

RESEARCH ARTICLE

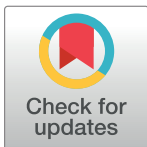
Automatic intra-subject registration and fusion of multimodal cochlea 3D clinical images

Ibraheem Al-Dhamari^{1*}, Rania Helal², Olesia Morozova¹, Tougan Abdelaziz², Roland Jacob³, Dietrich Paulus¹, Stephan Waldeck⁴

1 Computer Vision Department, Koblenz University, Koblenz, Germany, **2** Radiodiagnosis Dept., Ain Shams University, Cairo, Egypt, **3** HNO Plus, Koblenz, Germany, **4** Interventional Radiology and Neuroradiology Dept., Military Hospital Koblenz, Koblenz, Germany

✉ These authors contributed equally to this work.

* idthamari@uni-koblenz.de



OPEN ACCESS

Citation: Al-Dhamari I, Helal R, Morozova O, Abdelaziz T, Jacob R, Paulus D, et al. (2022) Automatic intra-subject registration and fusion of multimodal cochlea 3D clinical images. PLoS ONE 17(3): e0264449. <https://doi.org/10.1371/journal.pone.0264449>

Editor: Kumaradevan Punithakumar, University of Alberta, CANADA

Received: August 4, 2021

Accepted: February 11, 2022

Published: March 2, 2022

Copyright: © 2022 Al-Dhamari et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The proposed method's source code is provided as a public open-source and can be downloaded from a public server in 3D Slicer Software through which any clinical temporal bone imaging can be tested on. (download link: <https://github.com/MedicalImageAnalysisTutorials/SlicerCochlea>) This enables other researchers to reproduce these experiments and for people to benefit from the results. Also the minimal dataset including all the original results tables are added as supplementary material to the paper (publicly available). The

Abstract

Background

The postoperative imaging assessment of Cochlear Implant (CI) patients is imperative. The main obstacle is that Magnetic Resonance imaging (MR) is contraindicated or hindered by significant artefacts in most cases with CIs. This study describes an automatic cochlear image registration and fusion method that aims to help radiologists and surgeons to process pre- and postoperative 3D multimodal imaging studies in cochlear implant (CI) patients.

Methods and findings

We propose a new registration method, Automatic Cochlea Image Registration (ACIR-v3), which uses a stochastic quasi-Newton optimiser with a mutual information metric to find 3D rigid transform parameters for registration of preoperative and postoperative CI imaging. The method was tested against a clinical cochlear imaging dataset that contains 131 multimodal 3D imaging studies of 41 CI patients with preoperative and postoperative images. The preoperative images were MR, Multidetector Computed Tomography (MDCT) or Cone Beam Computed Tomography (CBCT) while the postoperative were CBCT. The average root mean squared error of ACIR-v3 method was 0.41 mm with a standard deviation of 0.39 mm. The results were evaluated quantitatively using the mean squared error of two 3D landmarks located manually by two neuroradiology experts in each image and compared to other previously known registration methods, e.g. Fast Preconditioner Stochastic Gradient Descent, in terms of accuracy and speed.

Conclusions

Our method, ACIR-v3, produces high resolution images in the postoperative stage and allows for visualisation of the accurate anatomical details of the MRI with the absence of sig-

clinical imaging studies used (not part of the minimal dataset) can't be publicly available because it is owned by the BWZK hospital. The medical image data and the 3D landmark points can be requested by email to Mrs. Ringsmut (BwZKrhsKoblenzKlinikVIIIRadiologie@bundeswehr.org). Personalized data can not be made publicly available due to data protection laws.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

nificant metallic artefacts. The method is implemented as an open-source plugin for 3D Slicer tool.

Introduction

The cochlea is a principal part of the inner ear that plays a crucial role in hearing, filtering frequency coded auditory signals, and transmitting them to the brain. The number of patients presenting with sensorineural hearing loss has increased over the years and Cochlear Implantation (CI) is gaining popularity as a treatment option for these patients. The electrode array in CIs simulate the function of the cochlea, which in many cases allows patients to communicate with others and enjoy a normal social life.

The preoperative CI imaging examination relies on Multidetector CT (MDCT) or more recently the Cone Beam CT (CBCT) for the assessment of the bony labyrinth. The Magnetic Resonance imaging (MR) is used mainly for evaluation of the membranous labyrinth and the intracranial vestibulocochlear nerve.

However, in the postoperative period, MR is contraindicated in most cases with CIs, and even if performed, assessment is impossible due to the pronounced metallic artefact of the CI electrode within the cochlea. Therefore postoperative assessment relies entirely on computed tomographic imaging (including MDCT or CBCT) or the conventional projection radiography (X-ray) [1]. In the postoperative stage, it is important to accurately assess the position of the intracochlear electrode in the scala tympani to rule out intracochlear malpositioning. MDCT and CBCT can only visualise the bony cochlea; a direct assessment of the cochlea scalae is only possible to a limited extent. Multimodal image fusion between preoperative MR imaging and postoperative MDCT/CBCT examination can provide a solution to this problem.

Image fusion combines different images to create a single more informative image. The new fused image has features from all the input images e.g. one can see the bones and the soft tissue simultaneously from fused CT and MR images, see Fig 1. There are many methods of image fusion [2]. The method we used is a simple matrix addition which is a very fast operation that takes only a small part of a second in today's computers e.g.:

$$I_{fused} = I_A + I_B \quad (1)$$

For an efficient fusion, images must be correctly aligned to the same physical space using image registration. Manually performing this process takes a lot of time and neuroradiological expertise. Automating this process using an image registration method not only saves a huge amount of time but can also significantly improve the repeatable quality of the results.

Image registration

The image registration [3–5] is the problem of finding a transformation T that aligns one or more images to a reference image. This transformation transforms the points of an image to the same location of the points in the reference image. Intra-subject medical image registration aligns images of the same patient while inter-subject registration aligns images of different patients. Fig 2 shows the main registration components: a fixed and a moving image, a cost function, an optimiser, and a transformation. The first transformation handles the issue of different resolutions by mapping both input images to a virtual domain in the images' physical

