Localizing landmark sets in head CTs using random forests and a heuristic search algorithm for registration initialization

Dongqing Zhang
Yuan Liu
Jack H. Noble
Benoit M. Dawant

Localizing landmark sets in head CTs using random forests and a heuristic search algorithm for registration initialization

Dongqing Zhang,* Yuan Liu, Jack H. Noble, and Benoit M. Dawant

Vanderbilt University, Department of Electrical Engineering and Computer Science, Nashville, Tennessee, United States

Abstract. Cochlear implants (CIs) use electrode arrays that are surgically inserted into the cochlea to stimulate frequency-mapped nerve endings to treat patients with hearing loss. CIs are programmed postoperatively by audiologists using behavioral tests without information on electrode–cochlea spatial relationship. We have recently developed techniques to segment the intracochlear anatomy and to localize individual contacts in clinically acquired computed tomography (CT) images. Using this information, we have proposed a programming strategy that we call image-guided CI programming (IGCIP), and we have shown that it significantly improves outcomes for both adult and pediatric recipients. One obstacle to large-scale deployment of this technique is the need for manual intervention in some processing steps. One of these is the rough registration of images prior to the use of automated intensity-based algorithms. Although seemingly simple, the heterogeneity of our image set makes this task challenging. We propose a solution that relies on the automated random forest-based localization of multiple landmarks used to estimate an initial transformation with a point-based registration method. Results show that it produces results that are equivalent to a manual initialization. This work is an important step toward the full automation of IGCIP. © 2017 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JMI.4.4.044007](https://doi.org/10.1117/1.JMI.4.4.044007)]

Keywords: image registration; landmark localization; regression forest.

Paper 17160PR received Jun. 6, 2017; accepted for publication Nov. 13, 2017; published online Dec. 8, 2017.

1 Introduction

Cochlear implants (CIs) use electrode arrays that are surgically inserted into the cochlea to treat patients with hearing loss. Individual contacts in the array stimulate frequency-mapped nerve endings, thus replacing the natural electromechanical transduction mechanism. Postoperatively, the CI is programmed. Key programming parameters are the assignment of a frequency range to each contact, i.e., what contact to activate when a range of frequency is present in the input signal and the level of activation. In current clinical practice, this is done by an audiologist who, blind to the position of the electrodes, relies on patients' subjective response to stimuli, e.g., whether or not they can hear a signal or rank pitches. This is a trial-and-error process that has remained essentially unchanged since the mid-80s and can be frustratingly long (dozens of programming sessions is not unusual). In the recent past, we have introduced a technique that we call image-guided CI programming (IGCIP).¹ This technique relies on a series of image processing steps that permit the segmentation of the inner ear structures in preoperative CT images and the localization of the implant in postoperative CT images. Using this information, we have designed techniques that assist audiologists in programming the implants. We have shown in retrospective studies performed on long-term recipients in both adults and children that image-guided programming leads to settings that are preferred over clinical settings that were considered to be optimum.^{2,3} IGCIP thus has the

potential to profoundly change the way CIs are programmed. However, one barrier to the large-scale clinical deployment and evaluation of our methods is the lack of full automation. Indeed, several steps in our sequence of algorithms still require manual intervention. Specifically, the segmentation of the inner ear anatomy in preoperative CTs relies on the registration between an atlas and preoperative CT images. The initial segmentation is obtained by projecting manual segmentations in the atlas to preoperative CTs using the computed transformation and is refined using an active shape model-based method.⁴ Pre- and postoperative scans also need to be registered for segmenting inner ear anatomy in postoperative CTs. The electrode arrays are localized using methods described in Refs. 5–7 in postoperative CTs. In addition, if a patient is not scanned preoperatively, the atlas and the postoperative CT need to be registered, and a shape library-based algorithm is used for the segmentation of the anatomy.⁸ The processing pipelines that are used to segment the inner ear anatomy and localize electrodes in postoperative CTs are shown in Fig. 1. For all atlas-to-subject registration tasks, we use a sequence of algorithms ranging from rigid-body to nonrigid registrations. In all cases, we rely on intensity-based techniques that use mutual information^{9,10} as similarity measure. Intensity-based algorithms need to be initialized to converge, and properties of our dataset make this process challenging. First, the image sets often have very different fields of view. Instead of covering the whole head, they can be limited to a slab covering the left and right temporal bones, or one half of the head, or even just one temporal bone. This is because

*Address all correspondence to: Dongqing Zhang, E-mail: dongqing.zhang@vanderbilt.edu

