1-1-2009

An analysis of medical image processing methods for segmentation of the inner ear

Catherine Todd
University of Wollongong in Dubai, cath@uow.edu.au

Mikhail Kirillov
University of Wollongong in Dubai

Muaaz Tarabichi
American Hospital Dubai

Fazel Naghdy
University of Wollongong, fazel@uow.edu.au

Golshah Naghdy
University of Wollongong, golshah@uow.edu.au

AN ANALYSIS OF MEDICAL IMAGE PROCESSING METHODS FOR SEGMENTATION OF THE INNER EAR

Catherine Todd
College of Informatics and Computer Science, University of Wollongong in Dubai
Knowledge Village, Dubai, United Arab Emirates

Mikhail Kirillov
College of Informatics and Computer Science, University of Wollongong in Dubai
Knowledge Village, Dubai, United Arab Emirates

Muaaz Tarabichi (MD)
Ear Nose and Throat Department, American Hospital Dubai
Oud Metha, Dubai, United Arab Emirates

Fazel Naghdy
School of Electrical, Computer and Telecommunications Engineering, University of Wollongong (Australia)
Northfields Avenue, 2522, NSW, Australia

Golshah Naghdy
School of Electrical, Computer and Telecommunications Engineering, University of Wollongong (Australia)
Northfields Avenue, 2522, NSW, Australia

ABSTRACT

This study explores software development methods and subsequent results for delineation of the inner ear using medical image processing techniques with clinical relevance such as for pre- and post- operative evaluations, surgical planning and exploration. Methods for data acquisition and segmentation of inner ear anatomy, specifically the cochlea, are analyzed. Segmentation methods for extracting and rendering the cochlea from Computed Tomography are implemented using an ITK/VTK approach, and results are provided for comparison. These include variations of region-growing, threshold-based and level set segmentation methods. The analysis focuses on image acquisition, registration and extraction of the complex cochlear spiral and surrounding anatomy, with previous comparisons reviewing a broad-spectrum of medical image segmentation strategies. The review is intended to provide a comparative analysis of recent methods in segmentation of middle and inner ear anatomy, and ensuing results in this field of medical image processing.

KEYWORDS

Image Segmentation, Visualization, Region-growing, Threshold Level-set.

1. INTRODUCTION

Computer-aided reconstruction of the temporal bone region has been implemented using imaging modalities including histology, Computed Tomography (CT) and Magnetic Resonance Imaging (MRI), to capture and visualize middle and inner ear data. CT is commonly used for pre- and post- operative evaluation of patients for surgical planning and assessment, such as for CI recipients (Ketten, Skinner et al. 1998; Boor, Maurer et al. 2000; Indrajit, Souza et al. 2003). CT can capture the intricate structures within the temporal bone region for virtual reconstruction. A number of sources have utilized high resolution spiral CT for three-dimensional modeling of the temporal bone, including the cochlea (Skinner, Ketten et al. 1994; Himi, Kataura et al. 1996; Ketten, Skinner et al. 1998; Hans, Grant et al. 1999; Boor, Maurer et al. 2000; Yoo, Wang et al. 2000).

Existing techniques for segmentation of anatomical structures from medical images include threshold-

based approaches, region growing (seeded and unseeded), clustering, deformable models (including snakes, geodesic and parametric active contours), level sets, model fitting, split and merge methodologies (Xianfen 2005; Kassim 2008), and watersheds (Wegner 1996). Structures are identified and segmented using either a manual, semi-automatic or fully automated approach. The latter method is desired to reduce time, cost and manual labor, however an approach resulting in highly accurate feature delineation is difficult to achieve due to complexity of structure shape, imperfections in image data and variability between patients (Kassim 2008). Complex structures can often be precisely defined only through manual segmentation, yet this is time-consuming and laborious. Manual and semi-automatic image processing methods are the most common for temporal bone reconstructions. This is mainly due to the vast number of complex shapes and variability of structure sizes within this region, as well as image artifacts. Interactive, semi-automatic segmentation is often performed by the user selecting a threshold value based on pixel intensities (Hans, Jackson et al. 2003). Recursive connectivity algorithms are also used, where a seed is selected on the ROI surface (Hans, Jackson et al. 2003) and areas not located on the surface are discarded. Region-growing algorithms have been applied (Seemann, Seemann et al. 1999; Yoo, Wang et al. 2000; Yoo, Wang et al. 2000) and artifacts manually removed (Seemann, Seemann et al. 1999). Erosion and dilation iterations are also performed (Rodt, Ratiu et al. 2002). Level set methods may be preferred to active surfaces as they can handle complex geometries and surface changes, yet require a high degree of initialization by the user (Taheri 2007). In this work, a variety of segmentation techniques are examined, including threshold level set and threshold region growing, for extraction of the cochlea and surrounding anatomy using spiral CT image sequences.

