guidance for various ear procedures.

There are six distinct tubular structures in the ear anatomy targeted in this dissertation, each of which has unique properties that make detection difficult. This led to the development of

a new algorithm for the segmentation of general tubular structures using novel, trained, anatomy specific models as presented in Chapter 11. The method was validated on these structures as well as on the optic pathways in MR/CT. The results of the algorithm demonstrated that it can be used to segment a variety of structures with a training set of exemplar segmentations and minor tuning of intuitive parameters, the selection of which could be automated in future work.

## BIBLIOGRAPHY


[ 1 ]. Arun, K. S., Huang, T. S., and Blostein, S. D., "Least square fitting of two 3-D point sets," *IEEE Trans. Patt. Anal. Machine Intell*. Vol. 9, no. 5, pp. 698-700, 1987.

[ 2 ]. A. Aschendorff, J. Kromeier, T. Klenzner, and R. Laszig, "Quality Control After Insertion of the Nucleus Contour and Contour Advance Electrode in Adults", *Ear & Hearing*, 28, 75S-79S, April 2007.

[ 3 ]. Aylward S.R., Bullitt E., Initialization, Noise, Singularities, and Scale in Height Ridge Traversal for Tubular Object Centerline Extraction, *IEEE Transactions on Medical Imaging*, 2002.

[ 4 ]. Bekes, G, Máté, E, Nyúl, L., Kuba, A., Fidrich, M., 2008. Geometrical model-based segmentation of the organs of sight on CT images, *Med. Phys.* 35, 735.

[ 5 ]. Bouix, S., Siddiqi, K., Tannenbaum, A., 2005. Flux driven automatic centerline extraction. *Medical Image Analysis* 9, 209-221.

[ 6 ]. de Bruijne, M., van Ginneken, B., Viergever, M.A., Niessen, W.J., 2003b. "Adapting active shape models for 3D segmentation of tubular structures in medical images," Proc. Inf. Process. Med. Imaging, pp. 136–147.

[ 7 ]. Cootes, TF, Cooper, D, Taylor CJ, and Graham, J, "Active Shape Models - Their Training and Application." *Computer Vision and Image Understanding.* Vol. 61, No. 1, Jan. 1995, pp. 38-59.

[ 8 ]. D'Haese, P., Duay, V., Li, R., du Bois d'Aische, A., Merchant T., Cmelak A., Donnelly E., Niermann K., Macq B., Dawant B., 2003. Automatic segmentation of brain structures for radiation therapy planning Proceedings of SPIE: Medical Imaging 2003: Image Processing. 5032, 517-526.

[ 9 ]. Dawant, B., Hartmann, S. L., Thirion, J. P., Maes, F., Vandermeulen, D., Demaerel, P., 1999. Automatic 3D Segmentation of Internal Structures of the Head in MR Images Using a Combination of Similarity and Free Form Transformations: Part I, Methodology and Validation on Normal Subjects. IEEE Trans. Med. Imaging 18, 909-916.

[ 10 ]. Dice, L. R., 1945. Measures of the amount of ecologic association between species. Ecology. 26, 297–302.

[ 11 ]. Dijkstra, E. W., 1959. A note on two problems in connexion with graphs. Numerische Mathematik 1, 269–271.

[ 12 ]. Dinggang, S., Davatzikos, C., 2002. HAMMER: hierarchical attribute matching mechanism for elastic registration. IEEE Trans. Med. Imag. 21, 1421-1439.

[ 13 ]. Feng, J., Ip, H. H. S., Cheng, S. H., 2004. A 3D Geometric Deformable Model for Tubular Structure Segmentation. MMM 10, 174-180.

[ 14 ]. Fischl, B., Salat, D., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., and Dale, A., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33, 341–355.

[ 15 ]. Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A., 1998. “Multiscale vessel enhancement filtering,” Proc. Med. Image Comput. Assist. Interv., vol. 1496, pp. 130–137.