Fig. 1 Computation pipelines used to segment the inner ear anatomy and to localize electrode arrays in postoperative CTs when (a) the preoperative CT scan is available and (b) when the preoperative CT scan is not available. Circles represent the atlas, preoperative (preop), or postoperative (postop) CTs. Rectangles represent operations and ellipsoids represent structures of interest. R_{A-pre} , R_{A-post} , and $R_{pre-post}$ represent the transformations from atlas to preoperative CT, atlas to postoperative CT, and preoperative CT to postoperative CT, respectively. ASM represents active shape model-based methods for segmenting inner ear anatomy given an initial segmentation projected from the atlas. The rectangles drawn with dashed lines highlight the pipeline's final outputs.

Fig. 2 (a) Axial, (b) coronal, and (c) sagittal views of four representative image volumes in our dataset. This shows the range of coverage and orientation that can be encountered. Images of patient #3 only include the left and right temporal bones; images of patient #4 only include one temporal bone.

for CI surgery, the main structure of interest is the ear and coverage is limited to avoid exposing patients to unnecessary radiation. Second, the orientation of the scans can vary substantially, exceeding the capture range of traditional registration algorithms. Figure 2 shows a few illustrative examples.

As can be seen, patient #4's scan is unusually tilted because a large gantry angle has been used to avoid radiation to sensitive organs like the eyes. Third, in our study, CT images can be acquired with a standard clinical scanner or with a Xoran xCAT[®] scanner, which is a low-dose flat-panel scanner. Images

Fig. 3 A comparison of images acquired with a (a) Xoran xCAT® scanner and (b) conventional scanner. Both are axial views of the ear region.

produced with this scanner typically have lower signal-to-noise ratio than images produced by conventional CT scanners. Figure 3 shows side-by-side one image acquired with a Xoran xCAT® scanner and one acquired with a standard scanner. Fourth, in postoperative CT volumes, the metallic implants have very high intensity and cause severe imaging artifacts. Figure 4 shows a postoperative CT image in which this is visible. All these factors make our dataset very heterogeneous and, as mentioned above, make the automatic initialization of intensity-based registration algorithms challenging. Our current approach is to initialize these algorithms interactively either by manually rotating and translating the images or by selecting homologous points that are used to compute a rigid-body transformation. While effective, these methods necessitate loading images in a viewer and interacting with the images. This may not appear to be a major hindrance, but it requires skills and experience with this type of image and with three-dimensional (3-D) image manipulation software. As a result, it cannot easily be done by an end user such as an audiologist. More importantly, it prevents the development of fully automatic pipelines that are necessary for clinical translation, which is our ultimate goal. In this paper, we present a method to replace this interactive step to bring us closer to full automation.

Fig. 4 An example of postoperative CT image showing the imaging artifact caused by the CI. P_1 , which is a voxel near the CI, has much higher intensity than P_2 , which is inside the bone.

Our method relies on the detection of landmarks and a large body of work has been dedicated to the localization of landmarks in medical images (see Ref. 11 for instance). A complete review of the literature on the topic would be beyond the scope of this paper but some work that is particularly germane to what is presented herein includes the work by Potesil et al.¹² and Donner et al.,¹³ in which spatial compatibility between landmarks is taken into consideration using a parts-based graphical model. To localize eight aortic valve landmarks in C-Arm CT images, Zheng et al.¹⁴ impose spatial constraints between these by fitting to the images a model that comprises the complete set of landmarks. It leads to an approximate position for each landmark that is then refined for each of them individually using probabilistic boosting tree classifiers. For the purpose of localizing a target point in deep brain stimulation procedures using MR images, Liu and Dawant¹⁵ first automatically detected the anterior commissure, the posterior commissure, and the midsagittal plane using the method described in Ref. 16, and perform a rigid registration of all images based on the three detected structures. They then use the coordinates of the voxels as a feature along with intensity contextual features for the training and testing of a random forest-based¹⁷ detector. Here, as others have proposed,^{16,18–22} we use random forest regressors to automatically localize a number of landmarks. These landmarks that have also been localized in an atlas are utilized to compute a rigid-body transformation that is used to initialize our sequence of intensity-based algorithms. Because our landmarks are used for initialization purposes, we do not need the accuracy others may require and we show that our approach is sufficient for the task at hand. But, as will be shown, our method also lacks specificity, i.e., it is not capable of unequivocally localizing each landmark. Rather, it produces several likely candidates, from which the correct landmark set needs to be identified. To do so, we find among the possible candidates the set of landmarks whose spatial configuration matches the known spatial configuration of the landmarks in the atlas. Our technique is validated on the three aforementioned registration tasks (see Fig. 1), i.e., atlas to preoperative images, atlas to postoperative images, and preoperative to postoperative images. We use a dataset that includes both conventional and xCAT® images. We show that our initialization technique leads to final results, i.e., results obtained after the intensity-based steps, that are equivalent to those obtained when initialization is performed manually, which is the procedure we currently use.