2. COCHLEAR SEGMENTATION IN ITK/VTK

2.1 Data Acquisition

Spiral CT images of human temporal bones of CI candidates, provided by the American Hospital Dubai, were analyzed in this work. Images were in DICOM format and had a spatial resolution of 768 x 768 pixels of 8-bit grey-scale, with 0.23 x 0.23 x 0.67mm voxel size (some datasets were reconstructed at 0.33mm).

2.2 Threshold Level Set Segmentation

Threshold level set segmentation is applied in this work in Image Toolkit (ITK). The method works by setting a pixel intensity range for the ROI within an upper and lower threshold. For this range, a level set equation is applied where a constrained connected-component surface evolution takes place based on a propagation term, P; calculated in (1), where g is an input image, U and L are Upper and Lower Thresholds.

$$P(x) = \begin{cases} g(x) - L & \text{if } g(x) < (U - L)/2 + L \\ U - g(x) & \text{otherwise} \end{cases} \quad (1)$$

Threshold level set segmentation begins propagation from one or more seed pixels contained within the ROI and area evolution continues until either a timing constraint is reached or there are no more voxels/pixels within the range of the threshold level for inclusion. Following extensive testing, optimal input parameters were defined for extraction of the cochlea from CT. In addition to selection of seed pixel(s), upper and lower threshold values were set to -800 and 900 respectively, for extraction of the human cochlea from a CT image that has not been pre-processed. The wide range of pixel intensities is essential if considerable noise exists within the scan. Furthermore, the threshold range will vary depending not only on the type of pre-processing technique applied (if any) but also the structure of interest tissue type (intensity).

The number of algorithm iterations was set to 1200 and an initial distance or neighborhood size surrounding the seed set to 5. The isosurface value was set to 0. Changing the maximum Root-Mean-Square (RMS) error (set to 0.04) provided an upper limit defining the number of allowable errors in during computations. A higher permissible error value enables greater numbers of pixel inclusions yet the boundary may extend to encapsulate unwanted regions. Propagation scaling (set to 1.0) provided an internal 'force'

effectively pushing the ROI from inside, trying to include as many pixels as possible thereby defining how far the ROI grew. Acting as an external force, curvature scaling (set to 1.0) constrained the spreading of the ROI by pushing on its boundary from the outside. Figure 1 demonstrates the result of Threshold level set applied to extract the human cochlea from a spiral CT scan sequence using the specified input parameters.

The algorithm propagation is affected by the initial neighborhood of the seed(s) selected, as well as the complexity and size of the structure, which also influences curvature scaling. Therefore the method requires the user to adjust parameters often or monitor algorithm propagation and correct it in real time. This demands additional time and input by the user, reducing the level of automation and increasing the complexity of the process. Figure 2 represents an attempt to extract the cochlea from another CT scan in the same image sequence, without varying the input parameters. Results show that segments of the cochlea are missing from the ROI. No pre-processing has been applied. Results demonstrate that Level Set algorithms vary in terms of segmentation output depending on the shape and size of anatomical structure, making them less suitable for applications where morphological complexity and patient variability is great, such as inner ear segmentation.

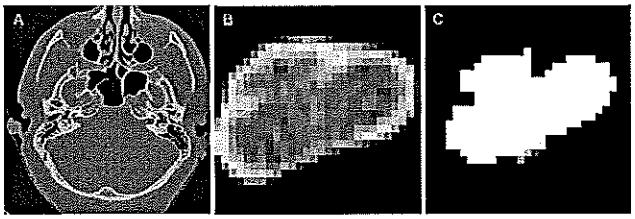

Figure 1. From left to right: Result of Threshold Level Set applied to a spiral CT image (left) containing a temporal bone. The cochlea is separated from surrounding anatomy (middle) and a subset of ROI pixels are segmented (right).