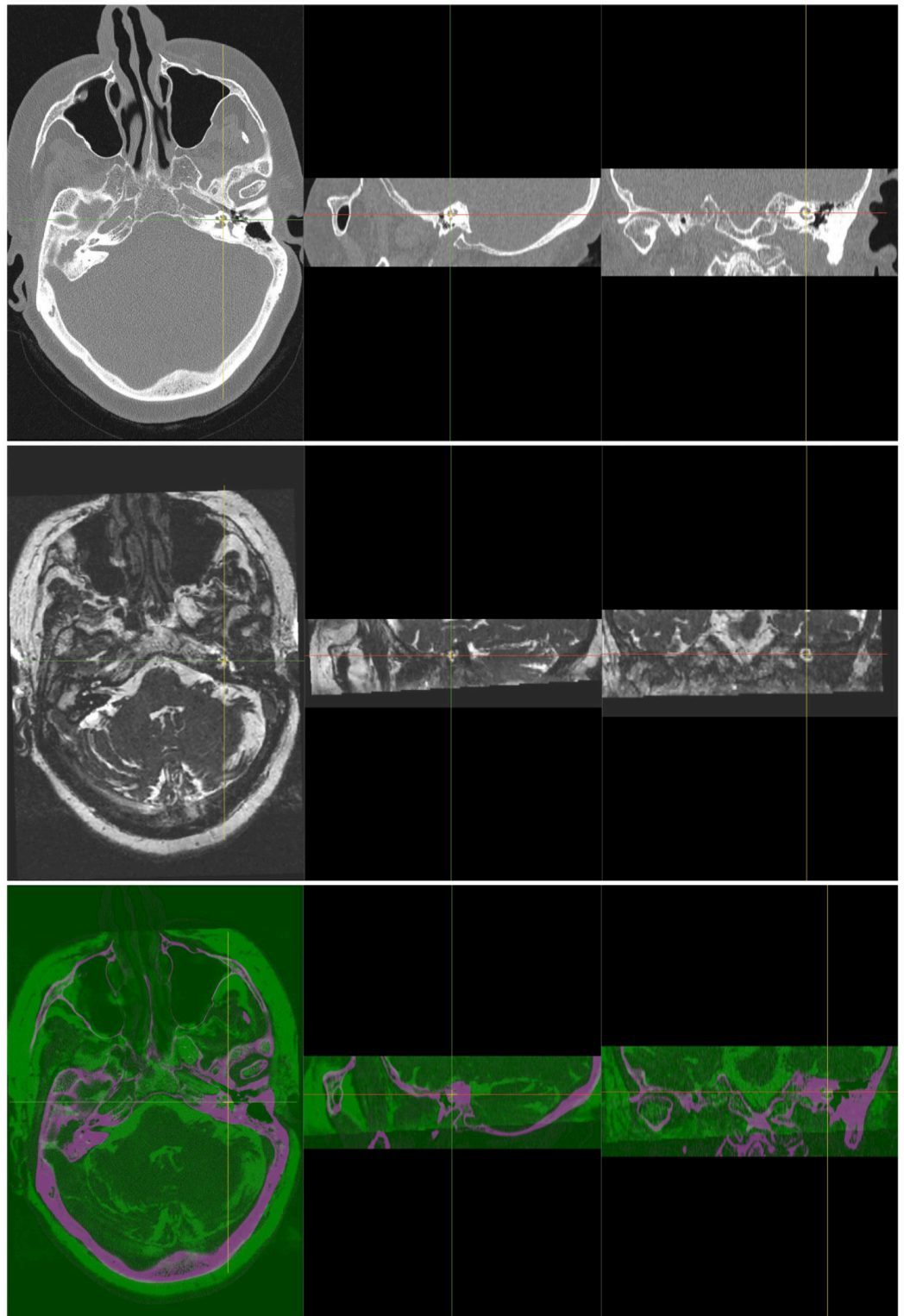


Fig 1. Two input 3D cochlea images of the same patient, CT as a fixed image (top) and MR as a moving image (middle). The bottom image is the registered and fused 3D image, with magenta colour representing the MR part and green colour representing the CT part. Each image has 3 views, from left to right: axial, sagittal and coronal.

<https://doi.org/10.1371/journal.pone.0264449.g001>

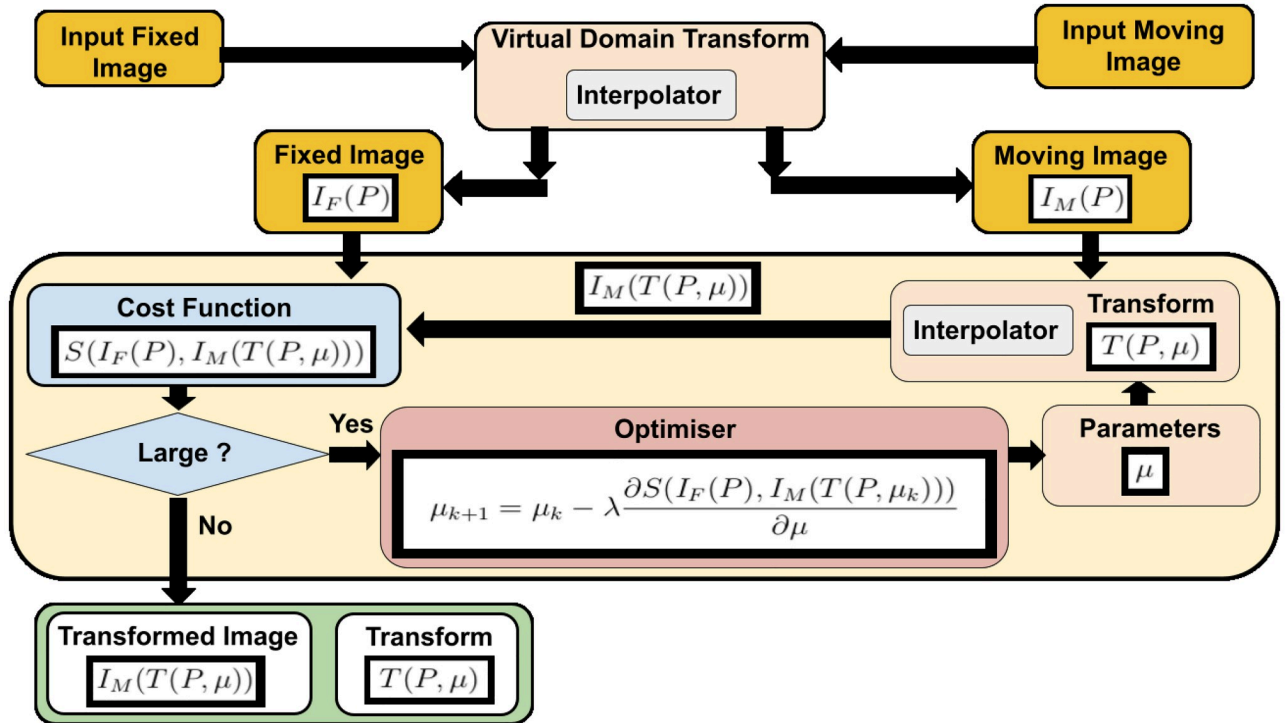


Fig 2. The main components of image registration: Input images, cost function, optimiser, and a transform. The virtual domain transform deals with the issue of different image resolution.

<https://doi.org/10.1371/journal.pone.0264449.g002>

space. This takes place before the optimisation process starts. In this implementation, medical image registration is not performed in the images' space [6].

Mathematically, the transform is a function of the image points P and the transform's parameters μ where it takes an image point (or a vector of points) and outputs the image transformed point (or a vector of transformed points). We can write this transform as $T(P, \mu)$ where P is the image coordinates vector. The second input to the transformation's function T is μ which is a vector that contains the transformation's parameters.

The transformation transforms an image called a moving image $I_M(P)$ to another image called a fixed image $I_F(P)$, where P is a vector of all image point coordinates. The image itself can be considered a function of points where the output is the intensity (or color) values of the point locations.

Finding the parameters of this transformation is a challenging task. Hence, the general registration problem is still unsolved and many papers are published every year trying to solve different aspects of this problem.

Optimisation techniques [7] are usually used to find these parameters. An optimiser, such as Gradient Descent (GD) [8], tries to find these parameters by minimising a cost function of the input images. Popular image registration cost functions use similarity metrics such as Mean Squared Error $S_{MSE}(I_F, I_M)$ [9] or Mutual Information $S_{MI}(I_F, I_M)$ [10]. Note that by minimising the similarity metric value, we are maximizing the similarity between the fixed and the transformed moving images.

The optimisation problem is stated as follows:

$$\tilde{\mu} = \arg \min S(I_F(P), I_M(T(P, \mu))), \tag{2}$$

where μ is the transformation's parameters' vector, I_F and I_M are the fixed and the moving images, P is the points' coordinate vector, S is the similarity metric function, and $T(P, \mu)$ is the transformation's function.

In the GD optimiser, the new parameters are updated using the derivative of the cost function with respect to the transformation's parameters, e.g.:

$$\mu_{k+1} = \mu_k - \lambda \frac{\partial S(I_F(P), I_M(T(P, \mu_k)))}{\partial \mu}, \quad (3)$$

where k is the optimisation iteration, λ represents the step-size or the learning rate of the GD. The derivative of the similarity metric with respect to the transformation's parameters is required for Eq (3). Using the chain rule, this derivative can be divided into two main parts:

$$\frac{\partial S}{\partial \mu} = \sum_{j=0}^{n-1} \left(\frac{\partial S(I_F(p_j), I_M(T(p_j, \mu_k)))}{\partial T(p_j)} \frac{\partial T(p_j, \mu_k)}{\partial \mu} \right), \quad (4)$$

where $\frac{\partial S}{\partial \mu}$ is the first order derivative of the similarity metric S with respect to the transformation's parameters μ at k optimisation's iteration, p_j is an image pixel coordinate, and n is the total number of the image's pixels. The first part of the right-hand side in Eq (4) requires the derivative of the similarity metric and the transformed moving image. The second part $\frac{\partial T(p_j, \mu_k)}{\partial \mu}$ is called J_μ and it is similar to the Jacobian matrix of the transformation as it represents the transformation's partial derivatives at each image point with respect to the transformation's parameters. The term J_μ is computed at each optimisation iteration. For a large image size and complex transformation, the computation requires a lot of time to complete. To solve this issue, the Stochastic Gradient Descent (SGD) optimisation approach [11, 12] uses samples instead of the whole image. Sometimes these samples are taken randomly. The stochastic approach allows fewer computations and provides better results in practice. The parameters updating rule of SGD is:

$$\mu_{k+1} = \mu_k - \lambda \frac{\partial S(\hat{I}_F(P), \hat{I}_M(T(P, \mu_k)))}{\partial \mu}, \quad (5)$$

where \hat{I}_F and \hat{I}_M represent samples from the input images I_F and I_M .

Aim of the work

The aim of this work is to propose and evaluate a new relatively easy and fast automatic cochlear image registration and fusion method utilizing simple computer hardware and software to align and fuse pre-and postoperative multimodal cochlear images in cochlear implant (CI) patients.

Materials and methods

Ethics statement

The use of all human datasets in our experiments were approved by the Ethics Committee of the Rheinland-Pfalz state, Germany (Landesaerztekammer Rheinland-Pfalz, Koerperschaft des oeffentlichen Rechts) with request number 2021-15895-retrospektiv.

