Automated segmentation of the incus and malleus ossicles in conventional 3D CT images

Francisco P. M. Oliveira[1], Diogo Borges Faria[2], João Manuel R. S. Tavares[3]

[1] *Institute for Nuclear Sciences Applied to Health (ICNAS), and Institute for Biomedical Imaging and Life Sciences (IBILI), Faculty of Medicine, University of Coimbra, Portugal*

[2] *HPP Medicina Molecular, SA., Faculdade de Engenharia, Universidade do Porto, Portugal*

[3] *Instituto de Engenharia Mecânica e Gestão Industrial, Departamento de Engenharia Mecânica, Faculdade de Engenharia, Universidade do Porto, Portugal*

Corresponding Author:

Professor João Manuel R. S. Tavares,

Departamento de Engenharia Mecânica

Faculdade de Engenharia da Universidade do Porto

Rua Dr. Roberto Frias, s/n, 4200-465 Porto, PORTUGAL

Phone: +351 225 081 487

Fax: +351 225 081 445

Email: tavares@fe.up.pt

url: www.fe.up.pt/~tavares

Automated segmentation of the incus and malleus ossicles in conventional 3D CT images

Abstract

This paper proposes a fully automated computational solution to segment the incus and malleus ear ossicles in conventional tri-dimensional (3D) X-ray computed tomography (CT) images. The solution uses a registration-based segmentation paradigm, followed by image segmentation refinement. It was tested against a dataset comprising 21 CT volumetric images of the ear acquired using standard protocols and with resolutions varying from 0.162×0.162×0.6 to 0.166×0.166×1.0 mm^3. The images used were randomly selected from subjects that had had a CT exam of the ear due to ear related pathologies. Dice's coefficient and the Hausdorff distance were used to compare the results of the automated segmentation against those of a manual segmentation performed by two experts. The mean agreement between automated and manual segmentations was equal to 0.956 (Dice's coefficient), and the mean Hausdorff distance among the shapes obtained was 1.14 mm, which is approximately equal to the maximum distance between neighbouring voxels in the dataset tested. The results confirm that the automated segmentation of the incus and malleus ossicles in 3D images acquired from patients with ear related pathologies, using conventional CT scanners and standard protocols, is feasible, robust and accurate. Thus, the solution developed can be employed efficiently in CT ear exams to help radiologists and otolaryngologists in the evaluation of bi-dimensional (2D) slices by providing the related 3D model.

Keywords: medical imaging, imaging analysis, ear segmentation, atlas-based segmentation, image registration.

1. Introduction

The human auditory system is made up of the hearing and balance organs. It can be divided into three main parts: outer ear, middle ear and inner ear. The auditory ossicles are three very small bones: malleus, incus and stapes, within the middle ear and are joined in a chain. Their function is to transmit sound energy from the tympanic membrane to the fluid-filled cochlea. Thus, they are fundamental for hearing.

Computational models automatically built from real ear images can be of considerable medical assistance. For instance, these models can help radiological diagnosis, surgical planning, demonstrating the relationship between ear structures and their function, simulate pathologies of the ear, improve prosthesis designs, just to name a few uses. Image segmentation is a precondition for efficient modelling and analysis of the ear. For example, the correct segmentation of ear structures in a tri-dimensional (3D) image is critical for biomechanical studies, surgical planning, surgical implantology, and medical diagnosis in general.[1-7]

Due to the influence of noise, partial volume effects, the low resolution of the images acquired using conventional X-ray computed tomography scanners and the standard protocols in relation to the dimensions of the structures, and the presence of image artefacts, it is challenging to segment ear structures effectively; even for experienced experts doing it manually. Semi-automated methods can be applied, but they require many user interventions, especially in the initialization step. Additionally, small differences in the usual segmentation algorithm parameters can cause dissimilar results.[8]

Automatic identification and recognition of structures or objects in images using registration-based segmentation, also known as atlas-based segmentation, has been recognized as an effective approach.[9] The main idea of atlas-based approaches is to use a mean image representing the data involved as a “template image” and then spatially

normalize the image under study according to that template image (or vice-versa). As such, the structures in the image under study assume coordinates close to the ones of the template image, which facilitates the posterior analysis. In the present case, if a segmentation of the template image is also available, then a rough automated segmentation of the image under study can be obtained easily after the registration of the image under study with the template image. If the registration algorithm used is not fully automated, then the advantage of the registration-based method is reduced to the traditional semi-automated segmentation method level.

In the literature, the segmentation of ear structures in clinical or research applications has mostly been achieved manually or semi-automatically by trained experts.[2, 5, 10-11] Inner ear segmentation methods based on registration have been developed.[12-15] However, as far as the authors know, none of those methods are totally automated since, at least, a manual pre-registration of the image under study with the template image is initially required.

Hence, the main goal of this work was to develop a fully automated computational solution to segment the incus and malleus ossicles in conventional volumetric CT images using a registration-based segmentation paradigm. Besides being fully automated, the proposed solution was developed to be robust to large spacings between the image slices, and was tested against images from real patients, some of them with ear pathologies and with the ossicles partially occluded.