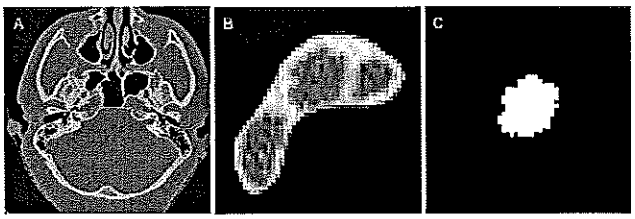

Figure 2. From left to right: Threshold level set applied to a different CT image (left) from the same dataset as in Figure 1. The cochlea is encased in a boundary (middle), yet after segmentation only part of the cochlea is extracted (right).

2.3 Region Growing Segmentation

Region growing is a simple yet effective approach for extraction of structures from medical scan sequences. The user may define one or more seeds within the ROI to initialize the region growing algorithm. Within each iteration, neighboring pixels are evaluated for possible inclusion into the region and based on some criterion (commonly, pixel intensity) are included or rejected. The main differences between region growing algorithms relate to methods for determining pixel neighborhoods, which pixels are to be added to the ROI and the algorithm for visiting neighboring pixels for evaluation.

2.3.1 Connected Threshold Region Growing

The main criterion of Connected Threshold for ROI pixel inclusion is a pixel intensity threshold range. The user specifies as input the seed index and lower and upper thresholds. Pixels are included into the ROI if their intensity values are in this range. In order to iterate through the image and establish an ROI, a Flood Iterator is applied for visiting neighboring pixels. Since the algorithm requires three input values, it is ideal for applications that require a more automated approach, minimizing user input. One limitation of region growing is that segmentation results depend on seed selection (location and intensity). A seed (located at 260, 400) was selected in a spiral CT scan of a temporal bone. Upper and lower threshold limits were set to -800

and 900 respectively. Connected Threshold was applied in ITK with a Curvature Flow algorithm for pre-processing to reduce noise and increase the quality of the segmentation result; shown in Figure 3.

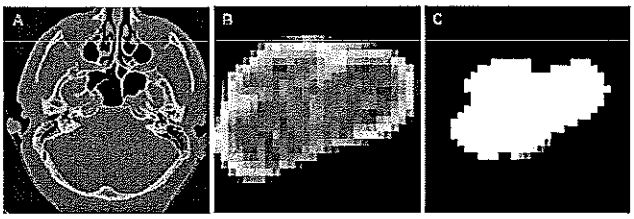

Figure 3. The result of applying Connected threshold region growing. From left to right: original CT image containing a temporal bone (left), an image containing the cochlea (middle) and segmentation result; comprised of ROI pixels (right).

2.3.2 Neighborhood Connected Region Growing

This segmentation technique is similar to Connected threshold; the main difference is that the criterion for determining whether a pixel is added to the ROI is based on pixel neighborhood (the size of which is defined by the user). A pixel is included into the region only when *all* neighborhood pixels have their intensities within the defined threshold range. The main reason for evaluating the neighborhood of each pixel is to make small structures (particularly noise or surrounding parenchyma) less likely to be added to the ROI. Neighborhood connected segmentation is therefore not suited for extraction of small anatomical structures, such as the cochlea. Figure 4 reveals this unsuitability: much of the cochlea is missing from the segmented ROI, due to the limitations of pixel inclusion. Input parameters were set as for Connected Threshold.

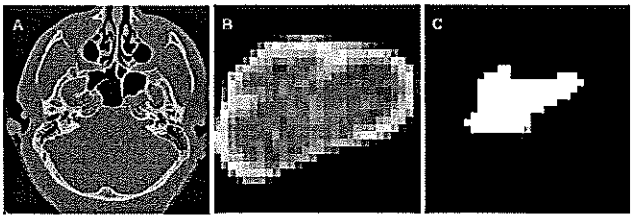

Figure 4. The result of applying Neighbourhood connected. From left to right: the original CT image (left), an image that contains a cochlea (middle) and segmentation result (right) which includes only a subset of pixels defining the cochlea.