We conducted a retrospective review from 2014 to 2020 using our institution's CI database to retrieve the preoperative and postoperative imaging data of patients who underwent CI surgery. Exclusion criteria were the presence of congenital cochlear anomalies or severe artifacts distorting the image quality.

Dataset

The study included 41 CI patients of different genders and ages with a total of about 131 pre-operative and postoperative imaging studies from different modalities (MR, MDCT and CBCT). All the patients' information was anonymised.

All preoperative MR images were done on a Siemens Avanto 1.5 Tesla device. The sequence used was transverse T2-spc 3D sequence with the following specifications: Repetition time (TR) 9.1 ms Echo time (TE) 3.62 ms, Flip angle 80 degrees; pixel bandwidth 130Hz/Px; thickness 0.8 mm; voxel size 0.7x0.5x;0.8 mm; FOV selection 200 mm; Matrix size 384x384; SNR 1.0; Acquisition time 3:30 min. Each used image has a size of $384 \times 512 \times 64$ voxels with $0.39 \times 0.39 \times 0.7$ mm spacing.

All preoperative MDCT images were done on a Siemens Sensation Cardiac 64 Slice. Each image was taken with these parameters: KV = 120; mAS = 140; matrix = 512×512 ; FOV = 65–75 mm; slice thickness = 0.6 mm; Kernel = U90 μ . Each MDCT image has a size of $512 \times 512 \times 58$ voxels with $0.12 \times 0.12 \times 1.0$ mm spacing.

Preoperative and postoperative CBCT images were done on a Morita 3D Accuitomo 170 (J. Morita). Each ear was imaged separately using the following parameters: 90-kV tube voltage; 5-mA current; a high-resolution mode (Hi-Res J.Morita) with a rotation of 180 degrees; voxel size of 0.125 mm and FOV of 80x80x80 mm. The final postoperative CBCT images have a size of $485 \times 485 \times 121$ voxels with $0.12 \times 0.12 \times 0.5$ mm spacing and the final preoperative CBCT images have a size of $483 \times 483 \times 161$ voxels with $0.12 \times 0.12 \times 0.3$ mm spacing. These CBCT images were probably cropped as they don't have isotropic spacing.

ACIR: Automatic Cochlea Image Registration

The multi-registration Automatic Cochlea Image Registration (ACIR-v1) was proposed in [13] to provide practical cochlea image registration in a few seconds. It combined different standard techniques and tuned parameters in a customized way that worked for cochlea images. It combined different registration elements in a hierarchical approach of two stages. In the first stage, a rigid registration was used while in the second stage, a non-rigid *B-spline* registration was used. Moreover, the metric was based on the input image types i.e S_{MSE} for mono-modal images i.e CBCT and CBCT, and S_{MMI} for multimodal images e.g. MDCT and MR, or MDCT and CBCT. Even though CBCT and MDCT are very similar, in practice, CBCT and CT had better image registration result when S_{MMI} was used. It used the Adaptive Stochastic Gradient Descent (ASGD) as an optimiser [14]. The step-size λ , in Eqs (3) and (5), was an important optimisation factor that had to be set manually. It had a large influence on the GD optimisation results. This factor is data-dependent, so finding a suitable value for different problems was challenging. ASGD is an optimiser that adapts the step-size factor λ automatically, using an image-driven mechanism, to predict its value. It replaces λ with a parameter δ that defines the maximum incremental displacement allowed between the optimisation's iterations. The δ parameter was computed using the voxel size of the input image. The mean voxel spacing (in mm) was found to be a good value.

The cochlear image registration problem in this study is an intra-subject image registration problem. Hence, it belongs to rigid image registration problems. For such images, the same transform applies to all pixels in the moving image. Finding the correct transform for any part of the moving image may solve the problem for other parts in the image. Some areas in these images have clear structures and less noise. Cropping the original images to one of these small areas and registering them is an efficient way to produce a transform that registers the original images. Another method, ACIR-v2, which uses prior knowledge of the images' characteristics was introduced in [15]. ACIR-v2 uses the ASGD optimiser to minimise negative Mattes'

mutual information S_{MMI} similarity metric of two images [16] by modifying 3D rigid transformation’s parameters.

ACIR-v1 and ACIR-v2 were GD methods. In GD, the second order derivative of the similarity metric function, also called the Hessian H , was the identity matrix I , hence, it was omitted from (3). In some cases where the Hessian matrix is ill-conditioned, this assumption leads to an inefficient optimisation result. In this case, one should use a Newton optimiser method where the second-order partial derivatives of the cost function are considered. This approach is very computationally expensive as the computation of the Hessian requires a lot of time and memory capacity.

The proposed idea

Quasi-Newton methods [7] use an approximation of the Hessian of the cost function which reduces the computation time. The update function in quasi-Newton optimiser is:

$$\mu_{k+1} = \mu_k - \lambda \left(\frac{\left(\frac{\partial S_k(I_F(P), I_M(T(P, \mu_k)))}{\partial \mu} \right)}{\left(\frac{\partial^2 S_k(I_F(P), I_M(T(P, \mu_k)))}{\partial \mu^2} \right)} \right), \tag{6}$$

where k is the current optimisation iteration, $\frac{\partial S_k}{\partial \mu}$ is the first order derivative of the similarity metric function at iteration k , λ is the step-size and $\frac{\partial^2 S_k}{\partial \mu^2}$ is a symmetric positive definite approximation of the Hessian matrix. A popular quasi-Newton update rule is Broyden-Fletcher-Goldfarb-Shanno (BFGS) [7] which uses the first-order derivatives to update the inverse Hessian directly. This produces a linear rate convergence. The BFGS update rule is described in Eq (7):

$$H_{k+1} = V_k^T H_k^{-1} V_k + \frac{\mu'_k \mu'^T_k}{g'^T_k \mu'_k}, \tag{7}$$

where $V_k = I - \frac{g'_k \mu'^T_k}{g'^T_k \mu'_k}$, I is the identity matrix, $g'_k = \frac{\partial S_{k+1}}{\partial \mu} - \frac{\partial S_k}{\partial \mu}$, $\mu'_k = \mu_{k+1} - \mu_k$, H_k^{-1} is the inverse matrix of $\frac{\partial^2 S_k}{\partial \mu^2}$, V^T is the transpose of V , and μ^T is the transpose of μ .

This still requires a long computation time and large memory. The memory issue is solved by the Limited-memory BFGS (LBFGS) [17] method which saves only a few numbers of the previous Hessian approximations. The computation time issue is solved using a stochastic approach.

This paper introduces the Automatic Cochlea Image Registration (ACIR-v3) method which is based on the stochastic quasi-Newton with Limited-memory BFGS (s-LBFGS) updating rule.

The proposed method replaces the ASGD optimiser in ACIR-v2 with the s-LBFGS optimiser [18]. The s-LBFGS combines the stochastic approach with the LBFGS approach. This allows for faster convergence and more robust results.

Evaluation method and statistical analysis

The ACIR-v3 method was compared to its previous versions (ACIR-v1, ACIR-v2), and three other optimisers i.e. ASGD, Fast Adaptive Stochastic Gradient Descent (FASGD) [19], and Fast Preconditioner Stochastic Gradient Descent (FPSGD) [20]. Since the result time recorded might differ according to the hardware used, all the experiments were done using the same hardware. The hardware used was a computer equipped with AMD Ryzen 3900 CPU, a 32 GB

memory and a Nvidia RTX2080Ti graphics card. We used the same parameters for all methods and the original implementations of ACIR-v1, ACIR-v2, ASGD, s-LBFGS, FPSGD by their authors which are provided as an open-source in 3D Slicer software and Elastix 5.0.0 toolbox [21, 22].

In the ASGD method, we used the original images and no cropping was involved. The ASGD parameters were: rigid transformation, no multi-stages, and no multi-resolution. The optimiser was changed in order to make fair comparisons between FASGD, FPSGD, ACIR-v2, and ACIR-v3. The comparison between the methods included three criteria: the image registration accuracy, the required time to align a pair of images, and the robustness of the method. The time was recorded using a fixed number of iterations ($n = 100$) and the pre-processing and post-processing steps, e.g. image cropping, were included. The robustness of the method was evaluated using the percentage of the missing results according to Eq (8):

$$\frac{N_{success}}{N_{success} + N_{fail}}, \quad (8)$$

where $N_{success}$ is the number of cases where the method produced a valid result, and N_{fail} is the number of cases where the method failed to produce a valid result e.g. the optimisation stopped.