2. Methods

2.1 Dataset

The initial image dataset comprised 25 axial 3D CT randomly selected images of adult ears (13 right ears and 12 left ears). All subjects were asked by their doctor to do a conventional CT exam since they had signs of possible ear pathology.

Later, two axial images of the right ear and two axial images of the left ear were excluded due to the occlusion, by soft mass, of most of the ossicles, possibly due to chronic otitis media. However, images with minor occlusions of ossicles due to soft mass were kept (Figure 1).

Thus, the final image dataset contained 21 axial images, 11 right ears and 10 left ears. The resolution of the images varied from $0.162 \times 0.162 \times 0.6$ to $0.166 \times 0.166 \times 1.0$ mm^3.

(insert Figure 1 about here)

2.2 Automated segmentation of the ear

The automated segmentation solution developed is based on the registration of the image under study with a previously built template image as a reference image (Figure 2). After the registration process, a geometric transformation that matches the image under study with the template image is obtained. Then, the inverse of this transformation is computed and applied to the segmented template image. Thus, obtaining a segmented image.

To improve the segmentation process, the initial segmented image is morphologically dilated using a kernel of approximately 0.3 mm wide on each axis. The goal of this dilatation is to guarantee with a previously built template image as a reference image.

The next step is to remove any voxels of the dilated segmented image that do not belong to the ossicles. This is achieved by making an AND logical operation of the dilated image with a binary image obtained from the original image under study by image

threshold. In the tests performed, the original images were binarized using a threshold level equal to 150 Hounsfield units (HU). The threshold level used had a low value (usually, bones have HU values > 1000)[16] because the structures to be segmented are very small compared to the resolution of the conventional CT systems, which usually cause severe partial volume effects around the structures.

Then, a closing morphological operation, using a 3D square kernel of approximately 0.3 mm wide, is performed on the axial plane to smooth the edges of the segmented structures. Finally, a connected component filter is used on the segmented image to remove any possible scattered voxels or disconnected clusters of voxels.

(insert Figure 2 about here)

2.2.1 Registration algorithm

The inter-subject registration of images representing the human ear based on conventional CT images is a challenging task, mainly for three reasons: 1) the small size of the structures involved in relation to the resolution of the images, 2) the large variety of the shapes to be segmented among subjects, and 3) the different fields of view associated to the images. To overcome these difficulties, a dedicated algorithm was developed to perform the inter-subject registration of anatomical structures of the ear.

Let's consider a reference image, i.e., the image that remains unchanged, and a moving image that should be geometrically transformed to match the reference image. The registration algorithm developed can be described by the following steps: 1) Computation of an initial shift correction along the three axes; 2) Computation of an optimal rigid transformation composed of a shift along the three axes and three rotation angles; 3) Computation of an optimal affine transformation that included 9 parameters

for the deformation and three for the shift; 4) Registration of the two input images using a curved geometric transformation based on a free-form deformation model.

Then, the geometric transformation found in Step 1 is used as the initial solution for the optimization algorithm in Step 2; the transformation found in Step 2 is then used as the initial solution in Step 3, and this is repeated until the Stop criterion is achieved.

It should be noted that the affine registration, conducted after the rigid registration, is used to compensate the small dimensions of the structures under study and thus strengthens the robustness of the registration process. Also Step 1 of the registration algorithm is fundamental to identify the ear in the volumetric CT image under evaluation.

The optimal shift is found by minimizing the sum of squared differences (SSD) between the voxel intensities of the two input images. The images to be registered may require different fields of view. Therefore, to have a good initial optimal shift, the SSD must only be computed in a region of interest (ROI). This ROI was defined in the template image and remained constant for any image to be segmented. Hence, the optimal shift (a) is found by minimizing:

$$SSD(a) = \sum_i \left(r(i) - m(i-a)\right)^2 w(i), \qquad (1)$$

where r is the reference image, m is the moving image, and w is the ROI mask over the template image. The strategy to optimize this cost function is to exhaustively compute the SSD for all integer shifts and then choose the one that gives the smallest value. However, this direct computation for all cost function values is too expensive. Thus, these values are computed by transforming the cost function into a sum of convolutions that are evaluated in the frequency domain.[17] Thus, Equation (1) can be expanded to:

$$SSD(a) = \sum_i r(i)^2 w(i) - 2\sum_i \left(r(i)w(i)\right)\overline{m}(a-i) + \sum_i w(i)\overline{m}(a-i)^2 , \qquad (2)$$

where $\overline{m}(x) = m(-x)$. The second and third sums in Equation (2) can be efficiently evaluated in the frequency domain using the Fast Fourier transform (FFT) since they represent convolutions. The first sum can be discarded since it does not depend on *a*, and the goal is to find the shift *a* that minimizes the SSD. It should be noted that, since the SSD is computed only in a ROI containing only a small volume around the middle ear of the template image, the problem of different fields of view can be overcome.