[ 16 ]. Frangi, A.F., Rueckert, D., Schnabel, J.A., Niessen, W.J., “Automatic construction of multiple-object three-dimensional statistical shape models: Application to cardiac modeling," IEEE Trans on Med. Imag., Vol. 21, No. 9, pp. 1151-66, 2002.

[ 17 ]. Gensheimer, M., Cmelak, A., Niermann, K., Dawant, B., 2007. Automatic delineation of the optic nerves and chiasm on CT images. Proceedings of SPIE: Medical Imaging 2007: Image Processing. 6512, 41.

[ 18 ]. B. F. Green, “The orthogonal approximation of an oblique structure in factor analysis,” Psychometrika, vol. 17, pp. 429-440, 1952.

[ 19 ]. Green JD Jr, Marion MS, Erickson BJ, et al. Three-dimensional Reconstruction of the temporal bone. *Laryngoscope* 1990; 100: 1–4.

[ 20 ]. Hanssen, N., Burgielski, Z., Jansen, T., Lievin, M., Ritter, L., von Rymon-Lipinski, B., Keeve, E., 2004. Nerves - level sets for interactive 3D segmentation of nerve channels. IEEE International Symposium on Biomedical Imaging: Nano to Macro 1, 201-204.

[ 21 ]. Isambert, A., Dhermain, F., Bidault, F., Commowick, O., Bondiau, P., Malandain, G., Lefkopoulos, D., 2008. Evaluation of an atlas-based automatic segmentation software for the delineation of brain organs at risk in a radiation therapy clinical context. Radiother. Oncol. 87 (1), 93-9.

[ 22 ]. C. James, K. Albegger, and R. Battmer, et. al, “Preservation of residual hearing with cochlear implantation: How and why,” Acta Oto-Laryngologica, 125:5, pp. 481-91, 2005.

[ 23 ]. Jun BC, Song SW, Cho JE, et al. Three-dimensional reconstruction based on images fromspiral high-resolution computed tomography of the temporal bone: anatomy and clinical application. *Laryngology & Otology* 2005;119:693–8.

[ 24 ]. Kass, A. Witkin, D. Terzopoulos, “Snakes: Active Contour Models,” *Int’l Jour. Of Computer Vision*, 321-331, 1988.

[ 25 ]. Labadie, R.F., Balachandran, R., Mitchell, J., Noble, J.H., Majdani, O., Haynes, D.S., Bennett, M., Dawant, B.M., Fitzpatrick, J.M., 2010, “Clinical Validation Study of Percutaneous Cochlear Access Using Patient Customized Micro-Stereotactic Frames,” Otology & Neurotology, 31(1):94-99.

[ 26 ]. Labadie, R.F., Noble, J.H., Dawant, B.M., Balachandran, R., Majdani, O., Fitzpatrick, J.M., 2008, “Clinical validation of percutaneous cochlear implant surgery: initial report,” Laryngoscope, 118:1031-9.

[ 27 ]. E. Lehnhardt and R. Laszig, “Specific surgical aspects of cochlear implant-soft surgery,” Adv. in coch. Imp., pp. 228-9, 1994.

[ 28 ]. Lesage, D., Angelini, E., Bloch, I., Funka-Lea, G., 2009. A review of 3D vessel lumen segmentation techniques: Models, features, and extraction schemes. Medical Image Analysis 13, 819-845.

[ 29 ]. Li, H., Yezzi, A., 2007. Vessels as 4D Curves: Global Minimal 4D Paths to Extract 3D Tubular Surfaces and Centerlines. IEEE Trans. Med. Imag. 26 (9), 1213-1223.

[ 30 ]. Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P., 1997. Multimodality image registration by maximization of mutual information. IEEE Trans Med Imag. 16, 187-198.