2.3.3 Confidence Connected Region Growing

This segmentation method is based on the iterative computation of the mean (m) and standard deviation (σ) of pixel intensities included in the current region and is implemented in ITK. A range about the mean is established using a value (f), defined by the user, multiplied with the standard deviation. Neighboring pixels with intensities within this range are included in the region. After the first iteration, the mean and standard deviation of the current region are updated to give a new inclusion range and the process is repeated. This continues until there are either no more pixels to add to the region or a maximum number of algorithm iterations are reached. The inclusion criteria for a neighboring pixel (X) is represented by (2), where I() is the image containing pixel X. The algorithm therefore operates by analyzing and iteratively computing a threshold range based on statistical information associated with regional pixel intensities, beginning with the seed. Since this process is automatic, the algorithm is most suitable for image segmentation that requires limited user input, however it is highly sensitive to noise, therefore requiring image pre-processing.

$$I(X) \in [m - f\sigma, m + f\sigma] \qquad (2)$$

Figure 5 represents the output of this method applied to a temporal bone CT scan, for extraction of the cochlea. The result reveals that data is missing from the ROI, both within and on its boundary.

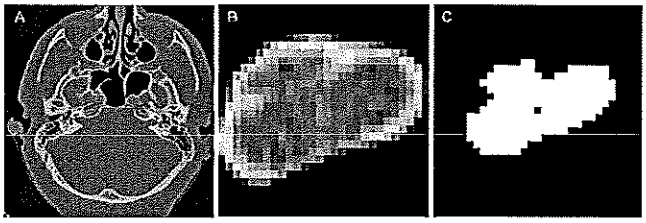

Figure 5. From left to right: segmentation results (right) of applying Confidence Connected Region Growing to a spiral CT scan (left). A region defining the cochlea is shown (middle).

2.3.4 Isolated Connected Region Growing

This segmentation me The Isolated Connected region growing algorithm requires at least two seed pixels to be specified and only one value for intensity threshold; the lower one. The algorithm tries to grow the region connected to the first seed location, and not connected to the second seed. In order to achieve this, the value for the upper threshold, that separates the seeds, is calculated automatically using a binary search. This method is mostly used when boundaries of an anatomical structure are indistinct. Since the shape (boundary) of the cochlea *is* distinct due to a high difference between the encasing bone and the fluid-filled chambers, as well as the membranous CP, this method does not perform well. It also requires additional user input for specification of a second seed. The intensity value of this seed affects the upper threshold; therefore if a second seed is selected in an area of high contrast with the structure of interest, the result may differ greatly between trials, as illustrated in Figure 6. To generate this result, two different second seeds of varying pixel intensity are selected, from one application of Isolated Connected region growing for extraction of the cochlea to the next. Pre-processing is not applied. Figure 9 (bottom, left) is the result of choosing seed pixels 1 and 2 whilst Figure 9 (bottom, right) is the result of seed pixels 1 and 3. The cochlea is successfully segmented with seed pixels 1 and 3, yet it cannot be distinguished using seed pixels 1 and 2 which demonstrates unpredictable variation between trials. The Isolated region growing method is unlike most other region growing techniques in that it requires the user to select at least two seeds, which adds to the time and labor required to segment the anatomical structure of interest.

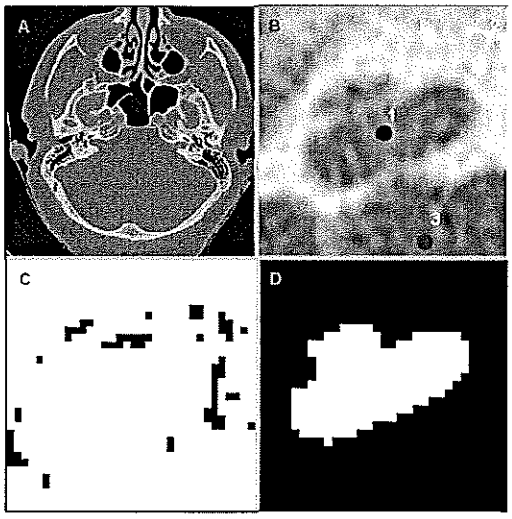

Figure 6. Segmentation of the cochlea using Isolated Connected region growing applied to a spiral CT temporal bone scan (left, top). Top to bottom, left to right: the original image (top, left), a magnified image region that contains the cochlea, with seed pixels 1, 2 and 3 marked (top, right), result of cochlear segmentation using seed pixels 1 and 2 (bottom, left), result of cochlear segmentation using seed pixels 1 and 3 (bottom, right).