The iterative optimization algorithm used searches for the parameters of the geometric transformation that optimize the image similarity measure under consideration. To enlarge the radius of convergence of the optimization algorithm, the optimal rigid transform is found using a multi-resolution or multi-scale strategy.

The curved registration is based on a free-form deformation model, with a regular grid of control points defined in order to determine the deformation involved. To control the deformation, the points of this grid are moved individually in the direction that optimizes the similarity measure used, thus attaining the local deformations. Grid points with 8 and 23 points along each axis were defined for the coarse and fine curved registrations, respectively. The deformation is propagated to all voxels by interpolation using cubic B-splines.[18] Thus, the coarse curved registration copes with the large deformations, while the fine curved registration deals with the small deformations. This solution, based on a coarse model with few grid points followed by a fine model with a higher number of grid points, is a multi-scale approach that always copes with all voxels of the image to be registered.

Two optimization algorithms are used: Powell's method in Steps 2 and 3, and a gradient descent based algorithm in Step 4.[19] In both cases, the goal of the optimization algorithm is to search for the parameters of the geometric transformations that maximize the similarity measure based on mutual information (MI):

$$MI = H(r) + H(m) - H(r,m), \quad (3)$$

where $H(r)$ and $H(m)$ are the Shannon's entropy of the voxel intensity in the reference and moving images, respectively; and $H(r,m)$ is the Shannon's entropy of the joint histogram of the reference and moving images.[20]

Powell's method is used in Steps 2 and 3, i.e. to search for the optimal rigid and affine transformations, since these transformations are global and thus a robust optimization algorithm is crucial to guaranty convergence.

A gradient descent based optimization algorithm is used in Step 4 to search for the optimal curved transformation, because the free-form deformation model used is defined by a large number of parameters, and thus a fast optimization algorithm should be used. The gradient descent based optimization algorithm is not as robust as Powell's method, but, since the influence of each parameter of the geometric transformation is local, it is sufficient to guarantee suitable registration results.

As for the SSD, the MI similarity measure used in Steps 2, 3 and 4 is computed only in a small ROI containing the ear, so that structures not belonging to the ear do not have any influence in the registration process.

2.2.2 Template image

Two template images were defined to represent a mean human ear of a non pathologic population: one for the right ear and the other for the left ear. It should be noted that the template images were built just once.

To build the left template, we first selected a set of 8 images, without perceptible ear pathologies confirmed by visual inspection, from the dataset with voxel dimensions varying from 0.164×0.164×0.6 to 0.166×0.166×1.0 mm^3. Five of the selected images had a slice spacing equal to 1.0 mm, and the other 3 had a slice spacing of 0.6 mm.

Then, an initial reference image was randomly chosen from the set of 8 images selected, and a mean image was built by rigid registration of this image with the remaining 7 images. Afterwards, all selected images were again registered, but now using the mean image previously built as the reference image. This process was repeated until no significant differences were found between the successive mean images.

A similar procedure was performed to obtain the mean image based on affine geometric transformation and the mean image using curved geometric transformation. In the initial step of the process of obtaining the mean image based on affine transformation, the mean image obtained using just rigid transformation was used as the reference image. Similarly, to obtain the mean image based on curved transformation, the mean image obtained using the affine transformation was used in the initial step as the reference image.

In the affine registration process, the registration algorithm started by computing the shift transformation, then the rigid geometric transformation and only then the affine geometric transformation. For the curved registration, the registration algorithm followed the same steps, and computed the curved transformation after the computation of the affine transformation. The last mean image obtained by using the curved geometric transformation was then considered as the template image for the registration-based segmentation process.

In the process of building the right template image, all registered images were reconstructed using cubic voxels with spacing equal to 0.15 mm in the three directions, resulting in a template image with the same voxel spacings. The template image was built with a voxel size smaller than the voxel size of the images selected to build it to avoid losing any relevant information, and linear interpolation was employed in the image resampling.

Although the template image for the right ear could be obtained by the same process as for the left ear, we built it by using the mirror of the template image defined for the left ear, since it is a common practice in atlas building in order to avoid biased data.[21]

It should be noted that we only used 8 of the 21 images of the dataset to build the template image. Thus, there should be no bias in the segmentation results due to the randomly selected images for the template image.

2.2.3 Segmented template image

The segmented template image for the left ear was obtained by the manual segmentation of the template image defined for the left ear using the ITK-SNAP 2.2.0 software (www.itksnap.org).[22] Using this software, an expert defined the contours that embraced the incus and malleus by drawing a polygonal line, first on the axial plane and then on the sagittal and coronal planes to improve the segmentation. The segmented template image for the right ear was then obtained by mirroring the segmented image defined for the left side.

2.3 Accuracy assessment

The accuracy of this new registration-based segmentation solution was tested against the segmentation results obtained manually by two experts. Two similarity measures, Dice's coefficient and the Hausdorff distance, were used to compare the shapes attained automatically and manually.