Algorithm 1: ACIR-v3

- 1 **Input:** two cochlea images $I_F(P)$, and $I_M(P)$;
- 2 **Output:** a registered and fused image $I_{result}(P)$;
- 3 Locate the cochlea locations in input images;
- 4 $\tilde{I}_F, \tilde{I}_M = \text{Crop}(I_F, I_M)$;
- 5 Set transform $T = 3\text{D rigid transform}$;
- 6 Set $k = 0$;
- 7 **while** $-S_{MMI}(\tilde{I}_F(P), \tilde{I}_M(T(P, \mu)))$ is large and $k < 100$ **do**;
- 8 $\mu = \text{update the old } \mu \text{ using s-LBFGS.}$;
- 9 Set $k = k + 1$;
- 10 Transform the moving image, $I_{result}(P) = I_M(T(P, \mu))$;
- 11 Fuse the result, $I_{result} = I_{result} + I_F$

Two neuroradiology experts added two 3D landmark points (fiducial points) (in consensus) to all the imaging studies using the 3D Slicer software version 4.10 [23, 24]. The two landmark points represent the round window and the cochlear apex, see Fig 3. These landmarks were saved for each imaging study and were used later for validation of the image registration results. After aligning the images, the landmarks of the moving image were transformed using the same transformation. Thereafter, RMSE in mm was measured between these transformed landmarks and the related fixed image landmarks.

For justification, the results were repeated 3 times and the average values were used. The results have been divided into 4 groups based on image modalities to obtain more details on how each method performs on a specific modality. This allowed for more detailed evaluation as some methods may be biased to specific types of images. These groups were (CBCT, CBCT), (CBCT, MR), (CBCT, MDCT), and (MR, MDCT). In addition to these group results, the total results were presented to give a global evaluation of each method. The average RMSE and average time calculated for each registration process in each of the 6 methods were then used to compare (ACIR-v3) to each of the other 5 methods using independent T-tests.

Other optimisers (used for comparisons)

FASGD is another adaptive stochastic optimiser that estimates the step-size automatically using the observed voxel displacement. During the computation of the step-size, two ASGD

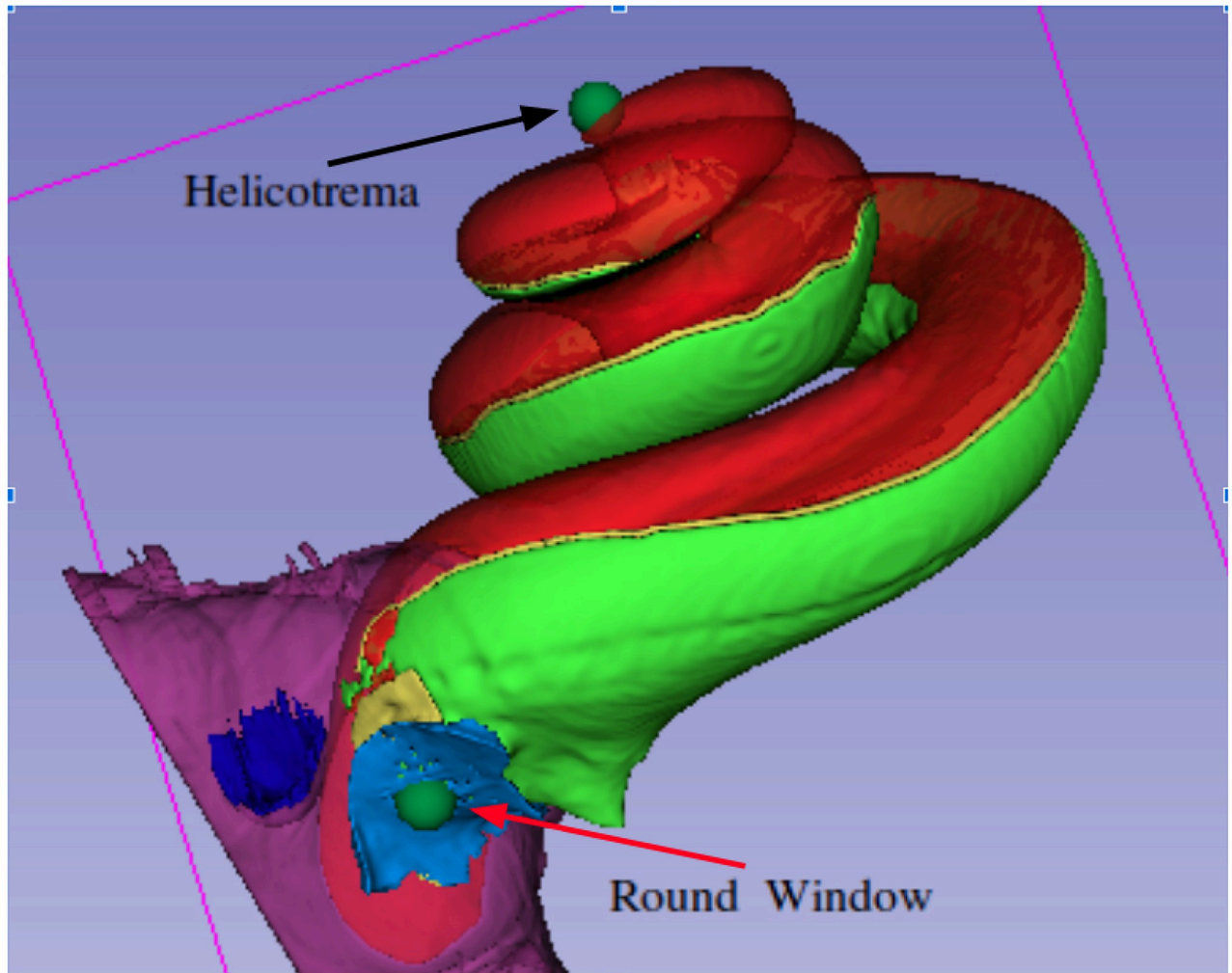


Fig 3. Cochlea 3D models and two landmarks. This model is generated in our lab by manual segmentation of a high quality μ CT image using 3D Slicer software. The red colour represent scala media and scala vestibuli. The green colour represent the scala tympani.

<https://doi.org/10.1371/journal.pone.0264449.g003>

parameters are fixed in FASGD, while they have to be computed in each iteration in ASGD. This makes FASGD faster but less adaptive than ASGD.

Preconditioned Stochastic Gradient Descent (PSGD) [25] is proposed to improve the rate of convergence of GD methods. It adds a preconditioning strategy to Robbins-Monro stochastic gradient descent (SGD_{RM}) [11] and ASGD methods. The updating rule for PSGD is:

$$\mu_{k+1} = \mu_k - \lambda M \frac{\partial \mathcal{S}_k(\hat{I}_F(P), \hat{I}_M(T(P, \mu_k)))}{\partial \mu} \quad (9)$$

The equation is similar to Eq (5), the only difference is adding the preconditioner matrix M which is a symmetric positive definite matrix with size equal to the number of the transformation's parameters, i.e. $|\mu|$. When M is the identity matrix, we get the standard SGD method as in Eq (5). In [25] the step-size is defined to be a non-increasing and non-zero sequence to

guarantee convergence:

$$\lambda_k = \begin{cases} 1 & \text{if } k = 0 \\ \frac{\eta}{((t_k + 1)/A) + 1} & \text{if } k > 0 \end{cases}, \tag{10}$$

where $t_0 = t_1 = 0$, $t_k = \max(0, t_{k-1} + \text{sigmoid}(\tilde{g}_{k-1}M\tilde{g}_{k-2}))$, \tilde{g} is an approximation of $\frac{\partial S_k}{\partial \mu}$, M is the preconditioner matrix, $A = 20$ is a decay speed factor, and η is a noise factor and is defined as:

$$\eta = \frac{E\|g^T M g\|}{E\|\tilde{g}^T M \tilde{g}\|} = \frac{E\|g^T M g\|}{E\|g^T M g\| + E\|\epsilon^T M \epsilon\|}, \tag{11}$$

where g is the exact gradient and ϵ is a random noise added to the exact gradient. Unfortunately, PSGD is limited to monomodal image registration where the fixed and the moving images are the same type e.g. MDCT. The FPSGD solved this issue and can be used for multimodal image registration. The authors mention that FPSGD is 2 to 5 times faster than SGD methods while retaining the same accuracy [20].

FPSGD estimates the diagonal entries of a preconditioning matrix M on the distribution of voxel displacements. This rescales the registration similarity metric and allows for more efficient optimisation. Let M be a preconditioning matrix with the same length of the parameter vector μ , and m_i is an element in the diagonal of M that can be computed using Eq (12):

$$m_i = \frac{\delta}{E(\|J_{\mu_i}\| \cdot \|\tilde{g}_i\|) + 2\sqrt{\text{Var}(\|J_{\mu_i}\| \cdot \|\tilde{g}_i\|)} + \epsilon}, \tag{12}$$

where δ is a pre-defined value representing the maximum voxel displacement, $E\|X\|$ is the expectation of the l_2 norm, $\text{Var}\|X\|$ is the variance of the l_2 norm, J_{μ_i} is the i th column in the Jacobian matrix $J_{\mu} = \frac{\partial T(P;\mu)}{\partial \mu}$ which is mentioned earlier in Eq (4), and ϵ is a small number to avoid dividing by zero.

