Automated Detection of Cochlear Structures on Optical Coherence Tomography Volume Images

Hee Yoon Lee
Department of Electrical Engineering
Stanford University, Stanford, CA
lhy2090@stanford.edu

Abstract - Segmentation of inner structures of cochlea from OCT volume image is important to improve patient (or animal) diagnosis and treatment. In this study we show that three membranes and three chambers in the cochlea can be automatically segmented with good accuracy with the aid of locational and morphological characteristics. The segmented membranes can be used for vibration measurements and the calculated volume of the chambers can aid diagnosing cochlear diseases such as Secondary Endolymphatic Hydrops (SEH) and blast injuries.

I. INTRODUCTION

Cochlea is the portion of the inner ear responsible for converting mechanical sound vibrations into electrical signals that are carried to the brain via the auditory nerve. Most of hearing-related diseases are originated from the cochlea, which makes it important to diagnose and find a treatment in physiological studies. Diagnosis of cochlear diseases is generally very invasive and imaging them in vivo with high resolution was impossible.

Optical Coherence Tomography (OCT) is a relatively new imaging technology that images subsurface structures of biological samples noninvasively. 3D imaging can be done almost in real-time in 3D and it has high resolution (several microns), and large imaging depth (2-3mm). OCT can also measure the vibration of the sample in a point by point manner in 3D spaces [1], which attracted researchers in this field to do both imaging and vibration measurement of cochlea. Inner structures of the cochlea could be viewed nicely with OCT [2] and the vibration of membranes along the spirally-shaped turns in the cochlea could be measured in vivo [3]. Fig. 1 shows a spirally turned cochlear model (a), 2D cross-sectional OCT image (b), and an enlarged view of one chamber with some important structures labeled (three chambers and three membranes).

When working in two-dimensional spaces, it is relatively quick and easy to do vibrational measurement: finding the membrane, illuminating the laser to the point, and measuring the vibration over time. However when it comes to 3D, it is time-consuming to find the region of interests manually frame by frame. Automatic detection of those particular membranes is crucial for animal experiments and for human diagnosis in the future [4]. In addition, monitoring the scala media (SM) can help clinician to diagnose some diseases such as SEH and blast injuries. To be specific, automated quantitative measurement of SM volume can aid the diagnosis.

In this study we show that the three membranes (RM, TM, BM) and the three chambers (ST, SM, SV) can be automatically segmented with good accuracy with the aid of morphological information. Quantitative volume measurement is applied to investigate the amount of volume expansion in SM after mouse have experienced a blast, compared to the mouse with normal hearing.

![Image](image1.jpg)

**Fig. 1** (a) Cochlear model, (b) 2D cross-sectional image (scale bar: 200µm x 200µm), (c) Enlarged view of one chamber. Reissner’s membrane (RM), tectorial membrane (TM), basilar membrane (BM), scala Tympani (ST), scala media (SM), scala vestibuli (SV) are labeled.

II. ALGORITHM

A. Data preparation

OCT volume datasets from the mouse cochlea in vivo are collected with custom-built system in the lab. All datasets are composed of the same imaging parameters, for example; Number of data points: 240 in depth (z), 200 in lateral dimension (x, y), total points = 9.6 MVoxels per dataset. Image dimension: 2mm (z) x 1.5mm (x) x 1.5mm (y). Image resolution: 8.4µm (z) and 7.5µm (x, y). Two dataset from two normal mice and two another dataset from two mice that experienced a blast 20 minutes before the data collection.

B. 2D Image processing

Overall flow diagram of image processing for region detection and membrane segmentation is shown in the Fig. 2 with intermediate processed images. Detection and segmentation are done frame by frame except for the first step.
a) Register a raw volume dataset and crop images to reduce the processing time and focus on the region of interest (red box).

b) Take a weighted average between the consecutive frames. New frame is calculated by:
   \[
   \text{frame2}(i) = [0.3 \ 0.4 \ 0.3] \ast \{\text{frame1}(i-1); \text{frame1}(i); \text{frame1}(i+1)\};
   \]
   This process is to reduce speckle noise on the image by sacrificing some details of the images blurred. It also minimizes the effect of pixel saturation which appears in a few frames as a white vertical line which may divide one blob into two pieces unnecessarily.