The Dice's coefficient, *d*, assesses the overlapping of two shapes:

$$d(V_m, V_a) = \frac{2|V_m \cap V_a|}{|V_m| + |V_a|}, \quad (4)$$

where V_m and V_a are the coordinates of non-zero voxels of the shapes, obtained manually and automatically, respectively; and the operator $|.|$ gives the number of non-zero voxels. Thus, if the two shapes are the same, *d* is equal to 1 (one) and if they are totally different, *d* is equal to 0 (zero). Thus, higher values of *d* represent better agreement between the segmented shapes.

The Hausdorff distance, *h*, evaluates how far two subsets of a metric space are from each other. That is, *h* is the maximum distance of all possible distances between any point on a shape to the closest point on a second shape, and is given as:

$$h(V_m, V_a) = \max\left\{\max_{X \in V_m}\left\{\min_{Y \in V_a}\left\{\|X - Y\|\right\}\right\}, \max_{Y \in V_a}\left\{\min_{X \in V_m}\left\{\|Y - X\|\right\}\right\}\right\}, \quad (5)$$

where *X* and *Y* are the coordinates of the voxels in the sets V_m and V_a, respectively, and here $\|.\|$ represents the Euclidian distance.

For visual evaluation purposes, two strategies were employed. The first one was based on the segmented images overlapping the corresponding original images. The other strategy was based on the building of the incus and malleus ossicles surfaces using the marching cubes algorithm.[23] Figures 3 and 4 show an example of a CT image set and the surface that was built, based on the segmentation of those images.

2.3.1 Manual segmentation

All images of the experimental dataset were manually segmented by two experts using the ITK-SNAP 2.2.0 software, by drawing a polygonal line on the axial plane. In the cases with unclear structural borders, the experts drew the contour based on their knowledge of the structures. After the manual segmentation, the binary images obtained were morphologically smoothed using a morphological closing filter approximately 0.3 mm wide to ensure the expected smoothness of the bones.

2.4 Implementation

The computational solution developed was fully implemented in C++ and tested on a notebook PC with an Intel I7-2670QM microprocessor, 6 GB of RAM, and running Microsoft Windows 7. The techniques of image processing, registration and visualization were implemented using the following free open source C++ libraries: CImg Library (http://cimg.sourceforge.net), Insight Toolkit (ITK) (www.itk.org) and Visualization Toolkit (VTK) (www.vtk.org).

3. Results

The Kolmogorov-Smirnov test indicated that the possibility of both the Dice's coefficient and Hausdorff distance following normal distributions could not be excluded (p = 0.571). Thus, to evaluate the significance of the difference between the groups, the Student's *t*-test was used.

Dice's coefficients, which were computed in order to compare the results for the fully automatically and manually segmented incus and malleus, are shown in Table 1. No significant differences were found for this coefficient in terms of the ear side (p = 0.270).

Although, as expected, the mean value of Dice's coefficient obtained for the CT images acquired with the lower slice spacing (0.6 mm) was higher than the one for the CT images acquired with a higher slice spacing (1.0 mm), the difference between the mean Dice's coefficients was not statistically significant (p = 0.237). The rendered surfaces based on the CT images acquired with a slice spacing of 1.0 mm were visually rough and small details were lost in some cases. In contrast, with the CT images acquired with a slice spacing equal to 0.6 mm, the surfaces obtained were reasonably smooth and

more fine details could be perceived (Figures 3 and 4). In terms of Dice´s coefficient, the lower results were obtained in two images with soft mass located around and over the ossicles.

The mean Hausdorff distance values calculated between the segmented images obtained using the proposed automated solution and those obtained manually by the experts are shown in Table 2. The mean Hausdorff distance calculated from the segmented shapes from the CT images acquired with a slice spacing of 0.6 mm was lower than the one acquired with a slice spacing of 1.0 mm (p = 0.012). The Maximum Hausdorff distance for the images with a slice spacing of 1.0 mm was 1.53 mm and the 95th percentile was 1.34 mm according to both experts.

To analyse the fact that the template image had been built with images from the dataset under evaluation, we computed Dice's coefficient and Hausdorf distance after removing those images. The results obtained were: 0.955±0.0221 and 1.102±0.1597 mm for Dice's coefficient and Hausdorf distance, respectively, which were very close to the ones obtained when the full dataset was used.

The accuracy of the segmentation solution was also visually evaluated by placing each segmented image over the original image using a tool of the ITK-SNAP software that controls the transparency of the segmented image and thus facilitates the visualization (Figure 5). In all cases, the segmented images obtained matched the structures of interest in the corresponding original image reasonably well.

With the implementation developed and the parameters used, the total computational time for the automated segmentation was between 2 and 3 minutes for each 3D CT image; most of this time was required for the registration process.

(insert Figures 3, 4 and 5, and Tables 1 and 2 about here)

4. Discussion and Conclusions

The results had a good agreement between the segmented images obtained by the proposed automated method and the segmented images attained manually by the experts; the mean Dice's coefficient was approximately equal to 0.96. This is even more significant since the images were acquired using a conventional CT scanner and standard protocols as well as in some cases the ossicles were partially occluded.