Results

Out of the 131 imaging studies included in the experiments, 124 registration processes (CBCT, CBCT), (CBCT, MR), (CBCT, MDCT), and (MR, MDCT) were performed using the 6 included methods (ASGD2009, FASGD2015, FPSGD2019, ACIR-v1, ACIR-v2, and ACIR-v3). The mean RMSE, the mean time used for each method, and the results of their comparison to ACIR-v3 using the independent samples T-test are shown in Tables 1 and 2. N represents the number of registration processes, and SD represents the standard deviation. The registration processes failed in FASGD2019 in 2/124 and in FPSGD2019 in 13/124.

In Figs 4 and 5, visual samples of CBCT to CBCT, and CBCT to MDCT image results of all 6 methods are shown. The fixed image is in magenta and the moving image is in green. The distance between the ground truth and the registered ground truth landmarks is also shown. When more green and magenta can be seen, the registration process is inadequate as seen in the bottom right image of Fig 4 where FPSGD2019 method is used. When the green and magenta are mixed, the registration process is successful as seen in the bottom left of Fig 4 where ACIR-v3 is used. Similarly, Figs 6 and 7 show visual samples of CBCT to MR, and MR to MDCT image registration and fusion results. We selected different modalities and different views to give a visual sample of the results.

Table 1. Independent samples T-test for the RMSE of each method in comparison to ACIR-v3.

Method	N	RMSE (mm)	SD	p-value
ASGD2019	124	14.98	8.63	<0.0001
FASGD2015	122	2.77	11.85	0.03
FPSGD2019	111	1.45	8.39	0.194
ACIR-v1	124	10.95	9.497	<0.0001
ACIR-v2	124	0.36	0.17	0.125
ACIR-v3	124	0.41	0.30	-

N: number of registration experiments. RMSE: root mean squared error. SD: standard deviation. p-value: statistical samples t-test p-value comparison to ACIR-v3.

<https://doi.org/10.1371/journal.pone.0264449.t001>

The three charts in Fig 8 show a comparison of the 6 methods in terms of accuracy, time, and robustness. ASGD, ACIR-v2, and ACIR-v3 methods were 100% robust.

Discussion

There are few scientific papers related to cochlea image registration. Manual image registration and fusion are usually done by doctors which requires much effort and time as described in [26]. The authors proposed a manual procedure for image registration and fusion of CT and MR of the temporal bone using bony surgical landmarks. This procedure took 13 minutes per scan.

In [27], an automatic cochlea image registration for CT was proposed for Percutaneous Cochlear Implantation (PCI) surgery. The authors reported a maximum of 0.19 mm error when using their method. The method was complex as it involved using segmentation and two-stage registration. The computation time required to complete the registration was approximately 21 minutes which was larger than the time required for manual registration (approximately 10 minutes based on our experiments). The authors investigated only complete head CT scans so the functionality of their method with multi-modal images or small images focused on the cochlea aspect is not clear.

In the other research, high-resolution μ CT data was used in [28, 29] where two methods were proposed for cochlea image registration. The first method was based on heat distribution similarity in a cubic B-splines registration model. The second method was based on skeleton similarity as an anatomical prior. Their methods required cropped and segmented images which was a time consuming process that took approximately 48 minutes per scan.

Table 2. Independent samples T-test for the time of each method in comparison to ACIR-v3.

Method	N	Time (seconds)	SD	p-value
ASGD2019	124	6.45	0.83	<0.0001
FASGD2015	122	4.76	1.23	0.37
FPSGD2019	111	4.70	1.19	0.601
ACIR-v1	124	15.97	3.37	<0.0001
ACIR-v2	124	4.65	1.2	0.86
ACIR-v3	124	4.62	1.2	-

N: number of registration experiments. SD: standard deviation. p-value: statistical samples t-test p-value comparison to ACIR-v3.

<https://doi.org/10.1371/journal.pone.0264449.t002>

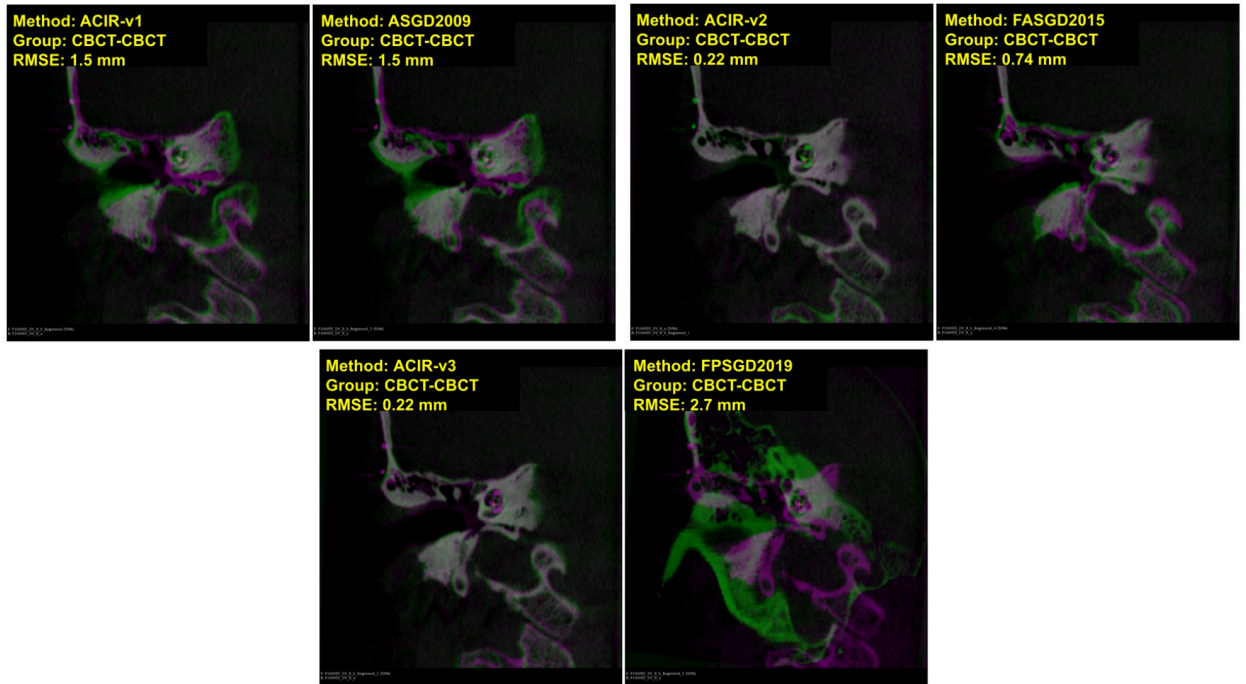


Fig 4. Sample results of registration CBCT to CBCT (coronal views). The figure shows image types, methods, and RMSE information. Magenta represents the fixed image and green represents the moving image.

