

Atlas-based Segmentation of Temporal Bone Anatomy

THESIS

Presented in Partial Fulfillment of the Requirements for the Degree Master of Science in
the Graduate School of The Ohio State University

By

Tong Liang

Graduate Program in Electrical and Computer Engineering

The Ohio State University

2017

Master's Examination Committee:

Dr. Bradley Clymer, Advisor

Dr. Kimerly Powell, Co-advisor

Copyrighted by

Tong Liang

2017

Abstract

In this research, we proposed an atlas-based automated segmentation approach which used 6 manually traced CT volumes (including the reference CT volume) to construct the atlas. We then registered the test CT volumes to a single reference volume. A series of morphological processes were implemented to segment the temporal bone anatomy. A 3D-view based tracing tool ‘VolEditor’ was used to provide the manual tracing results of the test data to compare with our automatic segmentations. The validation results of our method on 20 test clinical CT volumes (10 left, 10 right) resulted in average DICE similarity coefficients over 0.6 for cochlea, malleus and incus and ranged from 0.46 to 0.64 for facial nerve, semi-circular canals and vestibule.

The proposed method didn’t reach the precision required for surgical planning but was effective for segmenting structures required in surgical simulation software. Additionally, we determined that the 3D manual tracing tool, used in this study, resulted in segmentation errors that caused degradation of the validation results.

Acknowledgments

I would like to acknowledge and thank my thesis committee member Dr. Clymer for his introduction of this valuable research opportunity to me. Otherwise, I wouldn't be able to gain such important research experience during my stay at OSU.

I would like to acknowledge and thank my thesis committee member Dr. Powell for her patient guidance and suggestions in this work, her support was valuable in every step of my thesis research.

Words can't express my deep gratitude towards my parents Tiangang Liang and Caixia Wu. Without their support, I would not have the chance to study abroad in the OSU.

Vita

March 2012No.1 Middle School of Lanzhou, China

June 2015B.Eng. Electrical and Electronic

Engineering, University of Electronic Science and Technology of China

September 2015 to presentM.S. Electrical and Computer Engineering,

The Ohio State University

Fields of Study

Major Field: Electrical and Computer Engineering

Table of Contents

Abstract	ii
Acknowledgments	iii
Vita	iv
List of Tables	viii
List of Figures	ix
Chapter 1 Introduction	1
1.1 Problem Statement	1
1.2 Related Work.....	2
1.3 Motivations and Contributions.....	5
1.4 Organization of Thesis	5
Chapter 2 Data and Methods	6
2.1 Data Information	6
2.2 Manual Segmentation.....	7
2.3 Work Flow of the Proposed Atlas-based Approach.....	9
2.3.1 Registration Process	9
2.3.2 Image Segmentation of Temporal Bone Anatomic Structures.....	14
2.4 Introduction of Validation Metrics.....	17

2.4.1 DICE.....	18
2.4.2 Volumetric Similarity	18
2.4.3 Hausdorff Distance	19
2.4.4 Average Hausdorff Distance	19
2.4.5 Mahalanobis Distance.....	20
2.4.6 Principal Component Axes Similarity	21
2.5 Comparison between the Automatic Segmentation with a Baseline Approach	22
2.6 Evaluation of Manual Tracing Results Generated by VolEditor	23
Chapter 3 Results	24
3.1 Validation Results	24
3.2 Comparison between the Automatic Segmentation and a Baseline Approach	28
3.3 Evaluation of Manual Tracing Results Generated by VolEditor	32
3.3.1 Comparing the 3D-view Manual Segmentation Results with Otsu’s Automatic Segmentation of Bone Mask.....	32
3.3.2 Comparing the 3D-view Manual Segmentation Results Traced by VolEditor with 2D-view Manual Segmentation Results	35
Chapter 4 Discussion and Conclusion	37
4.1 Analysis of Validation Results	37
4.2 A Comparison between Automatic and the Baseline Approach	39

4.3 The 3D-view Based Manual Tracing Method.....	40
4.4 Conclusion and Suggestions for Future Work	41
References	42

List of Tables

Table 1 Manual adjustment records of left volumes.....	10
Table 2 Manual adjustment records of right volumes	11
Table 3 The confusion matrix	18
Table 4 Validation results of left volumes	26
Table 5 Validation results of right volumes.....	27
Table 6 Principal component axes similarity of left volumes.....	28
Table 7 Principal component axes similarity of right volumes	28
Table 8 Non-overlapping percentage.....	34

List of Figures

Figure 1 Example of Coordinate system adopted by ImageJ	7
Figure 2 Examples of original image (left) and transferred image (right).....	8
Figure 3 Example of 3D-view based segmentation results.....	8
Figure 4 Work flow of registration process	9
Figure 5 Example of manual intensity adjustment with ImageJ.....	12
Figure 6 Work flow of segmentation process	14
Figure 7 Work flow of clean operations and segmentation of bony surface	14
Figure 8 Example of principal component axes of cochlea	21
Figure 9 Example of orientation angles	21
Figure 10 DICE.....	29
Figure 11 Volumetric Similarity.....	29
Figure 12 Hausdorff Distance	30
Figure 13 Average Hausdorff Distance	30
Figure 14 Mahalanobis Distance	31
Figure 15 Example of voxels in the manual segmentation that do not overlap with the bone mask	33
Figure 16 Example of voxels in the manual segmentation that do not overlap with the bone mask	33

Figure 17 Comparison between 2D-view based segmentation and 3D-view based
segmentation 35

Chapter 1 Introduction

1.1 Problem Statement

Image segmentation is a technique applied to segment the input image into regions with specific properties and extract those regions out for further processing. It is often the first step of image analysis. Through image segmentation, the region of interest (ROI) is labeled and its representation is simplified for further study. Common methods of image segmentation are based on the pixel values' intensity difference between different regions. These methods work well when dealing with segmentation problems where different regions' pixel values differ significantly from each other. But, when it comes to the segmentation need of temporal bone anatomy, the pixel intensities of different bony structures are very close to each other. Obviously, this makes the intensity-based image segmentation unsuitable for this particular situation. Thus, we propose an atlas-based image segmentation approach for the segmentation of temporal bone anatomy.

Atlas-based image segmentation is a type of method which does not depend on the pixel intensity difference between different regions to segment the image. This method segments the image into different regions of interest based on a user-defined atlas. The atlas serves as a source of information of the regions of interest that need to be segmented from the images. In our approach, this atlas is developed by combining

several manual segmentations of temporal bone anatomy from a set of selected “normal” reference images. Then image registration techniques are applied to align the atlas with the CT images of unknown subjects. Ideally, the manually segmented anatomic structures of the atlas will overlap with structures in the unknown CT images as much as possible.

The segmentation results obtained from atlas based segmentation of temporal bone are mainly used to enhance 3D visualization and surgical simulation in virtual environment.

1.2 Related Work

In previous 3D visualization studies of the middle and inner ear, manual segmentation methods [7] were used that took many hours to perform. In a semi-automated volume growing segmentation approach, proposed by M. Seemann et al. [8], a threshold interval density value was selected for each structure. An experienced investigator was needed to manually mark one voxel per slice that belonged to the structure of interest. Then the adjacent voxels whose intensity values were in the defined threshold interval would be segmented as part of the structure. If neighboring voxels were segmented that did not belong to this structure, they were manually removed by the investigator. This semi-automatic approach indeed shortened the segmentation process of middle and inner ear to about 40 minutes. But it still required a considerable amount of work from an experienced investigator. Segmentation times on the order of one hour to one day were observed by Chan et al. [2] using the commercially available software package Amira to prepare data for their surgical simulator. A further review of publicly

available software packages by Hassan et al. [9] resulted in the observation that one segmentation approach was inadequate for all anatomical structures of interest and would require an experienced user 1-2 hours' effort to prepare the data for patient specific surgical simulation. Our approach currently takes less than five minutes to perform and could be further optimized for real-time applications.

A temporal bone surgical simulation using Voxelman [1] used a semi-automated global segmentation for 3D reconstruction. The authors found that the time consumption is significant (on average 20 min) when uploading a new CT volume to the surgical simulator due to the segmentation process. The case rehearsal in Arora et al.'s research included ossicular chain surgery, cochlear implantation and congenital anomalies. They observed that rehearsals of surgical applications that involved facial nerve (e.g. the cholesteatoma surgery) were not possible due to lack of delineation of soft tissue. Our segmentation method delineates the facial nerve channel and incorporates the location of the facial nerve.

Higher resolution images available from microCT imaging systems provide better delineation of fine structures such as the stapes and corda tympana. However, this is at the expense of increased image size. Lee et al. [3] developed a volume visualization and haptic interface system for measurement of 3D structures.

In the field of automatic segmentation of medical images, atlas-based methods are well accepted and widely used. This method often served as the 1st step of automatic segmentation to localize the anatomic structures in unknown subjects. Noble et al. [4] applied atlas-based registration to automatically segment the following anatomic

structures: labyrinth, ossicles, and external auditory canal. This method registered the atlas image of a “normal” subject with the unknown subject’s image, and superimposed the manually labeled contour of the atlas onto the unknown subject. An active contour approach was then applied to acquire the final delineation of structures. The advantage of this atlas-based registration is that it is easy and fast to apply. But this simple method relies on the assumption that the anatomic structures of the manually picked “normal” subject are topologically similar to the structures of the unknown subjects. Due to variations in the region of pneumatized bone surrounding the facial nerve and chorda tympani, this assumption does not hold. Therefore, in the segmentation of these two structures, simple atlas-based registration only serves as the first step of segmentation. This step roughly localizes the anatomic structures and further algorithms were proposed to segment the facial nerve and chorda tympani.