No statistically significant differences were found between the Dice's coefficients obtained from the CT images acquired with a slice spacing of 0.6 mm, and the ones acquired with a slice spacing of 1.0 mm. However, the Hausdorff distance revealed a higher agreement for the CT images acquired with a slice spacing of 0.6 mm than for the ones acquired with a slice spacing of 1.0 mm (p = 0.012). This confirms that the segmented images obtained automatically by the proposed computational solution and the segmented images obtained manually by the experts are very similar for both subsets of images. In fact, the difference found in the Hausdorff distance between the values obtained for the two image subsets is mainly explained by the slice spacing associated to each subset. It should also be noted that, for 3D CT images acquired with a slice spacing of 1.0 mm, the Euclidean distance between two neighbouring voxels can be greater than 1.0 mm.

The results obtained, when the images used to build the template image were excluded from the dataset, are almost equal to the ones obtained using the full dataset. This confirms that there was not a bias in the results due to the fact that the template image was built using images from the same dataset.

Although two previous works were found on the automated segmentation of the ossicles,[15, 24] the results cannot be compared. In the work of Hsieh and co-workers there

are no results comparing their approach with a ground truth.[24] Noble and co-workers used a dataset with a significantly lower slice spacing than the slice spacings of the dataset used in this work.[15] Moreover, the dataset used here included images with ossicles partially occluded. However, just for a rough comparison, a maximum Hausdorf distance of 1.53 mm was obtained with our solution, while Noble and co-workers report a maximum value of 1.77 mm using a dataset composed by images with a resolution of $0.35 \times 0.35 \times 0.4$ mm^3.[15]

Initially, we had the intention of comparing our automated results with ones obtained using standard semi-automated segmentation methods. However, none of the solutions tested, including algorithms based on statistical confidence, region growing, level set and active contours, worked satisfactory in all of the images.

Also we had an intention to segment the stapes bone in the CT images, but it was abandoned since the stapes were not visible in most of the images acquired with a slice spacing of 1.0 mm. On the other hand, the incus and malleus were segmented as a group because they were overlapping, and in most of the images of the dataset used, the experts could not detect the individual borders successfully.

The computational solution was also tested on coronal conventional CT images of the ear. In all cases, the solution performed the segmentation quite well. However, since those images had a large spacing between slices, they produced very rough surfaces and thus were discarded.

The solution adopted to diminish the influence of the different fields of view in the registration accuracy worked very well, since all images, including images of the complete head, were automatically and accurately registered with the template image.

The computational framework developed can be used to segment other structures of the ear in images, such as the cochlea and semicircular canals. With an effective

implementation, the segmentation of these structures can be carried out simultaneously with the segmentation of the incus and malleus ossicles by using the same geometric transformation computed for the segmentation of the ossicles.

The actual implementation of our computational solution results from several tests and experiments. For instance, in the registration process, besides the mutual information measure, the cross-correlation and intensity-based mean squared difference measures were also tested; different numbers of bins to compute the mutual information were analysed; and different optimization parameters, initialization strategies, ROIs defined to compute the similarity measure were used and studied.

In conclusion, the fully automated segmentation of the incus and malleus ossicles in 3D images acquired using conventional CT scanners and standard protocols was shown to be feasible and robust, even for patients with ear pathologies and when the ossicles are partially occluded. Thus, it can be used to help radiologists and otolaryngologists evaluating the bi-dimensional slices by adding useful 3D surface models built automatically from the patients CT data.

Acknowledgements

The authors are grateful to HPP Medicina Molecular, in Portugal, for providing the experimental data used.

This work was partially done in the scope of the project with reference PTDC/BBB-BMD/3088/2012, financially supported by Fundação para a Ciência e a Tecnologia (FCT) in Portugal.

References

1. Gentil F, Parente M, Martins P, et al. The influence of muscles activation on the dynamical behaviour of the tympano-ossicular system of the middle ear. *Comput Method Biomech Biomed Eng* 2013; 16(4): 392-402.

2. Gentil F, Parente M, Martins P, et al. The influence of the mechanical behaviour of the middle ear ligaments: a finite element analysis. *Proc IMechE Part H: Journal of Engineering in Medicine* 2011; 225(1): 68-76.

3. Gan RZ, Reeves BP, Wang X. Modeling of sound transmission from ear canal to cochlea. *Ann Biomed Eng* 2007; 35(12): 2180-95.

4. Majdani O, Rau TS, Baron S, et al. A robot-guided minimally invasive approach for cochlear implant surgery: preliminary results of a temporal bone study. *Int J Comput Assisted Radiol Surg* 2009; 4: 475-86.

5. Gerber N, Bell B, Gavaghan K, et al. Surgical planning tool for robotically assisted hearing aid implantation. *Int J Comput Assisted Radiol Surg* 2014; 9(1): 11-20.

6. Crescenzio FD, Fantini M, Ciocca L, et al. Design and manufacturing of ear prosthesis by means of rapid prototyping technology. *Proc IMechE Part H: Journal of Engineering in Medicine* 2011; 225(3): 296-302.