[ 31 ]. Maks CB, Butson CR, Walter BL, Vitek JL, McIntyre CC. Deep brain stimulation activation volumes and their association with neurophysiological mapping and therapeutic outcomes, J Neurol Neurosurg Psychiatry, (In Press), 2008.

[ 32 ]. Manniesing, R., Viergever, M. A., Niessen, W. J., 2007. Vessel Axis Tracking Using Topology Constrained Surface Evolution. IEEE Trans. Med. Imag. 26, 309-316.

[ 33 ]. Nakashima S, Sando I, Tkahashi H, et al. Computer-aided 3-D reconstruction and measurement of the facial canal and facial nerve. I. Cross-sectional area and diameter: preliminary report. *Laryngoscope* 1993; 103: 1150–6.

[ 34 ]. Noble, J.H., Datteri, R., Dawant, B.M., D'Haese, P.F., 2010, "Automatic segmentation of the optic tracts for computer assistance of deep brain stimulation procedures," 2010 IEEE Intl. Symp. on Biomedical Imaging: From Nano To Macro, pp. 864 – 867.

[ 35 ]. Noble, J.H., Dawant, B.M., 2009, "Automatic segmentation of the optic nerves and chiasm in CT and MR using the atlas-navigated optimal medial axis and deformable model algorithm," Proceedings of the SPIE Conf. on Med. Imag., 7259: 725916.

[ 36 ]. Noble, J.H., Dawant, B.M., Warren, F.M., Labadie, R.F., 2009, "Automatic Identification and 3D Rendering of Temporal Bone Anatomy," Otol Neurotol., 30(4):436-42.

[ 37 ]. Noble, J.H., Majdani, O., Labadie, R.F., Dawant, B.M., Fitzpatrick, J.M., 2010, "Automatic Determination of Optimal Linear Drilling Trajectories for Cochlear Access Accounting for Drill-Positioning Error," Intl. J. of Med. Robotics and Comp. Assist. Surg., 6(3):281-290.

[ 38 ]. Noble, J.H., Rutherford, R., Labadie, R.F., Majdani, O., Dawant, B.M., 2010, "Modeling and segmentation of intra-cochlear anatomy in conventional CT," Proc. of the SPIE conf.on Med. Imag., 7623: 762302.

[ 39 ]. Noble, J., Warren, F.M., Labadie, R.F., Dawant, B.M., 2008. Automatic segmentation of the facial nerve and chorda tympani in CT images using spatially dependent feature values. Medical Physics 35 (12), 5375-5384.

[ 40 ]. Olabarriaga, S. D., Breeuwer, M., Niessen, W.J., 2003. Minimum cost path algorithm for coronary artery central axis tracking in CT data. Lecture Notes in Comp. Sci.-MICCAI 2879, 687-694.

[ 41 ]. Srivatsan Pallavaram, Benoit M. Dawant, Michael S. Remple, Joseph S. Neimat, Chris Kao, Peter E. Konrad, Pierre-François D'Haese, "Effect of brain shift on the creation of

functional atlases for deep brain stimulation surgery," Int J CARS, Springer, DOI 10.1007/s11548-009-0391-1, 2009.

[ 42 ]. Press, W. H., Flannery, B. P., Teukolsky S. A., Vetterling, W. T., 1992. Numerical Recipes in C, 2nd ed. Cambridge University Press, Cambridge, U. K., pp. 412-419.

[ 43 ]. Rohde, G. K., Aldroubi, A., Dawant, B. M., 2003. The adaptive bases algorithm for intensity-based nonrigid image registration. IEEE Trans. Med. Imag. 22, 1470-1479.

[ 44 ]. Rohlfing, T., Maurer Jr., C. R., 2003. Nonrigid image registration in shared-memory multiprocessor environments with application to brains, breasts, and bees. IEEE Trans. on Info. Tech. in Biomed. 7, 16-25.