3. COMPARISON OF RESULTS

In medical image processing, segmentation results are often analyzed qualitatively, via subjective evaluation by a specialist, which is subject to human error, bias and may take considerable time. Quantitative error measures provide a more objective approach for evaluation of segmentation algorithms and usually fall into one of two categories: supervised and unsupervised. The former requires a *priori* knowledge about the ROI; usually this information is a reference image denoting the exact region to be extracted from the image which is then used as a benchmark to compare against segmentation algorithm results; yielding a measure of variance or error as an indication as to algorithm performance. Unsupervised evaluations do not require a reference image but instead use a set of features or characteristics against which to measure the segmentation result. A supervised approach has been applied to evaluate the performance of the threshold level set and region growing algorithms for segmentation of the cochlea implemented in this work. The error of over-merged and under-merged pixels is calculated, as well as pixel distance error for over-merged pixels, for each image in the dataset that contains the cochlea. Each segmentation algorithm is applied to 33 datasets in total (pertaining to 33 cochleae from CI candidates). The desired ROI corresponding to the image undergoing segmentation is extracted semi-automatically to give a reference image. This process requires the user to interactively set parameters for edge-detection and threshold levels as well as seed placement for interactive region growing. The reference image is then quantitatively compared with the segmentation result to give measures of error. The error mean and standard deviation for all image sets tested are summarized in Table I.

The degree of error associated with under-merging, e_u, is calculated in (3) by considering the number of non-segmented pixels, j, that lie within or on the ROI boundary in the reference image as a percentage of the total number of pixels, k, that comprise the ROI. An error value, e_u, of 0% signifies that all pixels comprising the ROI were successfully segmented by the applied algorithm. To obtain the pixel count j, the binary image of the segmented region is subtracted from the binary reference image.

$$e_u = 100 \times \left(\sum_{m=1}^{j} m\right) / \left(\sum_{n=1}^{k} n\right) \quad (3)$$

Over-merging, as calculated in (4), considers those segmented pixels that lie outside the ROI, as a percentage of the total number of pixels, k, that comprise the ROI. An over-merging error value, e_o, of 0% indicates that all pixels segmented from the image belong to the ROI and none exist outside of it. A high percentage error means that a high proportion of pixels were segmented that did not belong to the ROI. To obtain the pixel count, i, the image containing the segmented region is subtracted from reference image, to give a value i representing all pixels of value -1 (non-segmented pixels lying outside the ROI).

$$e_o = 100 \times \left(\sum_{m=1}^{i} m\right) / \left(\sum_{n=1}^{k} n\right) \quad (4)$$

Pixel distance error, e, considers the Euclidean distance of a misclassified pixel from the closest pixel that is contained within the ROI (taken as the boundary of the ROI in the reference image). The error is calculated in (5) (Yasnoff 1977) as the summation of the squared distances of all (k) misclassified pixels, d_m.

$$e_d = \sum_{m=1}^{k} d_m^2 \quad (5)$$

Table 1 contains a summary of results that shows the mean and variance (standard deviation, σ) in percentage of under-merging (eu), over-merging (eo) and pixel distance error (ed) for each segmentation method applied in the work, within each dataset and for all 33 cochleae datasets analyzed in the work.