7. Counter P. Implantable hearing aids. *Proc IMechE Part H: Journal of Engineering in Medicine* 2008; 222(6): 837-52.

8. Ferreira A, Gentil F, Tavares JMRS. Segmentation algorithms for ear image data towards biomechanical studies. *Comput Method Biomech Biomed Eng* 2014; 17(8): 888-904.

9. Oliveira FPM, Tavares JMRS. Medical image registration: a review. *Comput Method Biomech Biomed Eng* 2014; 17(2): 73-93.

10. Bradshaw AP, Curthoys IS, Todd MJ, et al. A mathematical model of human semicircular canal geometry: a new basis for interpreting vestibular physiology. *J Assoc Res Otolaryngol* 2010; 11: 145-59.

11. Lee DH, Chan S, Salisbury C, et al. Reconstruction and exploration of virtual middle-ear models derived from micro-CT datasets. *Hear Res* 2010; 263: 198-203.

12. Christensen GE, He J, Dill JA, et al. Automatic measurement of the labyrinth using image registration and a deformable inner ear atlas. *Acad Radiol* 2003; 10(9): 988-99.

13. Noble JH, Labadie RF, Majdani O, et al. Automatic segmentation of intracochlear anatomy in conventional CT. *IEEE Trans Biomed Eng* 2011; 58(9): 2625-32.

14. Noble JH, Warren FM, Labadie RF, et al. Automatic segmentation of the facial nerve and chorda tympani in CT images using spatially dependent feature values. *Med Phys* 2008; 35(12): 5375-84.

15. Noble JH, Dawant BM, Warren FM, et al. Automatic identification and 3D rendering of temporal bone anatomy. *Otol Neurotol* 2009; 30: 436-42.

16. Feeman TG. *The Mathematics of Medical Imaging: A Beginner's Guide*. New York: Springer, 2010.

17. Orchard J. Efficient least squares multimodal registration with a globally exhaustive alignment search. *IEEE Trans Image Process* 2007; 16(10): 2526-34.

18. Rueckert D, Sonoda LI, Hayes C, et al. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans Med Imaging* 1999; 18(8): 712-21.

19. Press WH, Teukolsky SA, Vetterling WT, et al. *Numerical Recipes: The Art of Scientific Computing*. 3rd ed. New York: Cambridge University Press, 2007.

20. Collignon A, Maes F, Vandermeulen D, et al. Multimodality medical image registration by maximization of mutual information. *IEEE Trans Med Imaging* 1997; 16(2): 187-98.

21. Grabner G, Janke AL, Budge MM, et al. Symmetric atlasing and model based segmentation: an application to the hippocampus in older adults. In: *9th International Conference on Medical Image Computing and Computer Assisted Intervention - MICCAI 2006* (ed Larsen R, Nielsen M, Sporring J); 2006 October 1-6; Copenhagen, Denmark: Springer-Verlag.

22. Yushkevich PA, Piven J, Hazlett HC, et al. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage* 2006; 31(3): 1116-28.

23. Lorensen WE, Cline HE. Marching Cubes: a high resolution 3D surface construction algorithm. *Comput Graph* 1987; 21(4): 163-9.

24. Hsieh M-S, Lee F-P, Tsai M-D. A virtual reality ear ossicle surgery simulator using three-dimensional computer tomography. *J Med Biol Eng* 2010; 30(1): 57-63.

FIGURE CAPTIONS

Figure 1: Example of two slices from two patients included in the dataset. (The arrows indicate the soft mass around and over the ossicles.)

Figure 2: Flowchart of the automated segmentation solution proposed.

Figure 3: Eight consecutive slices of a conventional 3D CT image of the ear with a voxel size of $0.164 \times 0.164 \times 0.6$ mm^3. (The incus and malleus ossicles are inside the ellipses drawn in each slice. Only the areas of interest of each slice are shown.)

Figure 4: 3D visualization of the incus and malleus surface built from the automated segmentation of the 3D CT image shown in Fig. 3. (This surface was built using the marching cubes technique included in the computational framework developed.)

Figure 5: Visualization of an example of the segmented image obtained by the automated framework developed. (The segmented image is represented in red colour and is on top of the original. Three views and the surface reconstructed using ITK-SNAP software are depicted.)

TABLE CAPTIONS

Table 1: Dice's coefficients between the shapes automatically and manually segmented (mean ± standard deviation).

Table 2: Hausdorff distance in millimetres between the shapes automatically and manually segmented (mean ± standard deviation).