Processes from (c) to (j) are applied to each frame \((i^{th})\) identically. We will name the frame variable before the processing as ‘frame1’ and the variable after the processing as ‘frame2’.

c) Blur the image by filtering with a two dimensional Gaussian filter.
   \[
   h = \text{fspecial}('gaussian', 5, 1);
   \text{frame2}(i) = \text{imfilter}(\text{frame1}, h, \text{'replicate'});
   \]


d) Interpolate the image bi-linearly to increase number of pixels in the image.
   \[
   \text{frame2} = \text{imresize}(\text{frame1}, 2, \text{'bilinear'});
   \]
   This step is to have the structures contain more pixels such that the following image processing can be done more precisely and accurately. For example, when circular structuring element is applied, it has to be very small thus less precise and less quantized, eventually looking like a diamond.

e) Morphologically close holes.
   \[
   \text{se} = \text{strel}('disk', 3);
   \text{frame2} = \text{imclose}(\text{frame1}, \text{se});
   \]
   It is identical to a dilation followed by an erosion, using the same structuring element. Disk-shaped structuring element is used to smooth out the holes effectively and the operation is iterated 6 times in total. Main purpose of this process is to close the tunnel under TM. This tunnel may be detected together with SM region, eventually causing overestimation of the SM volume and wrong segmentation of TM.

f) Blur the image again with Gaussian filter
   \[
   h = \text{fspecial}('gaussian', 4, 1);
   \text{frame2}(i) = \text{imfilter}(\text{frame1}, h, \text{'replicate'});
   \]

g) Binarize the image with the intensity threshold determined by Otsu’s method [5].
   \[
   L = \text{graythresh}(\text{frame1});
   \text{frame2} = \text{im2bw}(\text{frame1}, L\ast \text{alpha});
   \]
   The threshold obtained by Otsu’s method is lowered a little bit by a factor of ‘alpha’ in order to separate the three chambers apart from each other clearly, getting rid of the chances of the chambers to be combined. From trials and errors, ‘alpha’ is decided to be 0.6–0.8.

h) Smooth the blob regions by repeating ‘dilate-erosion’.
   \[
   \text{se} = \text{strel}('disk', n);
   \text{frame2} = \text{imdilate}(\text{frame1}, \text{se}); \text{frame2} = \text{imerode}(\text{frame2}, \text{se});
   \]
   Since OCT image suffers from speckle (scattering) noise, the edges of the blob regions tend to be wiggly. To flatten out those edges, dilation and erosion is recursively applied. The radius of the disk-shaped structuring element \((n)\) is progressively increased from 3 to 5.

i) Detect 3 chambers (ST, SM, SV) with correct labels.
   \[
   L = \text{blabel}(1-\text{frame}, 4);
   \]
   Binarized image is inverted and blob detection algorithm is performed to detect isolated regions. If the previous steps were performed correctly, the three chamber regions must have been detected together with the region on top and some additional miscellaneous small regions. The number of detected regions was reduced to 4 by ranking them according to the area of the blob and ignoring those with smallest area.

---

**Fig. 2.** Overall flow diagram of image processing for region detection and membrane segmentation. Each step show: original volume dataset (a), frame averaged (b), blurred (c), resized (d), hole closed (e), blurred again (f), binarized (g), dilate-erosion (h), region detected (i), membrane segmented (j).
Among 4 regions, the three chambers could be correctly labeled based on the center location of each blob. For example, the blob with the smallest x-component of the center coordinate is labeled as ST since it is always located on the left-most area. This algorithm needs to be adjusted if the image is taken at different angle or at different side of spiral turn.

j) Segment 3 membranes (RM, TM, BM) with correct labels

A few assumptions are made for the membrane segmentations: 1. each membrane is located in the proximity to or between the 3 chambers, 2. vertical or horizontal length of each membrane is fixed, 3. membrane is continuous.