As is known, facial nerve channel and chorda tympani are tubular structures, and Nobel et al. [5] proposed a model-based spatially-variant method which specifically targets tubular structures in pneumatized bone. In this approach, a set of training volumes was used to create the models of facial nerve channel and chorda tympani. The advantage of this method is that the minimum cost path algorithm, adopted in this approach, compensated for small registration errors in the atlas projection procedure. And the corresponding cost function, which was adopted to find the minimum path, was incorporated with spatial information of the structures of interest to compensate for the variation of intensity along the structures. This effectively improved the segmentation results of pneumatized tubular structures. The limitation of this algorithm is that it

assumes the location distribution of tubular structures of the model are similar to that of unknown CT volumes. Therefore, the training data set should be large enough to capture the variation in location for this method to work well.

1.3 Motivations and Contributions

The motivation for this research is the need for automatic segmentation of temporal bone anatomy for surgical simulations and case specific pre-surgical planning. Anatomic structures like the facial nerve and chorda tympani are crucial structures that a surgeon must avoid when performing surgery such as during a cochlear implant [6]. Thus, it is very important to segment these structures in the preoperative CT scans for surgical planning.

1.4 Organization of Thesis

The rest of this thesis is organized in the following manner. In chapter 2, we provide the work flow that we proposed for the segmentation approach followed by an introduction of the validation metrics that we adopted to evaluate the performance of our method. In chapter 3, we provide validation results of our approach. In chapter 4, we discuss the performance of our approach based on the experimental results and potential improvements in future work.

Chapter 2 Data and Methods

2.1 Data Information

A total of 27 bones (14 left, 13 right) obtained from the OSU body Donation program and the University of Cincinnati were used for the study. Six of the bones (3 left, 3 right) were used to develop the automated registration parameters and the region-of-interest (ROI) masks used for the automated segmentation. One was used as the reference for registration. The remaining bones were used for testing the atlas-based segmentation approach.

Clinical X-ray computed tomography (CT) images of the bones were acquired at 140 kVp and exposure times of 1000 mA s using a Siemens 64-bit detector Somatom Sensation™ (Siemens Healthcare GmbH, Erlangen Germany). Axial slices were collected over a 119-mm field-of-view (FOV) with an in-plane voxel size of 0.232 mm and slice thickness of 0.4 mm. The images were oriented so that all the volumes are displayed starting from superior to inferior and anatomically from left to right. A single left image was randomly chosen to be used as the reference image (4532L) for registration. This image was flipped horizontally for spatial registration of the right bone images.

A demonstration of the image coordinate system adopted in ImageJ is shown in figure 1. The image coordinate system mentioned in the following sections refer to this coordinate system.

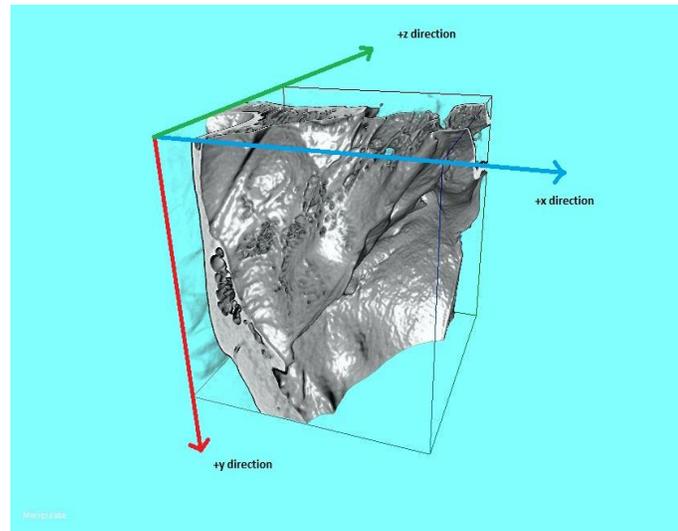


Figure 1 Example of Coordinate system adopted by ImageJ

All the manual segmentations were performed using VolEditor software that was developed in-house (Ohio Supercomputer Center, Columbus Ohio). The structures were manually segmented in a 3D view, and then verified by an experienced observer in a 3D view. Figure 1 is an example of such a 3D view provided by VolEditor.

2.2 Manual Segmentation

In the manual segmentation process, a nonlinear transfer function of gray levels is applied to adjust the 3D visualization results before manual tracing. The curve of this transfer function is manually selected. Figure 2 shows an example of the processed result.

The left image volume is the original image and the right image is the result after application of the transfer function. The corresponding transfer function for each volume is shown on the bottom left corner of the images in figure 2. After setting the transfer function, the original image volume is enhanced such that the image volume should contain bony structures only. This enhanced image is used for manual segmentation of bony structures using a painting tool with region size of 3 x 3 x 3. Figure 3 shows an example of manual segmentation result in 3D view.

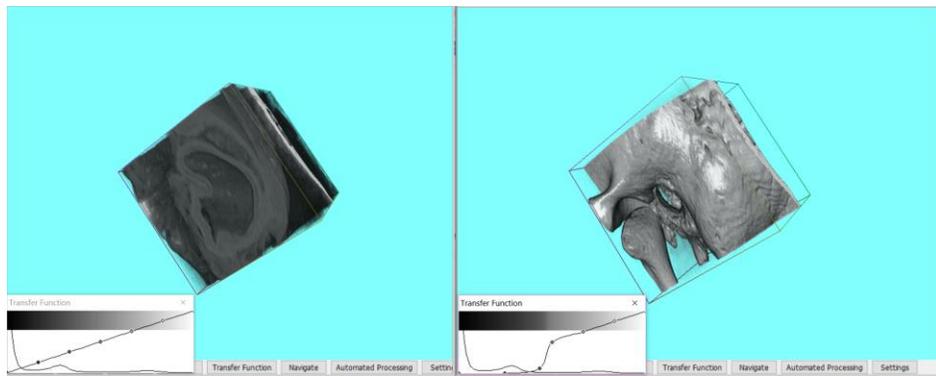


Figure 2 Examples of original image (left) and transferred image (right)

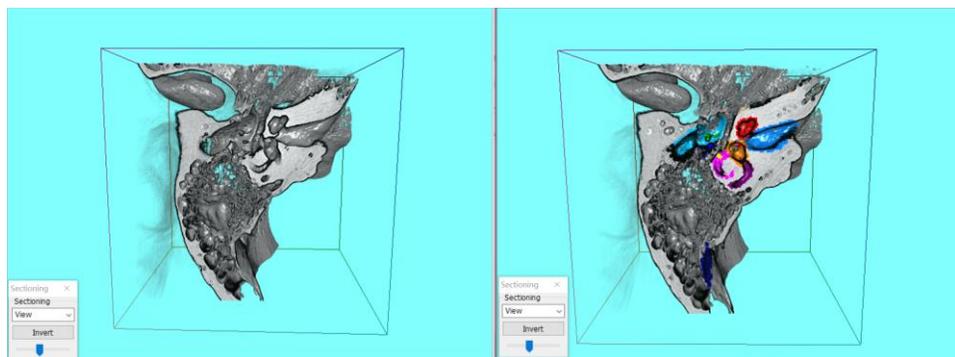


Figure 3 Example of 3D-view based segmentation results

2.3 Work Flow of the Proposed Atlas-based Approach

In our method, the segmentation procedure consisted of two major parts. The 1st part is the registration process of CT volumes and the 2nd part is the segmentation process. In the following sections, these two parts are explained in detail.

2.3.1 Registration Process

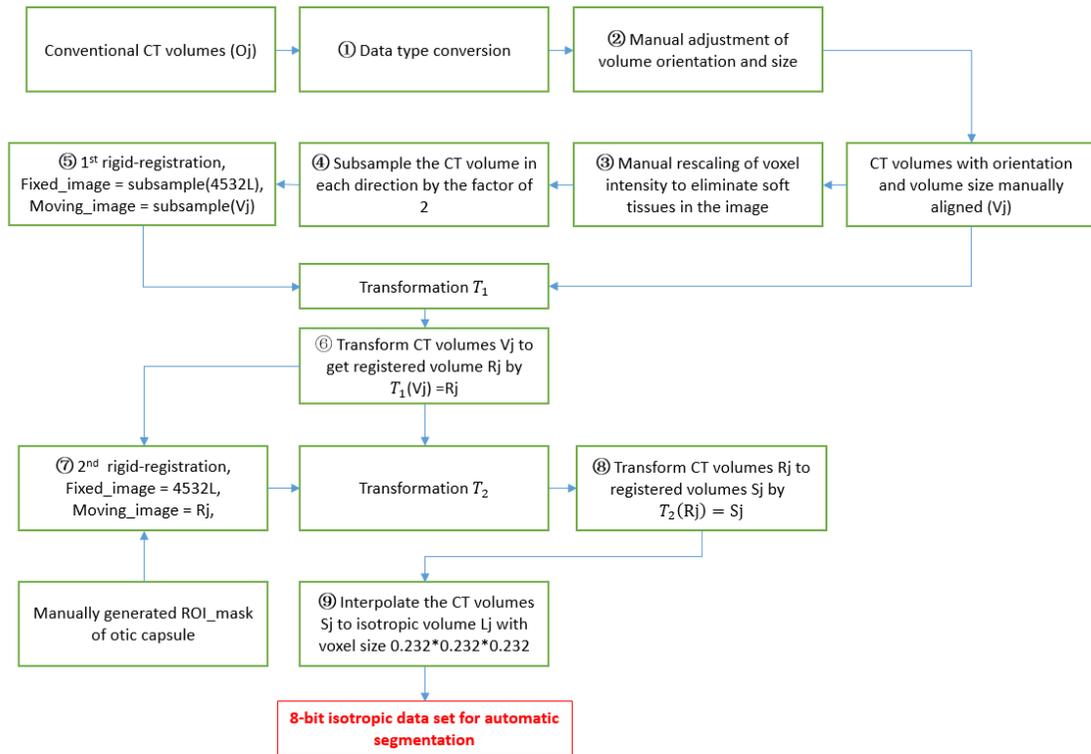


Figure 4 Work flow of registration process

As shown in figure 4, the registration process consists of 9 steps:

In step ①, the clinical CT image voxel data type is converted from 16-bit to 8-bit to save computational resource.