TABLES

Table 1

	Automated versus Expert 1	Automated versus Expert 2	Average (automated versus Expert 1, automated versus Expert 2)
All images	0.958±0.0195	0.954±0.0194	0.956±0.0187
Images acquired with a slice spacing of 0.6 mm (9 images)	0.962±0.0230	0.958±0.0235	0.960±0.0230
Images acquired with a slice spacing of 1.0 mm (12 images)	0.956±0.0161	0.951±0.0148	0.953±0.0139

Table 2

	Automated versus Expert 1	Automated versus Expert 2	Average (automated versus Expert 1, automated versus Expert 2)
All images	1.14±0.192	1.13±0.187	1.13±0.181
Images acquired with a slice spacing of 0.6 mm (9 images)	1.02±0.202	1.03±0.199	1.02±0.188
Images acquired with a slice spacing of 1.0 mm (12 images)	1.22±0.126	1.21±0.128	1.22±0.120

FIGURES

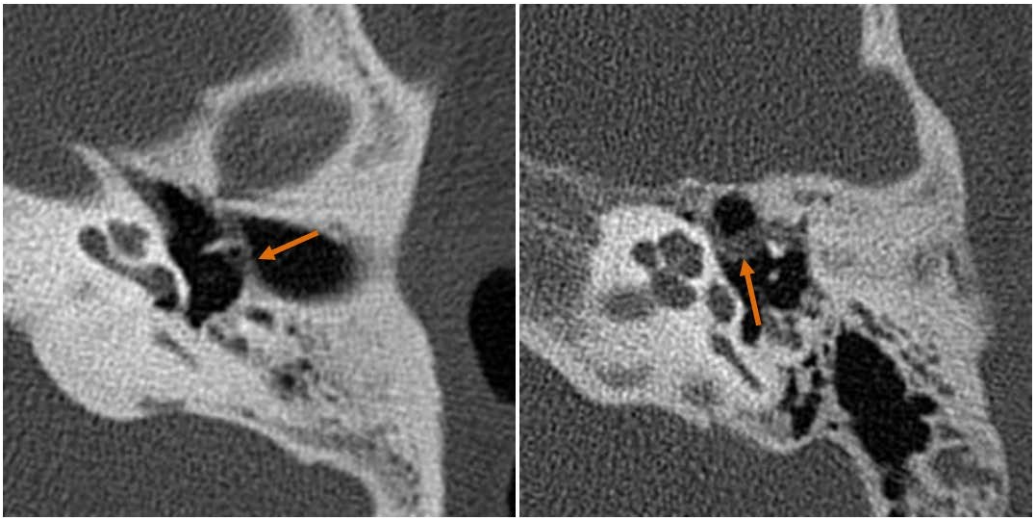

Figure 1

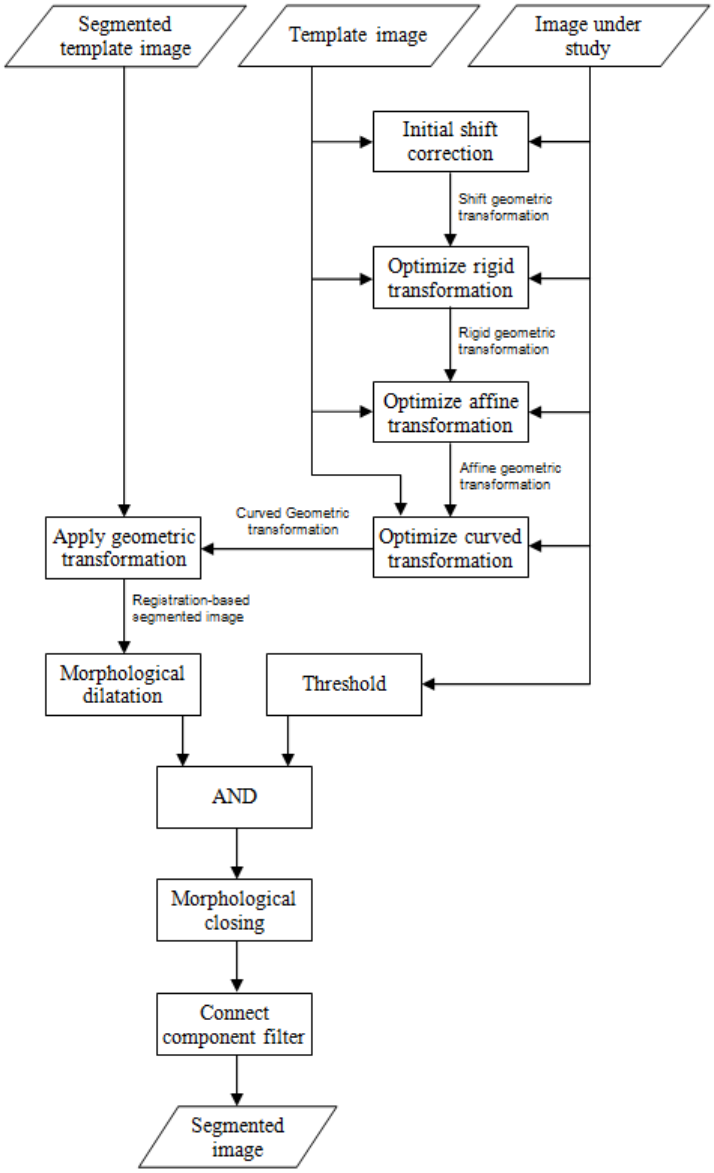
Segmented template image
Template image
Image under study
Initial shift correction
Shift geometric transformation
Optimize rigid transformation
Rigid geometric transformation
Optimize affine transformation
Affine geometric transformation
Curved Geometric transformation
Apply geometric transformation
Optimize curved transformation
Registration-based segmented image
Morphological dilatation
Threshold
AND
Morphological closing
Connect component filter
Segmented image