<https://doi.org/10.1371/journal.pone.0264449.g004>

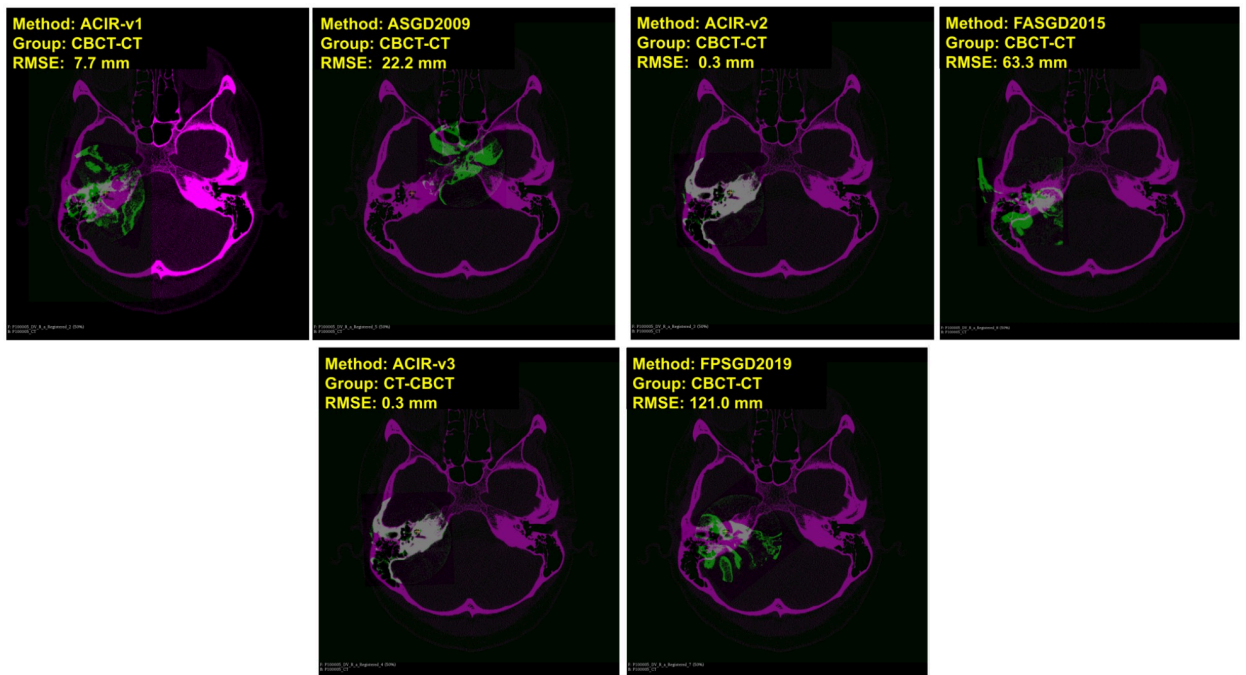


Fig 5. Sample results of registration CBCT to MDCT (axial views). The figure shows image types, methods, and RMSE information. Magenta represents the fixed image and green represents the moving image.

<https://doi.org/10.1371/journal.pone.0264449.g005>

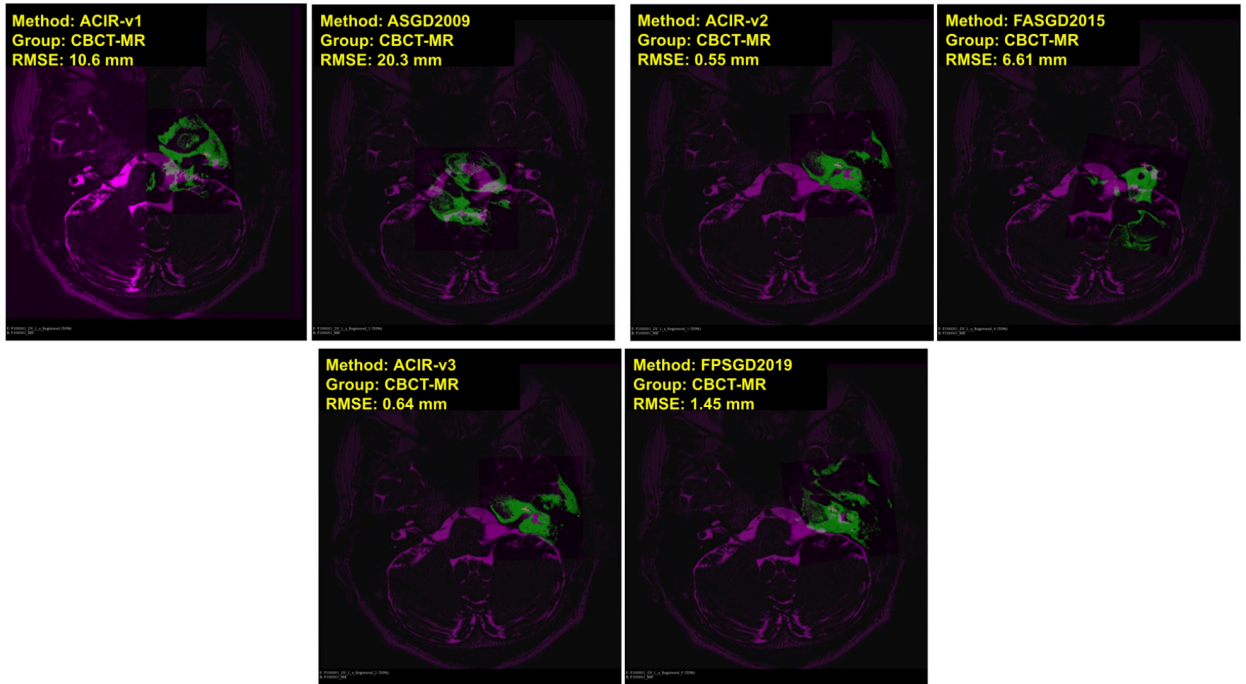


Fig 6. Sample results of registration CBCT to MR. The figure shows the axial views with image types, methods, and RMSE information. Magenta represents the fixed image and green represents the moving image.

<https://doi.org/10.1371/journal.pone.0264449.g006>

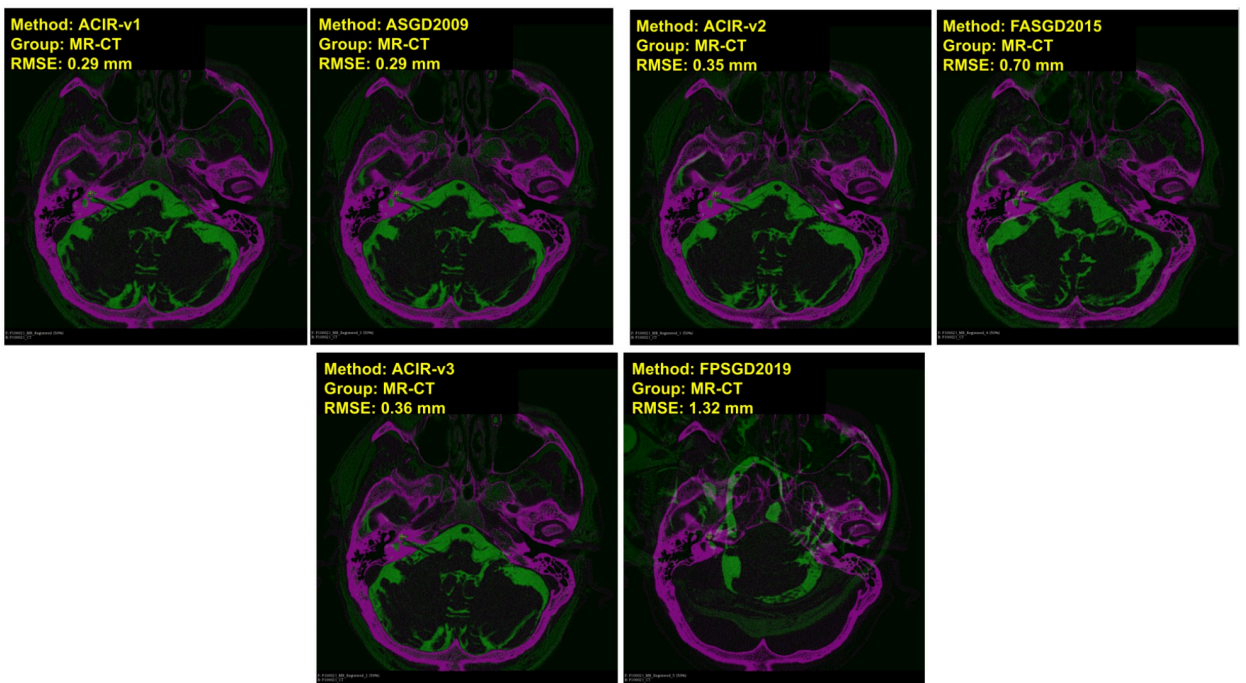


Fig 7. Sample results of registration MR to MDCT. The figure shows the axial views with image types, methods, and RMSE information. Magenta represents the fixed image and green represents the moving image.