In step ②, the training and test CT volumes are manually adjusted to roughly align their orientation with that of the atlas volume, and all the CT volumes size are either cropped or padded to the same size as the reference volume (4532L). These adjustments are made through ImageJ [16], an image processing program developed at the National Institutes of Health. The resulting volume from this step is referred as V_j . The manual orientation and volume size adjustments of the test data set are recorded in Tables 1 and 2.

File name	Flip along y axis	Flip along z axis	Crop/Pad slices	Original number of Slices
4532L (reference)				273
08CL_3 (atlas)		X	3	270
08DL_3 (atlas)	X		59	214
08FL_2 (atlas)	X	X	-25	298
08HL_3	X			273
08ML		X	3	270
08QL		X	29	244
6183L	X		-23	296
6270L	X		52	221
BYU960L1	X		69	204
BYU960L2	X		86	187
RDR974L	X		84	189
rmf328L	X		102	171
ZCO867L	X		74	199

Table 1 Manual adjustment records of left volumes

File name	Flip along y axis	Flip along z axis	Crop/Pad slices	Original number of slices
4532L.flipx (reference)				273
9413R (atlas)	X			273
UMO609R (atlas)	X		66	207
UTQ235R (atlas)	X		55	218
08AR_3	X			273
08BR_3		X	3	270
08ER		X	-25	298
08GR		X	3	270
08JR_3		X	3	270
2517R	X			273
3203R	X			273
3689R	X			273
5687R	X		1	272
8871R	X			273

Table 2 Manual adjustment records of right volumes

In step ③, a copy of each volume V_j is rescaled manually with ImageJ to suppress the visible soft tissue voxel values to 0s. The subjective standard in this linear rescaling is to keep malleus and incus visible in the CT image and suppress the soft tissue as much as possible. The linear intensity adjustment is given by the following ramp function:

$$T(x) = \begin{cases} 0, & \text{if } x \leq T_1 \\ \frac{255}{T_2 - T_1} (x - T_2) + 255, & \text{if } T_1 < x < T_2 \\ 255, & \text{if } x \geq T_2 \end{cases}$$

Figure 5 shows an example of manual intensity adjustment result.

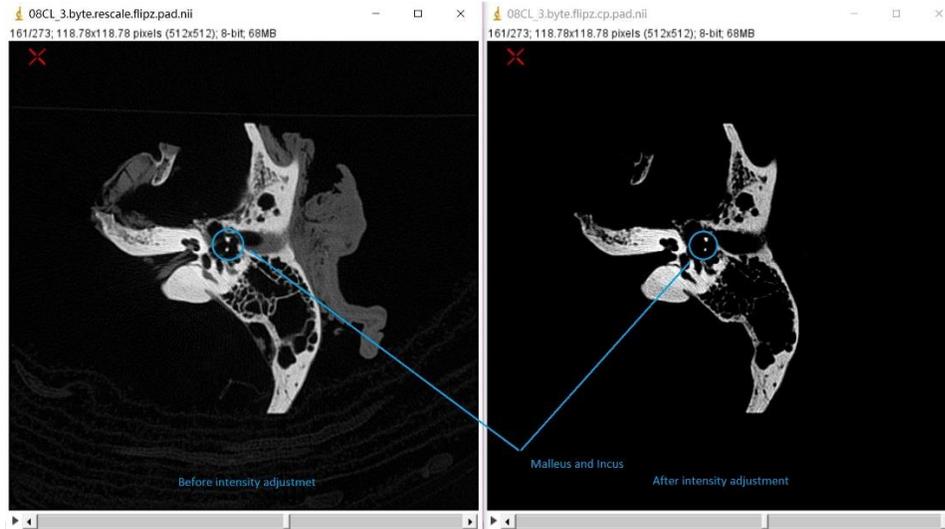


Figure 5 Example of manual intensity adjustment with ImageJ

In step ④, the rescaled image volumes are subsampled in all x, y, and z directions as output to the next step for the 1st coarse rigid body registration.

In step ⑤, as shown in figure 4, the subsampled reference volume is set as the fixed image and the other subsampled volumes are set as the moving images in the 1st registration procedure. The result of this step is the transformation T_1 from the moving image to the fixed image.

In step ⑥, the original size volumes V_j are transformed with T_1 to get the coarsely registered volumes referred as R_j . Then R_j is output to the following step for refined registration.

In step ⑦, R_j is registered to the reference volume with a manually made ROI mask of cylinder shape centering around the otic capsule of the reference image. This 2nd

refined image registration aims to register the otic capsules of the reference and the other volumes to a very close space for the following segmentation.

In step ⑧, the transformation T_2 acquired from last step is used to obtain the refine-registered volumes S_j .

Both registrations in step ⑤ and ⑦ are rigid body registration implemented by elastix [10] with mutual information (MI) as the metric. This method maximizes the mutual information [11] between the two CT volumes A and B, where MI is given in the following equation:

$$MI = H(A) + H(B) - H(A, B),$$

where $H(\cdot)$ denotes the Shannon entropy of one image and $H(\cdot, \cdot)$ denotes the joint Shannon entropy between two images.

In step ⑨, the volumes S_j with voxel size $0.232*0.232*0.4 \text{ mm}^3$ are interpolated to isotropic volume L_j with voxel size $0.232*0.232*0.232 \text{ mm}^3$. After this step, the 8-bit isotropic volumes are sent to the next processing stage for automatic image segmentation.