[ 45 ]. Rueckert, D., Sonoda, L., Hayes, C., Hill, D. L. G., Leach, M. O., Hawkes, D. J., 1998. Nonrigid Registration Using Free-Form Deformations: Application to Breast MR Images. IEEE Trans. Med. Imag. 18, 712-721.

[ 46 ]. Santamaría-Pang, A., Colbert, C. M., Saggau, P., Kakadiaris, I. A., 2007. Automatic centerline extraction of irregular tubular structures using probability volumes from multiphoton imaging. Lecture Notes in Comp. Sci.-MICCAI 10, 486-494.

[ 47 ]. Sethian, J., Level Set Methods and Fast Marching Methods, 2nd ed. Cambridge University Press, Cambridge, U. K.

[ 48 ]. M.W. Skinner, T.A. Holden, B.R. Whiting, et. al, “In vivo estimates of the position of advanced bionics electrode arrays in the human cochlea,” Ann Otol Rhinol Laryngol Suppl., 197:2-24, April 2007.

[ 49 ]. Schrader, B., W. Hamel, D. Weinert, H. M Mehdorn. 2002. Documentation of electrode localization. Movement Disorders 17:suppl 3:S167:S174.

[ 50 ]. Sensi M et. al., Pallidal stimulation for segmental dystonia: long term follow up of 11 consecutive patients. Mov Disord. 2009 Sep 15; 24(12):1829-35

[ 51 ]. Studholme, C., Hill, D. L. G., Hawkes, D. J., 1999. An overlap invariant entropy measure of 3D medical image alignment. Pattern Recognition. 32 (1), 71-86.

[ 52 ]. Symon L, Momma F, Schwerdtfeger K, et al: Evoked potential monitoring in neurosurgical practice. Adv Tech Stand Neurosurg 14:25 70, 1986

[ 53 ]. Takagi A, Sando I, Takahashi H. Computer-aided three-dimensional reconstruction and measurement of semicircular canals and their cristae in man. *Acta Otolaryngol* 1989;107: 362–5.

[ 54 ]. Wang H, Merchant SN, Sorensen MS. A Downloadable Three-Dimensional Virtual Model of the Visible Ear. *ORL* 2007;69:63–7.

[ 55 ]. Wanna GB, Balachandran R, Majdani O, Mitchell J, Labadie RF. Percutaneous Access of the Petrous Apex In Vitro Using Customized Micro-stereotactic Frames based on Image Guided Surgical Technology. *Acta Oto-Laryngologica* 2009 Aug; 25:1-6.

[ 56 ]. Warfield, S., Zou, K., Wells, W., 2004. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans Med Imaging 23 (7), 903-921.

[ 57 ]. Wells III, W. M., Viola, P., Atsumi, H., Nakajima, S., Kikinis, R., 1996. Multi-modal Volume Registration by Maximization of Mutual Information. Medical Image Analysis 1, 35-51.

[ 58 ]. Wesarg, S., Firle, E. A., 2004. Segmentation of vessels: The corkscrew algorithm. Progress in Biomedical Optics and Imaging - Proceedings of SPIE 5370, 1609-1620.

[ 59 ]. West J, Fitzpatrick JM. "Comparison and evaluation of retrospective intermodality brain image registration techniques." *J Comput Assist Tomogr.* 1997 Jul-Aug;21(4):554-66.

[ 60 ]. Wu, Z., "Multivariate compactly supported positive definite radial functions," *Adv. Comput. Math.* 4, 283–292, 1995.

[ 61 ]. Yim, P. J., Cebral, J. J., Mullick, R., Marcos, H. B., Choyke, P. L., 2001. Vessel surface reconstruction with a tubular deformable model. IEEE Trans. Med. Imag. 20, 1411-1421.

[ 62 ]. Zijdenbos, A. P., Dawant, B. M., Margolin, R., 1994. Morphometric Analysis of White Matter Lesions in MR Images: Method and Validation. IEEE Transactions on Medical Imaging. 13 (4), 716-724.