Table 1. Results of Segmentation Algorithms

Segmentation Technique	eu (m)	eu (σ)	eo (m)	eo (σ)	ed (m)	ed (σ)
Connected Threshold	24.94	8.27	5.64	2.36	22.43	18.89
Neighborhood Connected	87.31	8.23	0.01	0.01	0.04	0.07
Confidence Connected	42.47	38.11	718.38	1426.20	479.83	908.46
Threshold Level Set	17.86	10.35	9.93	4.36	31.96	18.68
Isolated Connected	17.59	9.95	10.16	3.86	40.30	31.83

Overall, Connected threshold yielded the best results for extraction of the cochlea from a series of spiral CT scans, with a relatively low under-merging error (mean 24.94%, σ 8.27%), low over-merging error (mean 5.64%, σ 2.36%) and low pixel distance error (22.43, σ 18.89) in comparison to the other segmentation methods tested, for the same 33 image datasets. In addition, other parameters were considered during the testing, including amount of user interaction, number of input parameters required, as well as consistency of an algorithm between different datasets. Since the goal is to automate the process of segmentation as much as possible, one of the main criteria for selecting an algorithm was the number of user inputs required for the segmentation technique. Level set algorithm proved to be the most inconvenient due to the large number of parameters that require setup, such as propagation and curvature scaling, mean root square error value and/or number of iterations, and so on. Ideally, this algorithm performs best when its evolution is visually supervised and the parameters are adjusted accordingly. However, it would not be useful when used for clinical applications that are time-critical and surgeon input is to be minimized. In comparison, algorithms based on region growing have shown better results in terms of the number of user inputs required.

Confidence connected requires the least number of parameters; the seed pixel and initial neighborhood radius for calculation of threshold values. Yet it is sensitive to image noise, and so pre-processing is required prior to segmentation, increasing complexity and overhead. Connected threshold requires seed coordinates and lower and upper threshold values. This approach proved to be the most suitable both implementation-wise, and from an end-user perspective, as it requires minimal input values and enables heightened control for extraction of cochleae as well as surrounding structures. Neighborhood connected is similar to Connected threshold, but it requires another input value; the neighborhood radius. Isolated connected calculates the upper threshold value automatically, but it requires an extra seed pixel placed on a region of similar intensity that is not connected to the structure of interest, making it the least useful in the current application. This algorithm has limitations in cases where there is no isolation value available between two regions; where regions are connected and have similar intensity values, which is common in cochleae segmentation.

4. EXTRACTION OF SURROUNDING ANATOMY

Connected threshold region growing has also been implemented for segmentation of anatomical landmarks surrounding the cochlea; most importantly for CI pre-operative planning, the external ear canal. Parameter values for the segmentation are set at -1000 and -500 for lower upper thresholds respectively (no image smoothing was applied for this test). Results are shown in Figure 7. A user may segment any structure of interest and such extraction is possible through the use of pre-set values for bone, soft tissue, liquid and air. The user also has an option to set the threshold range for the anatomical landmark manually.

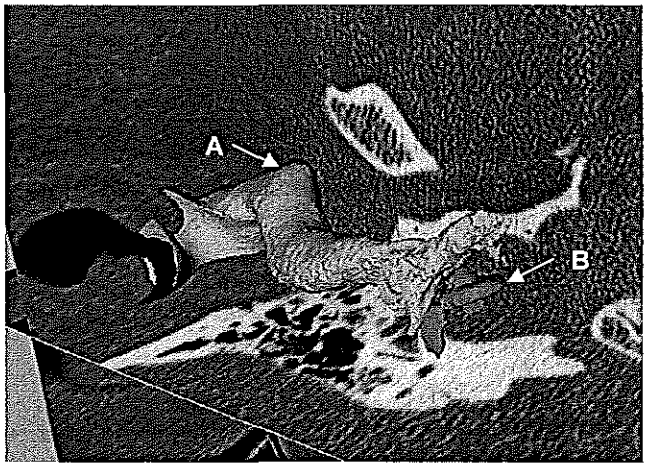

Figure 7. Three-dimensional visualization of the cochlea (B) and external ear canal (A) superimposed on spiral CT scans.