2.3.2 Image Segmentation of Temporal Bone Anatomic Structures

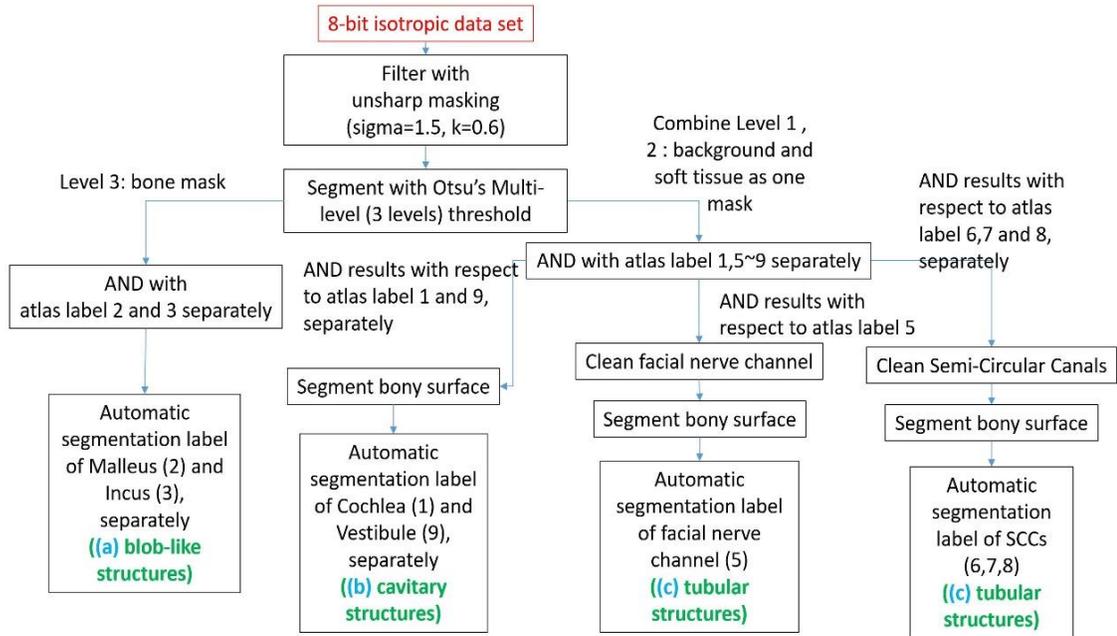


Figure 6 Work flow of segmentation process

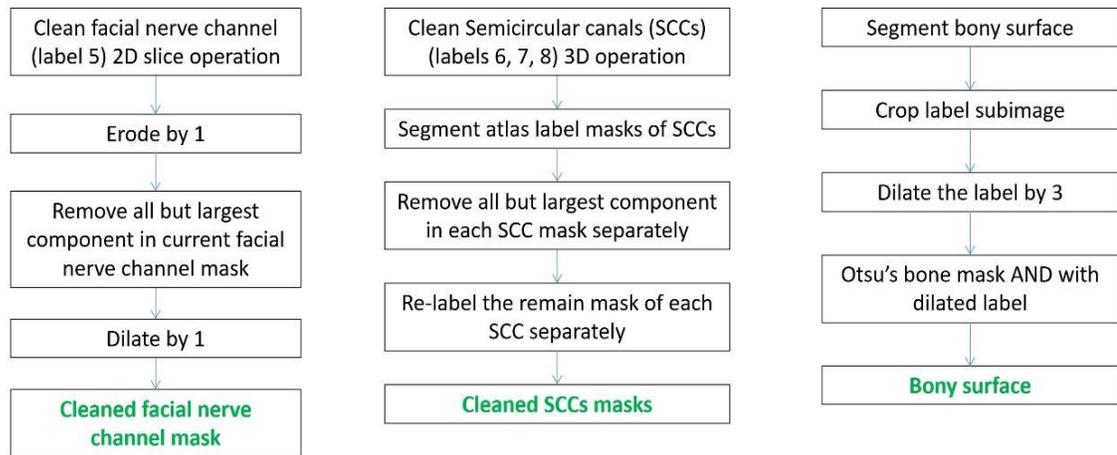


Figure 7 Work flow of clean operations and segmentation of bony surface

As can be seen in figure 6, the anatomic structures studied in this work are categorized as 3 groups (a), (b), and (c). All the image volumes go through the unsharp mask filter first. The unsharp mask filtering is defined by following equation:

$$g(x, y) = f(x, y) + k * (f(x, y) - \bar{f}(x, y)),$$

Where $k = 0.6$, $\bar{f}(x, y)$ is the CT slice image $f(x, y)$ blurred by a 2D Gaussian filter with $\sigma = 1.5$. Because $k < 1$, this filtering is actual de-emphasizing the edges of the image volumes. A three-level Otsu' Multi-level threshold [15] is then applied, where level 1 refers to background, level 2 refers to soft tissue, and level 3 is the bone. The next section introduces the different segmentation approaches used for blob-like, cavitory, and tubular structures:

a) Blob-like structures: Malleus and Incus

The bone level of Otsu's threshold is logically ANDed with the atlas label mask of Malleus (2) and Incus (3) separately.

b) Cavitory structures: Cochlea and Vestibule

The background level and soft tissue level of Otsu's threshold are combined as a single mask to logically AND with the atlas label mask of Cochlea (1) and Vestibule (9) separately. The ANDed results are sent to segment the bony surface surrounding the structures separately. The output label mask of corresponding bony surface is then label as Cochlea and Vestibule respectively.

c) Tubular structures:

Though, facial nerve channel and semicircular canals (SCCs) are both categorized as tubular structures, their topological characteristics are quite different. Therefore, a slightly different cleaning operation of floaters (tiny bone structures) near the tubular structures are adopted to remove floaters of facial nerve channel and SCCs after the separate Boolean logic AND operations between the combined Otsu's mask and the atlas label masks (5, 6, 7 and 8). Figure 7 shows the details of the cleaning operations and segmentation of the bony surface. As shown in figure 6, the results of cleaning operations are sent to segment bony surface. And the output results of segment bony surface operation are labeled as the corresponding automatic segmentation labels.

i. Facial nerve channel

For tubular structures like facial nerve channel, a morphological erode with size 1 per each slice is applied to the image volume, then the connected components are calculated slice-wise to remove all but largest component. After removing all the small floaters, another morphological dilate with size 1 per each slide is applied to the image volume. The remaining mask was labeled as facial nerve channel for segmentation of bony surface. The output bony surface mask is then labeled as the automatic segmentation of facial nerve channel.

ii. Semicircular canals

For cleaning of SCCs, we simply calculated the 3D connected components and the largest component is identified as the corresponding SCC

mask. Then this mask is sent to segment bony surface. The results surface mask is labeled as the corresponding automatic segmentation label.

iii. Segment bony surface

As shown in figures 6 and 7, the intermediate automatic segmentation labels are passed to segment the bony surface surrounding the structure. In this process, the whole image volume is cropped into a sub-volume that centered at the segmentation mask. This cropping process merely aims to reduce computational load. The previous automatic segmentation labels are then dilated by 3 and ANDed with the Otsu's bone mask. The resulting labels are regarded as the corresponding bony surface structures.

2.4 Introduction of Validation Metrics

Validation of segmentation results serves as a reference in the process of developing and evaluating a new method of image segmentation. According to the validation results, adjustment of the approach is made to improve the segmentation accuracy. In our experiment, multiple validation metrics were adopted to evaluate the accuracy of the proposed segmentation approach.

In the following section, six different metrics are introduced, DICE coefficient, volume similarity, Hausdorff Distance, average Hausdorff distance, Mahalanobis Distance, and principal component axes similarity. All of these metrics except principal component axes similarity were calculated using Taha et al.'s [12] quantitative evaluation tool. The principal component axes similarity was calculated with MATLAB.

2.4.1 DICE

The DICE coefficient was first proposed by LR. Dice [13] in 1945. It is considered a spatial overlap-based similarity measurement. The concept of confusion matrix is adopted for overlap-based metric. For a specific bone structure with predefined structure label \mathbf{X} , if the segmentation label of the voxel is \mathbf{X} , then it is regarded as a positive case. Otherwise, it is regarded as a negative case. The corresponding 4 scenarios of a confusion matrix are shown in table 3.

	Auto segmentation positive	Auto segmentation negative
Manual segmentation positive	TP	FN
Manual segmentation negative	FP	TN

Table 3 The confusion matrix

According to Dice LR, the DICE coefficient is calculated through the following equation:

$$DICE = \frac{2 * TP}{2 * TP + FP + FN}$$

2.4.2 Volumetric Similarity

Volume similarity (VS) metric is also defined with the 4 basic scenarios introduced previously. The following equation were used to calculate the volume similarity between auto and manual segmentations:

$$VS = 1 - \frac{|FN - FP|}{2 * TP + FP + FN} = 1 - \frac{||V_{auto}| - |V_{manual}||}{V_{auto} + V_{manual}}$$

Where V_{auto} represents the total volume of the auto segmentation mask and V_{manual} represents the total volume of the manual segmentation mask. Even though the 4

cardinalities are used in the above equation to calculate the VS, this is not an overlap-based metric. As shown in the equation, the total volume of the auto segmentation is compared with the total volume of the manual segmentation, the overlap volume between auto and manual segmentation has no effect on this metric.

2.4.3 Hausdorff Distance

Hausdorff Distance (HD) is commonly used as a validation metric of image segmentation which describes the dissimilarity between segmentations. It measures the spatial distance between 2 segmentations. The directed Hausdorff Distance between 2 sets of points A and B is given by the following equation:

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

where $\|a - b\|$ is some type of norm, i.e. Frobenius norm. And Hausdorff distance is defined by the following equation:

$$HD(A, B) = \max(h(A, B), h(B, A))$$

This metric is sensitive to outliers, as can be seen from the equation of HD. Thus, we also evaluated the Average Hausdorff Distance (AVD) as one of our validation metrics.

2.4.4 Average Hausdorff Distance

The Average Hausdorff Distance is the Hausdorff Distance averaged over all points in sets A and B. AVD is more stable and less sensitive to outliers. AVD is given by the following equation:

$$AVD(A, B) = \max(d(A, B), d(B, A)) ,$$

where $d(A,B)$ is the directed AVD defined by following equation:

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

2.4.5 Mahalanobis Distance

The original Mahalanobis Distance (MHD) is in contrast to the Euclidean distance. For 2 points x and y included in a point cloud K , MHD takes the information of the point cloud K into account when calculating the Mahalanobis distance. MHD is given by the following equation:

$$MHD(x, y) = \sqrt{(x - y)^T S^{-1} (x - y)}$$

Where S is the covariance matrix of the point set K . For our validation, a variant of MHD is adopted. According to G.J. McLachlan [14], MHD is modified to calculate the distance of the means of 2 point sets, A and B . It is given by the following equation:

$$MHD(A, B) = \sqrt{(\mu_A - \mu_B)^T S_{AB}^{-1} (\mu_A - \mu_B)}$$

Where, μ_A and μ_B are the means of point set A and B respectively. S_{AB} is the common covariance matrix of point sets A and B . It is defined by the following equation:

$$S_{AB} = \frac{n_A S_A + n_B S_B}{n_A + n_B}$$

Where, n_A and n_B are the number of points in set A and B respectively. S_A and S_B are the covariance matrix of set A and B respectively.

2.4.6 Principal Component Axes Similarity

The following figure 8 and 9 show examples of the cochlea's principal component axes and the orientation angles Alpha, Beta and Gamma between the principal axes of manual and automatic segmentation results of the cochlea, respectively.