Figure 2

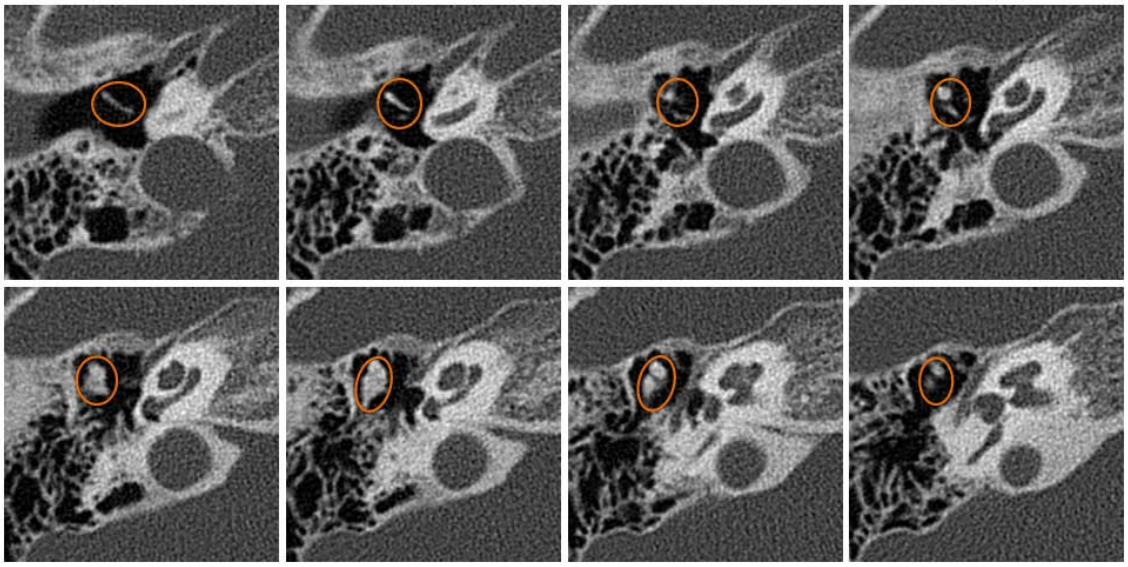

Figure 3

Figure 4

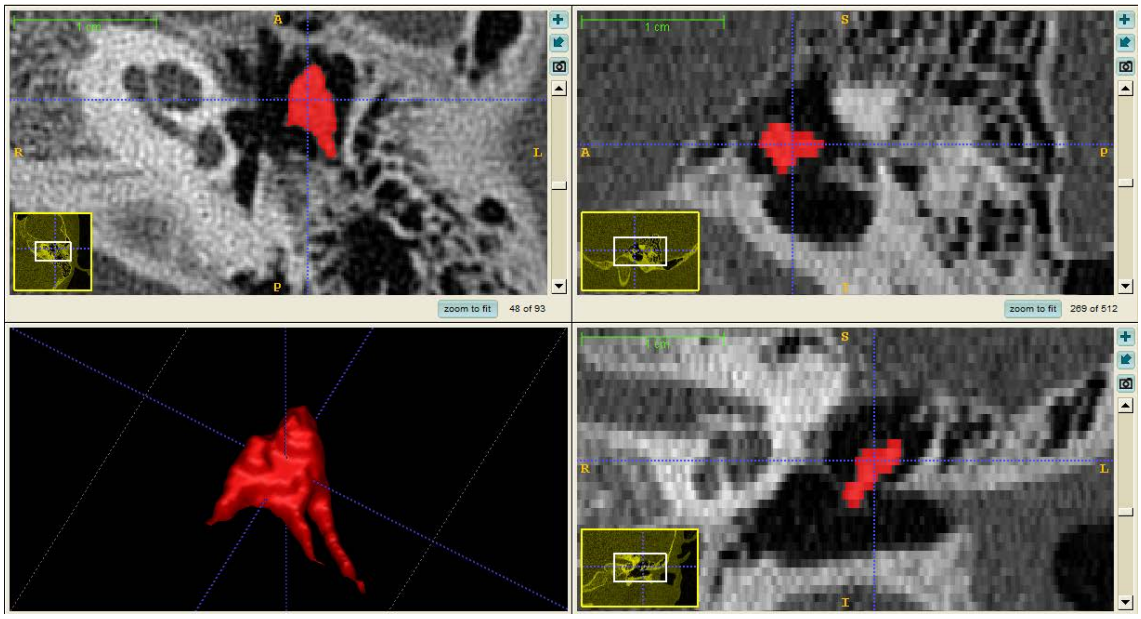
1 cm
A
R
L
P
zoom to fit
48 of 93
1 cm
S
A
P
I
zoom to fit
269 of 512
S
R
L
I

Figure 5