<https://doi.org/10.1371/journal.pone.0264449.g007>

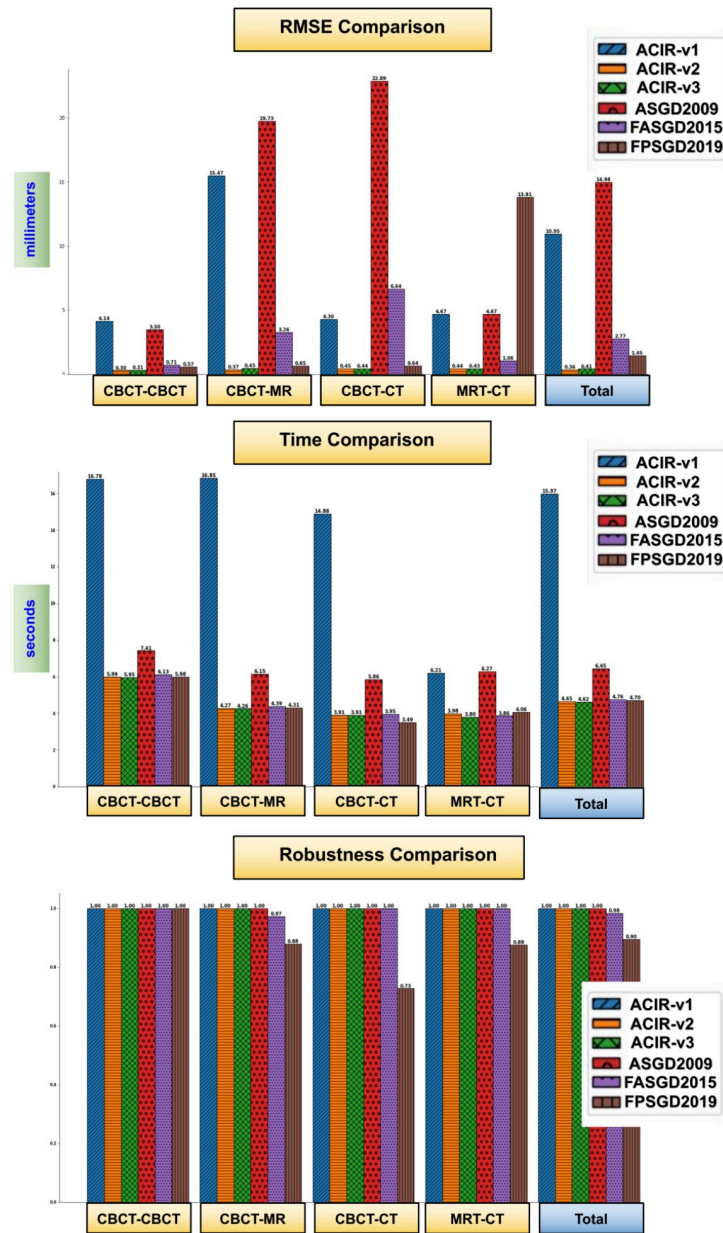


Fig 8. Results chart, the average RMSE, time, and robustness of each tested method grouped based on image-modalities. The robustness is computed based on Eq (8).

<https://doi.org/10.1371/journal.pone.0264449.g008>

Lately, a proposed cochlea registration method was used in [30]. The fusion of sequential CBCT was compared to the gold standard fiducial in order to analyse clinical CI migration. BRAINSFit tool in 3D Slicer was used which reduced the duration of the registration to less than 2 minutes. A mean error of 0.16 mm was reported. However, this method did not support multimodal images, and the images used did not include cochlea implants which made the registration problem less challenging.

Automatic Cochlea Image Registration and fusion (ACIR-v1) was the first attempt to achieve practical automatic cochlea image registration. The method’s robustness was very

high, only one case lacked results, but some images required a manual correction step which took a few seconds. The average RMSE error was less than 5 mm for all image types except (CBCT, MR) where it was 15.47mm. It seems that for these two image types, the noise level was too high which made the mutual information unclear and so it could not be maximised. The time required to process a pair of images was high because of the use of the *B-spline* transform which has thousands of parameters. Compared to ASGD, ACIR-v1 produced a 3.8% improvement in accuracy.

The cropping in ACIR-v1 divided the input images to include the targeted cochlea i.e. 50% of the image represents either the left or the right ear. In ACIR-v2, the cropping was reduced to the cochlea region which was around 10x10x10 mm. Additionally, the multi-resolution and the multi-registration approaches were removed. This produced better accuracy and faster computation. This showed that the cropping approach in ACIR-v2 was instrumental in achieving practical results and using ASGD alone was not enough.

In this study, the ACIR-v3 method has been proposed based on the quasi-Newton s-LBFGS method. The history of the method's development was outlined i.e. ACIR-v1, and ACIR-v2. The method presents a practical and fast solution for the multimodal cochlea 3D image registration and fusion problem. The method results have been validated using landmarks located by two experts and tested using 131 multimodal cochlea 3D images of 41 patients from CBCT, MDCT, and MR modalities. The method aligned all the images successfully in a few seconds. The results showed that using ASGD or s-LBFGS has higher accuracy than FASGD or FPSGD optimisers, especially concerning multimodal images. This is due to the FASGD method being less adaptive than ASGD, whilst the FPSGD method has higher sensitivity to noise. In low-resolution images with complicated structures, such as clinical images of the cochlea, more noise is included which leads to poor alignment when FPSGD is used. ACIR-v3 has 4.8% accuracy and 0.05% speed enhancement over the FPSGD method on average.

The comparison of the ACIR-v3 to its previous versions (ACIR-v1 and ACIR-v2), showed a statistically significant difference between ACIR-v3 and ACIR-v1 with much improvement in the time used and lower RMSE. However, there was no statistically significant difference between ACIR-v3 and ACIR-v2.

The comparisons of ACIR-v3 with the known methods showed no statistically significant differences in terms of the mean time used except for (ASGD2009), where ACIR-v3 was relatively faster ($p < 0.0001$). The comparison of the RMSE in ACIR-v3 to ASGD2009 ($p < 0.0001$) and FASGD2015 ($p = 0.03$) was statistically significant. The comparison of the RMSE of the FPSGD2019 to ACIR-v3 showed no statistical significance, yet the mean RMSE was relatively higher in FPSGD2019 (1.45 ± 8.4 mm compared to only 0.41 ± 0.3 mm in ACIR-v3). However, the registration process failed in 11 out of the 124 processes in FPSGD2019 which demonstrated instability.

Limitation

The image registration field is very rich. There are many methods published every year and one cannot cover all the related work in one paper. Hence we selected related work close to our problem which involves rigid transformation and cochlea images. However there is some interesting work which can be used for problems involving non-rigid image registration and different datasets as in [31–35]. Deep learning methods are increasing in popularity in medical image registration e.g. as in [36, 37]. However, they require expensive specialized hardware and thousands of medical images for training which are often unavailable. Moreover, the accuracy of these methods is poor in some tasks [38].

Conclusions

The proposed method's source-code is provided as a public open-source and can be downloaded from a public server. The raw cochlea medical image dataset cannot be made publicly available at present for legal reasons. However, all the original results tables are publicly available. This enables other researchers to reproduce these experiments and for people to benefit from the results.

Future work will include studying more s-LBFGS parameters and testing the method against different problems and on different datasets.

Supporting information

S1 Data.
(CSV)

Author Contributions

Conceptualization: Ibraheem Al-Dhamari, Rania Helal, Olesia Morozova, Tougan Abdelaziz, Roland Jacob, Dietrich Paulus, Stephan Waldeck.

Data curation: Ibraheem Al-Dhamari, Stephan Waldeck.

Formal analysis: Ibraheem Al-Dhamari, Rania Helal, Tougan Abdelaziz, Roland Jacob, Dietrich Paulus, Stephan Waldeck.

Investigation: Ibraheem Al-Dhamari, Roland Jacob, Dietrich Paulus, Stephan Waldeck.

Methodology: Ibraheem Al-Dhamari, Rania Helal, Tougan Abdelaziz, Roland Jacob, Dietrich Paulus, Stephan Waldeck.

Project administration: Tougan Abdelaziz, Roland Jacob, Dietrich Paulus, Stephan Waldeck.

Resources: Ibraheem Al-Dhamari, Roland Jacob, Dietrich Paulus, Stephan Waldeck.

Software: Ibraheem Al-Dhamari, Olesia Morozova.

Supervision: Tougan Abdelaziz, Roland Jacob, Dietrich Paulus, Stephan Waldeck.

Validation: Ibraheem Al-Dhamari, Rania Helal, Olesia Morozova, Tougan Abdelaziz, Roland Jacob, Dietrich Paulus, Stephan Waldeck.

Visualization: Ibraheem Al-Dhamari.

Writing – original draft: Ibraheem Al-Dhamari, Rania Helal, Olesia Morozova, Tougan Abdelaziz, Roland Jacob, Dietrich Paulus, Stephan Waldeck.

Writing – review & editing: Ibraheem Al-Dhamari, Rania Helal, Olesia Morozova, Tougan Abdelaziz, Roland Jacob, Dietrich Paulus, Stephan Waldeck.

References

1. Vogl T, Tawfik A, Emam A, Naguib, Nour-Eldin A, Naguib N, et al. Pre-, Intra- and Post-Operative Imaging of Cochlear Implants. *RoFo: Fortschritte auf dem Gebiete der Rontgenstrahlen und der Nuklearmedizin*. 2015; 187(11):980–989. <https://doi.org/10.1055/s-0035-1553413> PMID: 26327670
2. Wang Z, Ziou D, Armenakis C, Li D, Qingquan L. A comparative analysis of image fusion methods. *IEEE Transactions on Geoscience and Remote Sensing*. 2005; 43(6):1391–1402. <https://doi.org/10.1109/TGRS.2005.846874>
3. Hajnal J, Hill D. *Medical Image Registration*. 1st ed. CRC press; 2001.
4. Gonzalez R, Woods R. *Digital Image Processing*. 4th ed. Pearson; 2018.