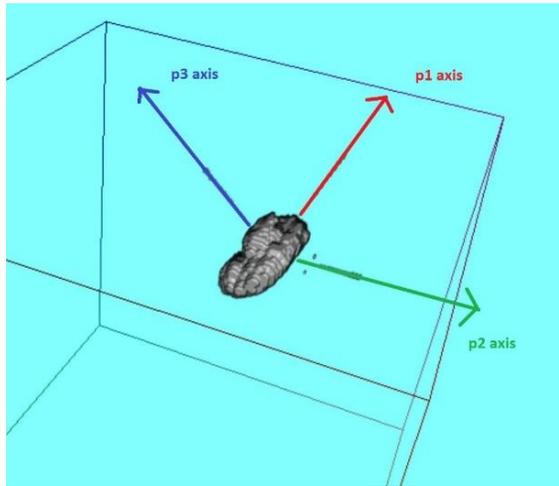


Figure 8 Example of principal component axes of cochlea

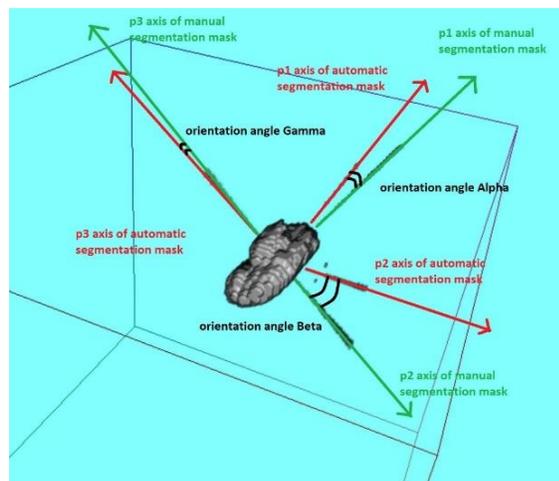


Figure 9 Example of orientation angles

The principal component axes similarity measures the difference in orientation angles between the principal component axes as shown figure 9. The smaller the difference in orientation angles, the more similar the two point clouds of automatic and manual segmentation labels are. The principal component axes are computed in the following manner. First, the coordinates of a structure's manual (or automatic) segmentation labels (x, y, z) are stacked into a matrix X as column vectors. Then the column vectors in X are treated as sample vectors. By applying eigenvalue decomposition of the corresponding sample covariance matrix S computed from X , we can obtain the principal axes (eigenvectors of S) of the segmentation mask. The principal axes (p_1 , p_2 , and p_3) can be used to visualize the shape and orientation difference between the manual and auto-segmentation masks. Figure 9 shows 2 set of principal axes, one set of axes are calculated from manual segmentation mask (green arrows) and the other set of axes are calculated from automatic segmentation mask (red arrows). The limitation of this validation method is that if the segmentation labels' locations vary in a way that is close to isotropic manner (e.g. the segmented structure has a spherical-like shape or cylinder-like shape), then the principal axes are not helpful, because the labels location distribution varies similarly in every direction.

2.5 Comparison between the Automatic Segmentation with a Baseline Approach

The proposed automatic segmentation approach was compared with a baseline approach which segmented the temporal bone anatomy by simply using the registered atlas as a segmentation mask and ANDing it with the bone mask of the test image without

any additional image processing. The comparison results were used to evaluate the performance of the additional image processing techniques adopted in the proposed automatic segmentation.

2.6 Evaluation of Manual Tracing Results Generated by VolEditor

The 3D-view based manual tracing results generated by VolEditor were evaluated by comparing them to the corresponding test image's bone mask calculated by Otsu's multi-level threshold method. Besides, comparison between 3D-view based tracing results of VolEditor and 2D-view based tracing results were also conducted through visual inspection to evaluate the manual segmentation results used in this work.

Chapter 3 Results

3.1 Validation Results

The median, mean, variance, minimum and maximum validation results of DICE, Volume Similarity (VS), Hausdorff Distance, Average Hausdorff Distance and Mahalanobis Distance calculated between automatic and manual segmentations of the test set are shown in Tables 4 (left) and 5 (right). The mean and standard deviation of orientation angles between principal component axes (principal component axes similarity) of automatic and manual segmentation labels for each anatomic structure are shown in Tables 6 (left) and 7 (right).

According to the validation results in Tables 4 and 5, the average DICE coefficients of cochlea, malleus, and incus were over 0.6; the average DICE coefficients of the facial nerve channel and the semi-circular canals ranged from 0.4 and 0.6. The lowest DICE coefficients were observed for the facial nerve channel, lateral SCC, and posterior SCC, of which, the minimum DICE coefficient dropped to 0.3015. The maximum variance of the DICE coefficient was observed to be 0.01 in the facial nerve channel.

The mean VS coefficients of all the anatomic structures were greater than 0.68 for each structure. The VS of facial nerve channel and SCCs were on average greater than

0.8. The variance of the VS coefficient for each structure was below 0.03. The lowest VS coefficient was observed in the incus, where it dropped to 0.4655.

The HD coefficients in the validation results are generally very large. This result shows that HD is highly sensitive to outliers in the manual segmentation and is therefore less useful to validate the automatic segmentation result. The average of directed Hausdorff Distance (AVD) coefficient for all structures was less than 1.9 on average, and the maximum AVD observed was 3.3858 in the facial nerve channel. The largest variance was 0.6455 in the facial nerve channel. The MHD coefficients were on average below 0.7 with a maximum variance 0.0951 observed for the lateral SCC. The maximum MHD was also observed for the lateral SCC.

According to Tables 6 and 7, most of the average orientation angles are less than 20 degrees, except for malleus of right temporal bone, the posterior SCC and vestibule of both sides of the temporal bone. The maximum average orientation angle observed was 50.2798 degrees in the alpha orientation angle of left temporal bones' vestibules. The minimum average orientation angle was observed for the malleus of left temporal bones with 4.0532 degrees. Average Orientation angles less than 10 degrees were observed in all structures except the vestibule.

Structure names	Cochlea	Malleus	Incus	Facial Nerve Channel	Lateral SCC	Posterior SCC	Superior SCC	Vestibule
DICE								
median	0.6220	0.6717	0.6261	0.4418	0.5272	0.5084	0.5882	0.5193
mean	0.6241	0.6706	0.6305	0.4623	0.5319	0.5273	0.5734	0.5245
variance	0.002984	0.002335	0.007581	0.010571	0.007044	0.004297	0.005941	0.004994
min	0.5531	0.6044	0.5074	0.3191	0.3750	0.4615	0.4642	0.4328
max	0.7184	0.7542	0.7656	0.6620	0.6397	0.6296	0.6865	0.6181
Volumetric Similarity (VS)								
median	0.8237	0.7372	0.7333	0.9111	0.8513	0.8807	0.9430	0.9143
mean	0.8315	0.7579	0.7249	0.8972	0.8656	0.8750	0.8799	0.8623
variance	0.006619	0.006890	0.013809	0.003815	0.009449	0.004851	0.011098	0.016693
min	0.7076	0.6385	0.5817	0.8179	0.6537	0.7360	0.7031	0.5580
max	0.9721	0.9025	0.9543	0.9824	0.9899	0.9839	0.9714	0.9906
Hausdorff Distance (HD)								
median	9.3372	4.7775	5.8735	11.9370	8.1542	15.2440	7.9671	11.4017
mean	10.6360	5.7198	8.2222	47.1780	8.6098	26.1294	8.1241	11.5677
variance	17.860954	7.759544	51.326455	5388.590953	3.523423	1511.726767	4.887193	2.320594
min	7.0000	3.1623	3.0000	7.6158	6.4031	7.3485	5.3852	9.0000
max	21.5639	12.0000	27.7308	189.7103	11.8743	136.3635	11.6619	14.2478
Average Hausdorff Distance (AVD)								
median	0.5757	0.4099	0.5046	1.3625	0.8316	1.2671	0.7296	1.0430
mean	0.6352	0.3965	0.5206	1.4253	0.9191	1.3214	0.7279	1.0973
variance	0.042687	0.003661	0.041357	0.509493	0.106978	0.216415	0.032523	0.054059
min	0.3962	0.2928	0.2474	0.6056	0.5512	0.7219	0.5125	0.8085
max	1.1208	0.4688	0.9175	3.0173	1.6338	2.0543	1.0924	1.6170
Mahalanobis Distance (MHD)								
median	0.2305	0.3924	0.3581	0.3176	0.5721	0.5853	0.4111	0.5957
mean	0.2163	0.3879	0.3867	0.3381	0.6150	0.5441	0.4862	0.5503
variance	0.008742	0.041375	0.055207	0.029253	0.051015	0.046448	0.059524	0.043584
min	0.0847	0.1336	0.0480	0.1194	0.2150	0.1221	0.1725	0.3037
max	0.3607	0.6860	0.7715	0.6687	0.9967	0.9119	0.9248	0.9505