- RM: center pixels between the right edge of ST and the left edge of SM.
- TM: starting from the bottom-left corner of SM and drawing a line along the bottom edge of SM until the assumed length of TM.
- BM: starting from the topmost corner of SV and drawing a line to the left of the top edge of SV until the assumed length of BM.

If the segmented line experiences a sudden jump of more than 7 pixels due to wiggly edges, the line is automatically extended with the same slope as the previous segmentation.

C. Segment smoothing

As shown in the Fig. 2 (j), the segmented membrane is usually wiggly and discontinuous due to uneven edges of chambers. However, the actual shape of the membranes must be smooth, round and continuous, therefore, we applied segment smoothing algorithm to the three-dimensional membrane segments.

If the $j^{th}$ element of the RM segment in the $i^{th}$ frame has a coordinates of $(x(i,j), y(i,j), z(i,j))$, this location is adjusted by taking weighted averages to its location and the neighborhood pixels’ coordinates. The new location coordinates will then be,

$$\begin{bmatrix} x(i-1,j) \ y(i-1,j) \ z(i-1,j) \\ x(i,j) \ y(i,j) \ z(i,j) \\ x(i+1,j) \ y(i+1,j) \ z(i+1,j) \end{bmatrix} \left[ \begin{array}{ccc} 0.25 & 0.5 & 0.25 \end{array} \right]$$

This location averaging is repeated 100 times to have a circular shape as compared in the Fig. 3.

![Fig. 3. En-face view of 3 membrane segmentations before (a) and after (b) the segment smoothing algorithm is applied.](image)

III. RESULTS

A. Volume measurement of SM partition

Since most of the diseases affect the volume of the SM partition, its regions are connectorized throughout the frames to build up a chamber volume. The detected volumes of normal and blast mice are shown in the Fig. 4. The number of frame for the blast mouse data was smaller than the normal mouse due to the limited viewing angle which often defers between mice and surgeries.

![Fig. 4. Volume rendered view of the 3D dataset (a) and the segmented portion of the SM chamber of normal (b) and blast (c) mice.](image)

The segmented membranes of normal and blast mice are displayed in 3D in the Fig. 5 with different colors: blue for RM, green for TM, and red for BM.

<table>
<thead>
<tr>
<th></th>
<th>Normal 1</th>
<th>Normal 2</th>
<th>Blast 1</th>
<th>Blast 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume (mm$^3$)</td>
<td>0.00939</td>
<td>0.00918</td>
<td>0.00762</td>
<td>0.00722</td>
</tr>
<tr>
<td>Area (mm$^2$)</td>
<td>0.02052</td>
<td>0.02007</td>
<td>0.03277</td>
<td>0.03107</td>
</tr>
</tbody>
</table>

Table 1. Measurement of volume and average area of the SM regions

B. Membrane segmentation

The segmented membranes of normal and blast mice are having only a few hundreds of pixels. Therefore these segmentation results can be used as a ROI mask for the vibration measurement in the cochlea. Without automatic segmentation, one might have to go through all voxels one by one, play the sound and measure the vibration, which consumes huge amount of time. This kind of segmentation with OCT images to aid clinician has been a fundamental to diagnose the progress of retinal diseases [6]
IV. CONCLUSION

We have successfully detected three chambers and the three membranes from the volume dataset of the cochlea. From the detected SM region, we found the average area was about 1.6 times larger for the blast mouse cochlea than for the normal mouse. Also, we achieved smooth circular-shaped membrane segments in 3D, which can be useful for vibration measurement of cochlea with OCT system.

All the processes are done in the MATLAB software and the processing time was about ~5s, which is too slow for the real-time application in clinic. Therefore as a next step, the algorithm needs to be more efficient in terms of timing. With more volume datasets, we can also extract the morphological information of each structure for better and faster segmentation.

ACKNOWLEDGMENT

The authors would like to thank to David Chen, course TA and project mentor, for providing valuable help and guidance in this project. We would also like to acknowledge Professor Bernd Girod for teaching us with a variety of image processing tools and algorithms, which have been certainly useful in this project as well as my phd research.

REFERENCES