5. CONCLUSION

In this study, Connected threshold region growing provided the best results in extraction of the human cochlea, as measured quantitatively using over-merging, under-merging and pixel distance measurements, as compared against Threshold level set, Neighborhood connected, Isolated connected and Confidence connected segmentation methods. Additional factors that influenced results when applying the segmentation algorithms with automatic threshold selection were the resolution of the dataset and placement of the seed(s). Changes in image resolution and seed location produced differences in results which were more apparent in the algorithms based on a neighborhood region. Neighborhood and Confidence connected algorithms showed weaknesses when the seed is placed near the structure of interest boundary. Region growing methods require less user interaction and are more efficient than level set methods due to simplicity in implementation, making them a suitable choice for more automated extraction of structures from medical images. They are, however, dependent on image quality and so pre-processing may be required. Segmentation methods based on level set are best suited for extraction of complex shapes that contain disconnected components yet require significant user input and time to monitor and correct the segmentation process, making them less viable for the systems where automatic segmentation is desired. Connected threshold region growing was also applied in this work for extraction and modeling of surrounding anatomical landmarks for pre-operative evaluations such as for determining the best entry trajectory for CI insertion. Connected threshold may be tested against other methods such as watersheds, to further compare performance and used for 3D medical visualizations.

REFERENCES

Boor, S., J. Maurer, et al. (2000). Virtual Endoscopy of the Inner Ear and the Auditory Canal. *Head and Neck Radiology* Vol. 42, pp 543-547.

Hans, P., A. J. Grant, et al. (1999). Comparison of Three-Dimensional Visualization Techniques for Depicting the Scala Vestibuli and Scala Tympani of the Cochlea by Using High-Resolution MR Imaging. *The American Journal of Neuroradiology*, Vol. 20, No. 7, pp 1197-1206.

Hans, P., A. Jackson, et al. (2003). Virtual Reality Modelling Language: Freely Available Cross-Platform Visualization Technique for 3-D Visualization of the Inner Ear. *The Journal of Laryngology and Otology*, Vol. 117, No. 10, pp 766-774.

Himi, T., A. Kataura, et al. (1996). Three-Dimensional Imaging of the Temporal Bone Using a Helical CT Scan and Its Application in Patients with Cochlear Implantation. *Otorhinolaryngology*, Vol. 58, pp 298-300.

Indrajit, I. K., J. D. Souza, et al. (2003). Imaging of Cochlear Implants. *Indian Journal of Radiology and Imaging*, Vol. 13, No. 4, pp 371-379.

Kassim, A. A., Yan, P (2008). Medical Image Processing. *Wiley Encyclopedia of Computer Science and Engineering.*

Ketten, D. R., M. W. Skinner, et al. (1998). In Vivo Measures of Cochlear Length and Insertion Depth of Nucleus Cochlear Implant Electrode Arrays. *Annals of Otology, Rhinology & Laryngology*, Vol. 107, No. 11, pp 1-16.

Rodt, T., P. Ratiu, et al. (2002). 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. *Head and Neck Radiology*, Vol. 44, pp 783-790.

Seemann, M. D., O. Seemann, et al. (1999). Evaluation of the middle and inner ear structures: comparison of hybrid rendering, virtual endoscopy and axial 2D source images. *European Radiology*, Vol. 9, pp 1851-1858.

Skinner, M. W., D. R. Ketten, et al. (1994). Determination of the Position of Nucleus Cochlear Implant Electrodes in the Inner Ear. *The American Journal of Otology*, Vol. 15, No. 5, pp 644-651.

Taheri, S., Ong, SH., Chong, V (2007). Threshold-based 3D Tumor Segmentation using Level Set (TSL). *IEEE Workshop on Applications of Computer Vision (WACV'07)*, pp 45-50.

Wegner, S. H., T; Oswald, H.; Fleck, E. (1996). The Watershed Transformation on Graphs for the Segmentation of CT Images. *IEEE Proceedings of ICPR '96*, pp 498-502.

Xianfen, D. S., C; Changhong, L; Yuanmei, W (2005). 3D Semi-Automatic Segmentation of the Cochlea and Inner Ear. *Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pp 6285-6288.

Yasnoff, W. M., Mui, J K. and Bacus, J. W. (1977). Error Measures for Scene Segmentation. *Pattern Recognition*, Vol. 9, pp 217-231.

Yoo, S. K., G. Wang, et al. (2000). Three-Dimensional Modeling and Visualization of the Cochlea on the Internet. *IEEE Transactions on Information Technology in Biomedicine*, Vol. 4, No. 2, pp 144-151.

Yoo, S. K., G. Wang, et al. (2000). Three-Dimensional Geometric Modeling of the Cochlea Using Helico-Spiral Approximation. *IEEE Transactions on Biomedical Engineering*, Vol. 4, No. 10, pp 1392-1402.