Table 4 Validation results of left volumes

structure names	Cochlea	Malleus	Incus	Facial Nerve Channel	Lateral SCC	Posterior SCC	Superior SCC	Vestibule
DICE								
median	0.6500	0.6579	0.6415	0.5481	0.5846	0.4665	0.6833	0.5906
mean	0.6366	0.6425	0.6185	0.5327	0.5632	0.4937	0.6466	0.5834
variance	0.004988	0.005720	0.008620	0.008807	0.007841	0.009588	0.008857	0.001811
min	0.4688	0.4762	0.4249	0.3015	0.3975	0.3383	0.4297	0.4941
max	0.7074	0.7537	0.7033	0.6184	0.6581	0.6564	0.7308	0.6362
Volumetric Similarity (VS)								
median	0.7728	0.7768	0.7141	0.8902	0.9306	0.8250	0.9662	0.8046
mean	0.7708	0.7953	0.6865	0.9095	0.9236	0.8243	0.9245	0.8425
variance	0.008539	0.005148	0.011441	0.003824	0.002883	0.026250	0.005948	0.006310
min	0.5496	0.6863	0.4655	0.8223	0.8446	0.6254	0.8058	0.7513
max	0.8749	0.9172	0.8145	0.9710	0.9937	0.9987	0.9931	0.9802
Hausdorff Distance (HD)								
median	9.4575	6.4803	7.4407	14.4821	9.5647	18.1489	8.4243	11.5199
mean	10.5075	7.1628	10.3872	32.5746	10.8170	17.3776	26.0297	11.3360
variance	20.323998	5.549219	56.340699	3187.441712	28.214925	9.799476	3142.060770	3.438791
min	5.6569	4.8990	4.1231	9.2195	6.3246	11.3578	6.0828	9.0000
max	19.8997	12.1655	26.4386	192.6707	24.9199	21.7486	185.4993	14.5258
Average Hausdorff Distance (AVD)								
median	0.5345	0.4786	0.4667	1.1171	0.8559	1.9134	0.5380	0.8377
mean	0.5733	0.5242	0.6438	1.2773	0.9753	1.8117	0.6758	0.8960
variance	0.025578	0.034043	0.329539	0.645467	0.178917	0.493983	0.199838	0.038299
min	0.4152	0.2935	0.3310	0.6363	0.5540	0.8743	0.3579	0.6598
max	0.9317	0.9253	2.2437	3.3858	1.6739	3.1511	1.8859	1.2915
Mahalanobis Distance (MHD)								
median	0.1866	0.4081	0.3183	0.3707	0.6148	0.8017	0.4014	0.5182
mean	0.2180	0.5665	0.3612	0.3841	0.6633	0.6802	0.4100	0.5283
variance	0.017133	0.085860	0.046600	0.065502	0.095080	0.079429	0.073307	0.033325
min	0.0528	0.2820	0.0761	0.1364	0.2574	0.1806	0.0964	0.1861
max	0.4905	1.1533	0.7968	0.9986	1.2021	0.9816	1.0800	0.8044

Table 5 Validation results of right volumes

structure (left) name	cochlea	malleus	incus	facial	lateral SCC	posterior SCC	superior SCC	vestibule
average orientation angle Alpha (degree)	12.4906	4.0532	10.5758	8.9118	10.7997	18.8296	9.4813	50.2799
average orientation angle Beta (degree)	15.9852	13.0105	11.7177	12.4356	13.325	21.5653	10.583	49.6207
average orientation angle Gamma (degree)	8.0674	12.8299	9.5381	8.6139	9.003	8.7779	5.6063	15.9019
std of Alpha (degree)	11.992	3.1357	8.6183	8.3182	8.8299	10.3002	4.7293	28.2669
std of Beta (degree)	11.2922	14.9956	7.174	7.2097	7.8705	7.9106	3.7586	26.9239
std of Gamma (degree)	5.4656	14.9294	8.2298	4.2939	6.7094	4.931	2.4113	7.115

Table 6 Principal component axes similarity of left volumes

structure (right) name	cochlea	malleus	incus	facial	lateral SCC	posterior SCC	superior SCC	vestibule
average orientation angle Alpha (degree)	11.8601	5.4828	15.9585	6.9481	13.4827	27.9108	9.4417	47.1377
average orientation angle Beta (degree)	13.1881	28.5659	16.166	16.4267	19.4198	31.8256	9.9658	46.7699
average orientation angle Gamma (degree)	5.1755	27.8883	5.2139	14.0731	14.9407	14.3628	6.5955	12.7075
std of Alpha (degree)	6.6456	2.8919	16.7866	3.7906	5.5976	22.657	6.0603	24.6299
std of Beta (degree)	5.7034	25.1977	16.9356	18.4886	10.4262	20.2581	6.2906	24.868
std of Gamma (degree)	3.287	25.4347	3.2114	19.48856	11.9239	7.629	3.1853	4.542

Table 7 Principal component axes similarity of right volumes

3.2 Comparison between the Automatic Segmentation and a Baseline Approach

Figures 10 to 14 show a comparison between average validation coefficients calculated from the segmentation results generated by the proposed automatic segmentation with additional image processing, and another set of segmentation results generated by the baseline approach with no extra image processing.