5. Yoo T. *Insight into Images: Principles and Practice for Segmentation, Registration, and Image Analysis*. AK Peters Ltd; 2012.
6. Hans Johnson, Matthew McCormick, Luis Ibanez, and the Insight Software Consortium. *The ITK Software Guide, Book 2: Design and Functionality, Fourth Edition*. 2019 May 28. [Cited 2022 February 16]. Available from <https://itk.org/ITKSoftwareGuide/html/Book2/ITKSoftwareGuide-Book2ch3.html>;
7. Nocedal J, Wright S. *Numerical Optimization*. 2nd ed. New York, NY, USA: Springer; 2006.
8. Snyman J. *Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms*. Springer Publishing; 2005.
9. Diez D, Cetinkaya-Rundel M, Barr C. *OpenIntro Statistics*. 4th ed. Electronic copy; 2019. Available from: <https://leanpub.com/openintro-statistics>.
10. Viola P, Wells W. Alignment by Maximization of Mutual Information. *International Journal of Computer Vision*. 1997; 24(2):137–154. <https://doi.org/10.1023/B:VISI.0000013087.49260.fb>
11. Robbins H, Monro S. A Stochastic Approximation Method. *The Annals of Mathematical Statistics*., 1951; 22(3):400–407. <https://doi.org/10.1214/aoms/1177729586>
12. Klein S, Staring M, Pluim J. Evaluation of Optimization Methods for Nonrigid Medical Image Registration Using Mutual Information and B-Splines. *IEEE Transactions on Image Processing*. 2007; 16(12):2879–2890. <https://doi.org/10.1109/TIP.2007.909412> PMID: 18092588
13. Al-Dhamari I, Bauer S, Paulus D, Lissek F, Jacob R. Automatic Multimodal Registration and Fusion of 3D Human Cochlea Images. In: *The 14TH INTERNATIONAL CONFERENCE ON COCHLEAR IMPLANTS CI2016*. Canada: American Cochlear Implant Alliance; 2016.
14. Klein S, Pluim J, Staring M, Viergever M. Adaptive stochastic gradient descent optimisation for image registration. *International Journal of Computer Vision*. 2009; 81(3):227–239. <https://doi.org/10.1007/s11263-008-0168-y>
15. Al-Dhamari I, Bauer S, Paulus D, Lessek F, Jacob R, Gessler A. ACIR: Automatic Cochlea Image Registration. In: *Proc. SPIE 10133, Medical Imaging 2017: Image Processing*. vol. 10133; 2017. p. 1–5.
16. Mattes D, Haynor D, Vesselle H, Lewellyn T, Eubank W. Non-rigid multimodality image registration. *Medical Imaging 2001: Image Processing*, SPIE Publications. 2001; p. 1609–1620.
17. Saputroa D, Widyarningsih P. Limited memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) method for the parameter estimation on geographically weighted ordinal logistic regression model (GWOLR). *AIP Conference Proceedings*. 2017; 1868(1):0400091–0400099.
18. Qiao Y, Sun Z, Lelieveldt B, Staring M. A Stochastic Quasi-Newton Method for Non-Rigid Image Registration. In: Navab N, Hornegger J, Wells W, Frangi A, editors. *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*. Cham: Springer International Publishing; 2015. p. 297–304.
19. Qiao Y, van Lew B, Lelieveldt B, Staring M. Fast Automatic Step Size Estimation for Gradient Descent Optimization of Image Registration. *IEEE Transactions on Medical Imaging*. 2016; 35(2):391–403. <https://doi.org/10.1109/TMI.2015.2476354> PMID: 26353367
20. Qiao Y, Lelieveldt B, Staring M. An Efficient Preconditioner for Stochastic Gradient Descent Optimization of Image Registration. *IEEE Transactions on Medical Imaging*. 2019; 38(10):2314–2325. <https://doi.org/10.1109/TMI.2019.2897943> PMID: 30762536
21. Klein Staring, Murphy, Viergever, Pluim. elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging* 2010. 2010; 29(1):196–205. <https://doi.org/10.1109/TMI.2009.2035616> PMID: 19923044
22. The elastix tool web site. [Cited 2022 February 16]. Available from <http://elastix.isi.uu.nl>;
23. Kikinis R, Pieper S, Vosburgh K. 3D Slicer: a platform for subject-specific image analysis, visualization, and clinical support. *Intraoperative Imaging Image-Guided Therapy*, Ferenc A Jolesz. 2014; 3(19):277–289. https://doi.org/10.1007/978-1-4614-7657-3_19
24. The slicer website [Cited 2022 February 16]. Available from <https://slicer.org>;
25. Klein S, Staring M, Andersson P, Pluim J. Preconditioned Stochastic Gradient Descent Optimisation for Monomodal Image Registration. In: Fichtinger G, Martel A, Peters T, editors. *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2011*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011. p. 549–556.
26. Bartling S, Peldschus K, Rodt T, Kral F, Matthies H, Kikinis R, et al. Registration and Fusion of CT and MRI of the Temporal Bone. *Journal of Computer Assisted Tomography*. 2005; 29(3):305–310. <https://doi.org/10.1097/01.rct.0000160425.63884.5b> PMID: 15891495
27. Reda F, Noble J, Labadie R, Dawant B. Fully Automatic Surface-Based Pre- to Intra-operative CT Registration for Cochlear Implant. In: Dawant B, Christensen G, Fitzpatrick M, Rueckert D, editors.

- Biomedical Image Registration. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012. p. 89–98. https://doi.org/10.1007/978-3-642-31340-0_10
28. Kjer H, Vera S, Fagertun J, Gil D, González Ballester, Miguel Angel, et al. Image Registration of Cochlear μ CT Data Using Heat Distribution Similarity. In: Paulsen R, Kim P, editors. Image Analysis. United States: Springer Science+Business Media; 2015. p. 234–245.
 29. Kjer H, Fagertun J, Vera S, Gil D, Ballester M, Paulsen R. Free-form image registration of human cochlear μ CT data using skeleton similarity as anatomical prior. Pattern Recognition Letters. 2016; 76:76–82. <https://doi.org/10.1016/j.patrec.2015.07.017>
 30. Dees G, van Hoof Marcand Stokroos R. A Proposed Method for Accurate 3D Analysis of Cochlear Implant Migration Using Fusion of Cone Beam CT. Frontiers in surgery. 2016; 3:2–2. <https://doi.org/10.3389/fsurg.2016.00002> PMID: 26835459
 31. Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. Medical Image Analysis. 2001; 5(2):143–156. [https://doi.org/10.1016/S1361-8415\(01\)00036-6](https://doi.org/10.1016/S1361-8415(01)00036-6) PMID: 11516708
 32. Zaveri T, Zaveri M. A Novel Multimodality Image Fusion Method Using Region Consistency Rule. In: Chaudhury S, Mitra S, Murthy CA, Sastry PS, Pal SK, editors. Pattern Recognition and Machine Intelligence. Berlin, Heidelberg: Springer Berlin Heidelberg; 2009. p. 321–326.
 33. Liao YL, Sun YN, Guo WY, Chou YH, Hsieh JC, Wu YT. A hybrid strategy to integrate surface-based and mutual-information-based methods for co-registering brain SPECT and MR images. Medical & Biological Engineering & Computing. 2011; 49(6):671–685. <https://doi.org/10.1007/s11517-010-0724-9> PMID: 21191661
 34. Sinibaldi R, Conti A, Sinjari B, Spadone S, Pecci R, Palombo M, et al. Multimodal-3D imaging based on μ MRI and μ CT techniques bridges the gap with histology in visualization of the bone regeneration process. Journal of Tissue Engineering and Regenerative Medicine; 12(3):750–761. <https://doi.org/10.1002/term.2494> PMID: 28593731
 35. Guidotti R, Sinibaldi R, Luca CD, Conti A, Ilmoniem RJ, Zevenhoven KCJ, et al. Optimized 3D co-registration of ultra-low-field and high-field magnetic resonance images. PLOS ONE. 2018; 13(3):1–19. <https://doi.org/10.1371/journal.pone.0193890> PMID: 29509780
 36. Hoopes A, Hoffmann M, Fischl B, Gutttag J, Dalca AV. HyperMorph: Amortized Hyperparameter Learning for Image Registration; 2021.
 37. Mok TCW, Chung ACS. Conditional Deformable Image Registration with Convolutional Neural Network; 2021.
 38. MICCAI Challenge Learn2reg results. [Cited 2022 February 16]. Available from <https://learn2reg.grand-challenge.org/evaluation/task-3-validation/leaderboard/>;