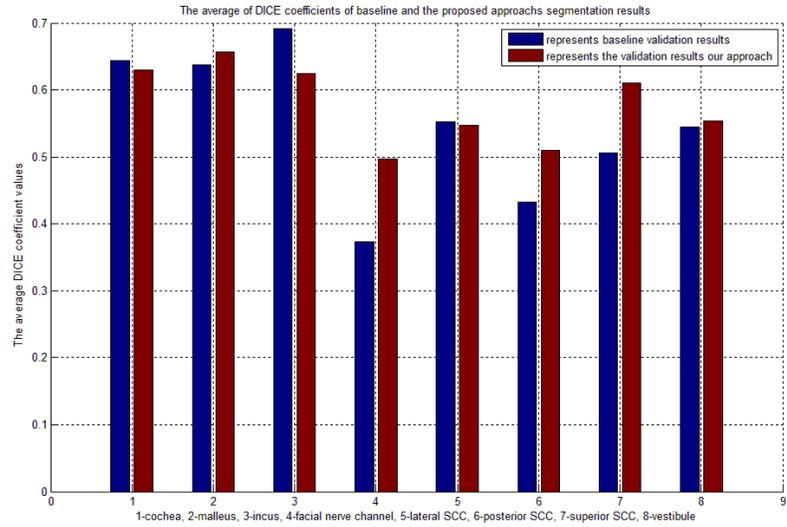


Figure 10 DICE

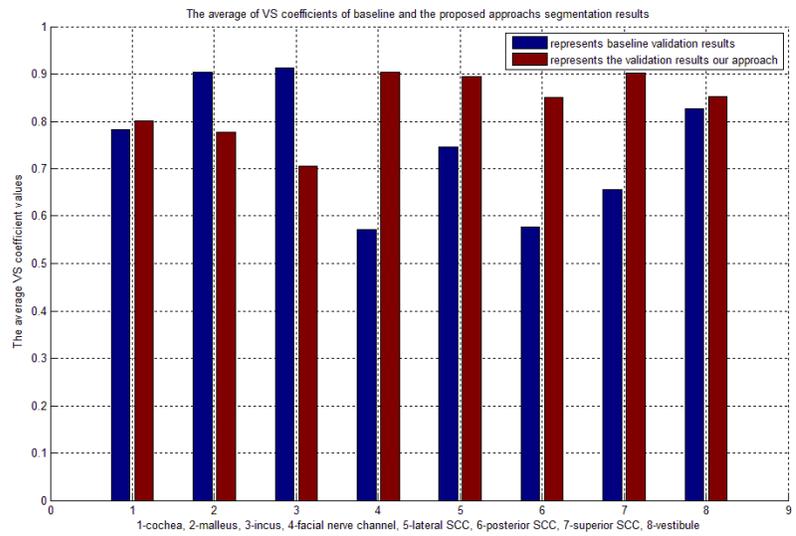


Figure 11 Volumetric Similarity

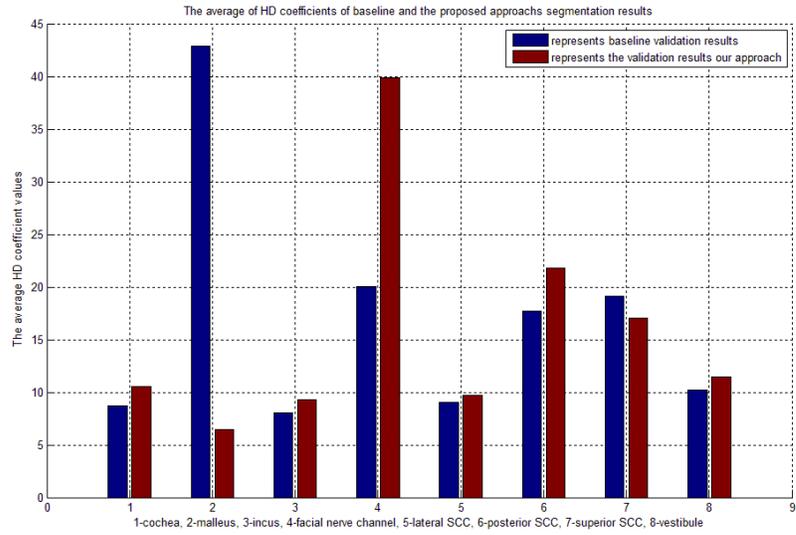


Figure 12 Hausdorff Distance

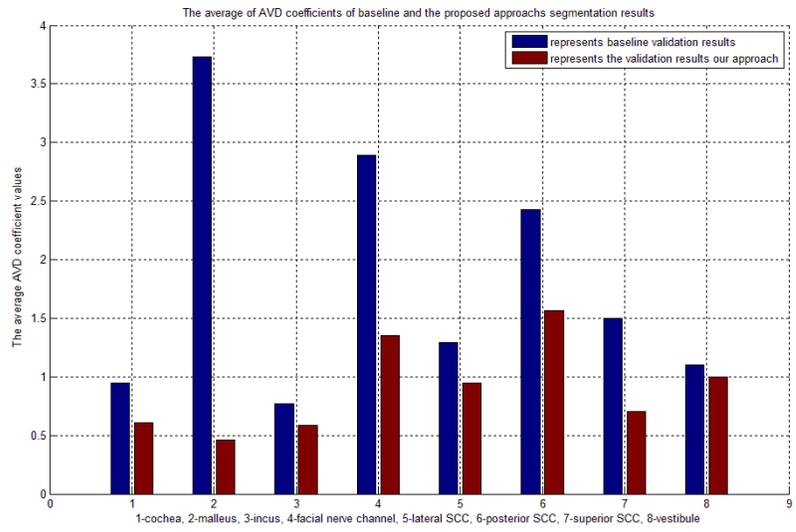


Figure 13 Average Hausdorff Distance

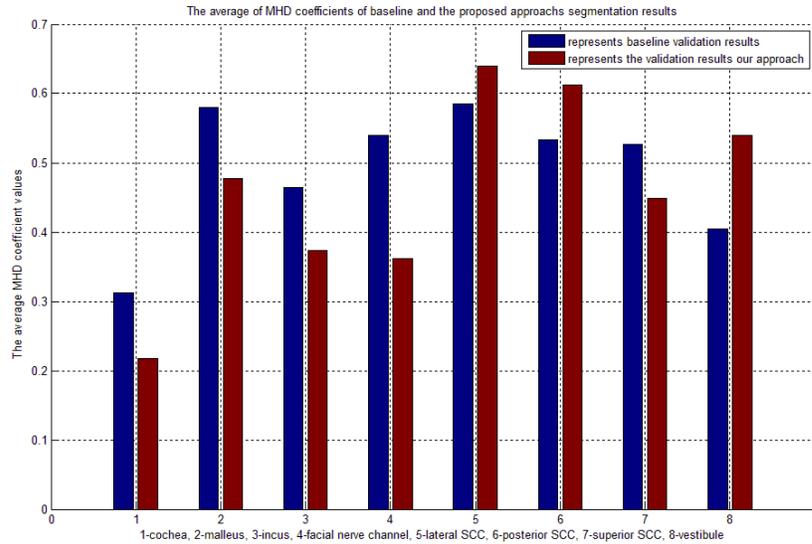


Figure 14 Mahalanobis Distance

The mean DICE coefficients of the automatic method were greater than the baseline values for incus, facial nerve channel, posterior SCC, superior SCC and vestibule. For the cochlea and lateral SCC, the DICE coefficients of the baseline method were greater than the automatic method's DICE coefficient. For the incus, the baseline's DICE coefficient was greater than the automatic method's.

The average VS coefficients of malleus and incus of baseline method are greater than the automatic approach. And for the rest of the anatomic structures in our research, the automatic method resulted in higher VS coefficients.

The HD coefficients of the automatic method are generally a bit higher than the baseline method, except that the baseline method resulted in much higher HD coefficients in the malleus and the automatic method resulted in much higher HD coefficients in the facial nerve channel. The automatic method resulted in lower AVD coefficients in all structures. The comparison of MHD showed that the automated method resulted in higher average MHD coefficients in lateral SCC, posterior SCC and vestibule.

3.3 Evaluation of Manual Tracing Results Generated by VolEditor

The evaluation of VolEditor contained two parts. The first part was compared the manual segmentation results with Otsu's automatic segmentation of bone mask. The second part compared the 3D tracing results of VolEditor to conventional 2D tracing results.

3.3.1 Comparing the 3D-view Manual Segmentation Results with Otsu's Automatic Segmentation of Bone Mask

Figure 15 and 16 show examples of manual segmentation labels that are not part of the bone mask of Otsu's multi-threshold in the automated segmentation approach. The right images in the figures are the original CT slices and the left images are the original CT images overlapped with the manual segmentation mask. The blue areas shown in the images represent the region where the manual segmentation mask overlaps with the bone mask, and the red areas represent the region where the manual segmentation mask does not overlap with the bone mask.

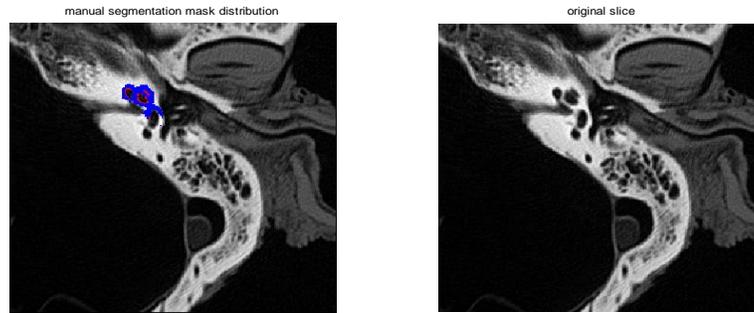


Figure 15 Example of voxels in the manual segmentation that do not overlap with the bone mask

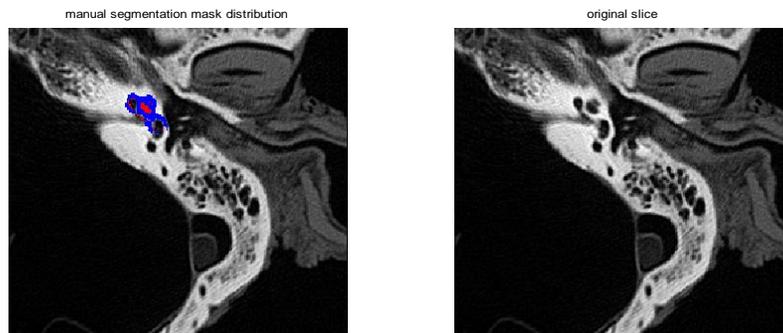


Figure 16 Example of voxels in the manual segmentation that do not overlap with the bone mask

The percentage of voxels in the manual segmentation that do not overlap with the bone mask of multi-threshold results from Otsu's method are computed in this work and the corresponding results are shown in table 8.

percentage of voxels in the manual segmentation that do not overlap with the bone mask of multi-threshold results from Otsu's method	mean	median	variance	max	min
cochlea	23.60%	20.10%	0.018138848	80.15%	14.80%
malleus	39.71%	40.81%	0.015725298	66.31%	18.40%
incus	41.60%	42.55%	0.010329359	72.14%	21.44%
facial nerve channel	31.29%	25.49%	0.027082725	66.26%	8.27%
lateral SCC	19.73%	15.06%	0.016050776	60.70%	7.40%
posterior SCC	13.49%	12.01%	0.003401964	31.51%	6.04%
superior SCC	15.56%	10.79%	0.017800412	68.46%	5.28%
vestibule	20.93%	15.42%	0.028936062	86.42%	7.87%
max	41.60%	42.55%	2.89%	86.42%	21.44%
min	13.49%	10.79%	0.34%	31.51%	5.28%
median	22.27%	17.76%	1.69%	67.39%	8.07%

Table 8 Non-overlapping percentage

According to table 8, the corresponding percentages for different structures range from 5.28% (superior SCC) to 86.42% (vestibule). On average, the posterior SCC has the lowest percentage of 13.49% in our test set, and incus has the highest percentage of 41.60%. The variance of vestibule's corresponding percentage is the highest in our validation results. For convenience purpose, the "percentage of voxels in the manual segmentation that do not overlap with the bone mask of multi-threshold results from Otsu's method" are referred as "the non-overlapping percentage" in the following sections.

3.3.2 Comparing the 3D-view Manual Segmentation Results Traced by VolEditor with 2D-view Manual Segmentation Results

The left image of Figure 17 is an example of 2D manual tracing result and the image on the right side is an example of 3D manual tracing result generated from VolEditor.

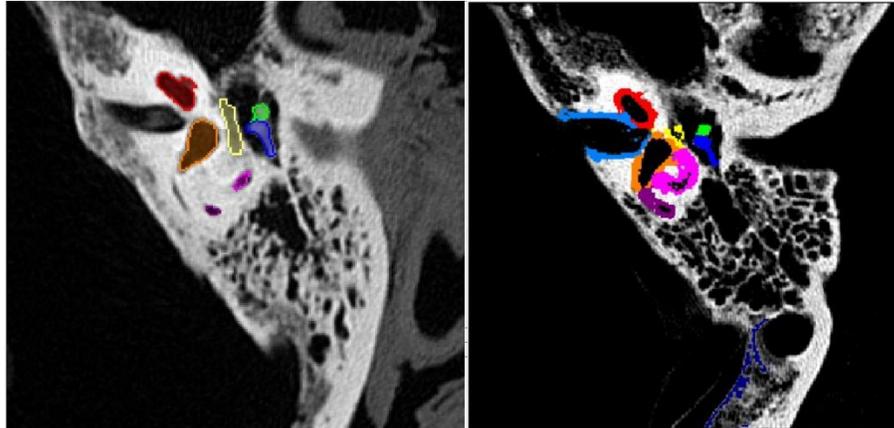


Figure 17 Comparison between 2D-view based segmentation and 3D-view based segmentation

Through visual inspection, we can easily notice that the 2D segmentation's contours are lying on the boundaries between bone structures and the background/soft tissue. In contrast, the 3D tracing results generally have fuzzy boundaries between bone structures and the background/soft tissue. Part of the segmentation labels of VolEditor are observed to spread into the bone. This is caused by the 3x3x3 kernel used in the labeling tool in VolEditor.

These results suggest that manual tracing of the structures using VolEditor is more likely to label voxels that belong to soft tissue and background masks segmented by Otsu's method as the bone structures. Additionally, the segmentation labels provided by

VolEditor may spread non-uniformly into the bones segmenting an irregular bony surface of the structure.

Chapter 4 Discussion and Conclusion

4.1 Analysis of Validation Results

In this work, an atlas-based method of segmentation of temporal bone anatomy is presented. We rely on a two-step image registration to register the manually constructed atlas of temporal bone anatomy to the test CT images with elastix [10]. Based on the overlapped atlas and the test images, a series of image processing techniques, introduced in chapter 2, are applied to the test images to locate the anatomic structures of research interest.

Our validation results in tables 4 and 5 showed that the automated segmentation of cavitory structures, like the cochlea and vestibule, resulted in varying results. The average DICE coefficients of the vestibule on both sides of the temporal bone are less than 0.6, while the cochlea's average DICE coefficients are greater than 0.6. There are three reasons which may have caused this phenomenon. One is that the ROI mask of the 2nd image registration is targeting the cochlea of the fixed image. Thus, the registration error of the cochlea would be expected to be less than that of vestibule, resulting in a better segmentation result of the cochlea than the vestibule. The other reason is that, compared to cochlea, the vestibule has more 'holes' in its' cavitory wall that could cause more errors in the segmentation procedure. Table 8 shows that the mean non-overlapping percentages of cochlea and vestibule are 23.60% and 20.93% respectively. The

percentage of bone mislabeled in the manual segmentation using VolEditor also contributes to the lack of overlap as seen in the DICE metrics in cochlea and vestibule

The malleus and incus, blob-like structures, have mean DICE coefficients that range from 0.6185 to 0.6706. This range is relatively good for the DICE validation results while the mean VS coefficients of malleus and incus is below average ($VS < 0.8$) when compared to the other structures evaluated in this study. As shown in table 8, the malleus and incus have the highest average non-overlapping percentage among all the anatomic structures in this study. Combining the mean VS coefficients of the malleus and incus with these highest percentages, we infer that the degradation of the VS coefficients was caused by VolEditor falsely labeling soft tissue and background voxels surrounding the malleus and incus as part of the corresponding segmentation results. Except for the VS metric, the validation results of the malleus and incus are above average in other metrics.

Our current method's robustness for tubular structure, such as the facial nerve channel and the semi-circular canals (SCC), is not good. For example, the mean DICE coefficients of these structures were between 0.4623 and 0.6466, and the variance of the DICE coefficients for the left facial nerve channel were the highest among all of the structures. Additionally, the maximum mean AVD and MHD were observed in either the facial nerve channel or the SCCs.

The principal axes orientation difference of the posterior SCC and vestibule are the highest (i.e. 50 degrees) when compared to the other structures in tables 6 and 7. This orientation difference between the automatic segmentation mask and manual mask is more stable and useful for structures like facial nerve channel. The orientation of the

vestibule's segmentation mask is generally more misaligned compared to other structures evaluated in the study. Combining the DICE validation results of the vestibule (0.5245 ~ 0.5834) with the principal axes orientation difference, indicated that that even when there is significant overlap between the manual and automatic segmentation masks, the automatic segmentation mask could still miss important shape/orientation information (the surface or boundary) as compared to the manual segmentation mask.

Overall, the validation results between automatic and manual segmentation showed that our method is good enough to provide visual enhancement in 3D visualization of temporal bone anatomy to help the surgeon identify the structures and provide segmentation results for surgical training or teaching, but due to the accuracy of our current method, it is not yet capable to provide precise segmentation results for further clinical use.

4.2 A Comparison between Automatic and the Baseline Approach

The segmentation results of the baseline method for malleus and incus had better validation results for the VS metric than other structures. The DICE coefficients were similar for both the automatic and baseline segmentations. This indicates that the topological variations of the malleus and incus are relatively low as compared to other anatomic structures analyzed in this study. The segmentation results of the automated method generally had close or better validation results compared to the baseline approach for the remaining structures, especially for structures like the facial nerve channel that has high topological variety.

4.3 The 3D-view Based Manual Tracing Method

As introduced in chapter 2, all the manual segmentation masks are generated using VolEditor. Compared to conventional 2D manual tracing method based on a 2D view, this software provides a 3D view for the user to label all the anatomic structures of interest. This is an attractive property because it greatly improves the speed of manual labeling and it's easier to train people how to label anatomic structures in a 3D view.

The current limitation of this software is that there's a transfer function that needs to be set at the beginning of the manual tracing for each test image. Manually setting the transfer function subjectively results in inconsistencies in the segmentation results. For example, the voxels which lay on the boundary of soft tissue and bone structures might be over-enhanced as a result of the transfer function and be segmented as soft tissue or bone when they should not be. Another problem with segmentations using VolEditor are labeling a 3D region pointed by the mouse with user selected size. Thus, when a user wants to label a specific region, all the voxels in this region with non-zero gray level will be labeled. It was observed that this feature could sometimes cause over-segmentation of the bony structures.

From examination of table 8, we discovered that the manual tracing labels in bony structures has spread into background and soft tissue masks of Otsu's multi-threshold. On average, each structures' manual segmentation's non-overlapping percentage was at least 13.29%.

These two problems are causing loss in our validation results. Our ongoing research with 2D manual segmentations also confirmed this observation.

4.4 Conclusion and Suggestions for Future Work

The performance of the proposed method is, on average, better than the baseline approach. The automatic segmentation results could be served as a preprocessing result for higher level segmentation approaches, i.e. model-based approach. The segmentation results of the current approach are not accurate enough for surgical planning, but was adequate for use in surgical simulation applications.

The following improvements could be made to the registration process: The manually rescaled CT volumes used for the 1st registration could be replaced with the bone level of Otsu's multi-level threshold to gain more consistency. A more refined 2nd registration procedure with an ROI_mask (introduced in chapter 2) specific for each anatomic structure could also potentially provide better segmentation results.

Finally, some improvements could be made to VolEditor. First, the transfer function in the VolEditor should be changed to an automatic procedure with well-defined standards to reduce inconsistency of manual segmentation results. Second, the spreading feature of the labeling tool of VolEditor could be modified, so that the labels would not spread in all directions. For example, if the selected region is on the surface of a bone structure, then the label should only spread in the neighboring voxels which are also on this surface, but not spreading into the inside region of the bone structure.

References

- [1] A. Arora et al, "Virtual reality case-specific rehearsal in temporal bone surgery: A preliminary evaluation," *Int. J. Surg.*, vol. 12, pp. 141-145, 2014.
- [2] S. Chan et al, "High-fidelity haptic and visual rendering for patient-specific simulation of temporal bone surgery," *Comput. Assist. Surg.*, vol. 21, pp. 85-101, 2016.
- [3] D. H. Lee et al, "Reconstruction and exploration of virtual middle-ear models derived from micro-CT datasets," *Hear. Res.*, vol. 263, pp. 198-203, MAY, 2010.
- [4] J. H. Noble et al, "Automatic Identification and 3D Rendering of Temporal Bone Anatomy," *Otol. Neurotol.*, vol. 30, pp. 436-442, JUN, 2009.
- [5] J. H. Noble et al, "Automatic segmentation of the facial nerve and chorda tympani in CT images using spatially dependent feature values," *Med. Phys.*, vol. 35, pp. 5375-5384, DEC, 2008.
- [6] J. H. Noble et al, "Determination of drill paths for percutaneous cochlear access accounting for target positioning error - art. No. 650925," *Medical Imaging 2007: Visualization and Image-Guided Procedures, Pts 1 and 2*, vol. 6509, pp. 50925-50925, 2007.
- [7] T. Rodt et al, "3D visualisation of the middle ear and adjacent structures using reconstructed multi-slice CT datasets, correlating 3D images and virtual endoscopy to the 2D cross-sectional images," *Neuroradiology*, vol. 44, pp. 783-790, SEP, 2002.
- [8] M. Seemann et al, "Evaluation of the middle and inner lear structures: comparison of hybrid rendering, virtual endoscopy and axial 2D source images," *Eur. Radiol.*, vol. 9, pp. 1851-1858, 1999.
- [9] K. Hassan et al, "Evaluation of software tools for segmentation of temporal bone anatomy," *Stud Health Technol Inform.*, vol. 22, pp. 130-133, 2016.
- [10] S. Klein et al, "elastix: A Toolbox for Intensity-Based Medical Image Registration," *IEEE Trans. Med. Imaging*, vol. 29, pp. 196-205, JAN, 2010.
- [11] F. Maes et al, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imaging*, vol. 16, pp. 187-198, APR, 1997.
- [12] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool," *BMC Med. Imag.*, vol. 15, pp. 29, AUG 12, 2015.
- [13] LR. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26(3), pp. 297-302, 1945.

[14] GJ. McLachlan, "Mahalanobis distance," *Resonance*, vol. 4, pp. 20-6, 1999.

[15] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, No. 1, pp. 62-66, 1979.

[16] CA. Schneider et al, "NIH Image to ImageJ: 25 years of image analysis," *Nat. Methods*, vol. 9, pp. 671-675, JUL, 2012.