Table 1 Dataset used in the study.

Image type	Preoperative CT		Postoperative CT	
	Conventional	Xoran	Conventional	Xoran
Number	228	59	16	80

2 Methods

2.1 Data

Table 1 shows the data that have been used in this study. They consist of 383 image volumes (one ear/volume is used) that were available at the time the experiments were made. 287 of those are preoperative images that have been acquired with a conventional scanner (228 image volumes) and with an xCAT[®] scanner (59 image volumes). The remaining 96 volumes are postoperative images also acquired with a conventional scanner (16 image volumes) and with an xCAT[®] scanner (80 image volumes). The acquisition parameters of the images in our dataset are shown in Table 2. As discussed below, two sets of experiments have been performed. In the first set of experiments, the training set consists of 83 preoperative image volumes acquired with a conventional scanner; the testing set consists of 145 preoperative volumes acquired with a conventional scanner and of 30 preoperative volumes acquired with an xCAT[®] scanner. In the second set of experiments, the training set also consists of 83 preoperative image volumes but 54 of those are acquired with a conventional scanner and 29 with an xCAT[®] scanner. In this set of experiments, the testing set consists of 175 preoperative volumes and of 96 postoperative volumes. 145 of the preoperative volumes are acquired with a conventional scanner and 30 with an xCAT[®] scanner (these are the same volumes that are used in experiment one). 16 of the postoperative images are acquired with a

conventional scanner and 80 with an xCAT[®] scanner. All images are downsampled to $2.25 \times 2.25 \times 2.25$ mm³ voxels both for the training of the regressors and for testing. This may appear coarse but, as the results will show, is sufficient for our main goal, i.e., the initialization of intensity-based registration algorithms.

The models are trained on the left ears. In each test image, localization of landmarks and registration are done for one ear. This is done because we have shown a high degree of symmetry between the right and the left ear.²³ When testing a volume that contains a right ear, the volumes are mirrored with respect to the middle sagittal image plane, which is typically the midsagittal plane of the head.

As discussed, our technique relies on the localization of landmarks. To produce good initialization results around the ear, these should surround it. They should also be visually distinguishable to permit their manual localization. The seven landmarks that meet these criteria and have been selected are shown in Fig. 5. Each of these has been localized manually in all volumes.

2.2 Random Forest-Based Localization of Candidate Landmarks

As proposed by Pauly et al.,²⁴ we compute a vector of textural features for each voxel. Specifically, we apply a displacement to a voxel x , calculate the mean intensities of a 3-D cuboidal region R_x^s centered on x and region $Q_x^{s,m}$ of the same size but centered on the displaced voxel, and subtract these two

$$f^m = \frac{1}{|R_x^s|} \left[\sum_{x' \in Q_x^{s,m}} I(x') - \sum_{x' \in R_x^s} I(x') \right], \quad (1)$$

where I is the intensity, and s is the current scale, i.e., a particular size of the cuboidal region. Four scales are used with corresponding cube sizes of $s = (2, 4, 8, 16)$. This process is repeated for $M = 2000$ different displacements (500 at each scale) to obtain the feature set $\{f^m\}_{m=1}^M$. At each scale, the

Table 2 Acquisition parameters for conventional and xCAT[®] CT images.

	kVp (V)	Exposure time (ms)	X-ray tube current (mA)	Slice size (pixels)	In-plane resolution (mm)	Slice thickness (mm)
Conventional CT	100 to 140	500 to 7450	30 to 450	410 × 410 to 768 × 768	0.13 to 0.49	0.22 to 2.5
Xoran CT	120	9600/11,200	6	640 × 640 to 812 × 812	0.3/0.4	0.3/0.4

Fig. 5 All seven landmarks, marked by crosses shown on (a) axial, (b) coronal, and (c) sagittal views.

displacements are obtained by uniformly sampling the $14 \times 14 \times 14$ voxel region centered at voxel x . With a truncated Gaussian function, each point is assigned a probability p to be the landmark as follows:

$$p(x) = e^{-\frac{d^2}{2\sigma^2}}, \quad (2)$$

where $\sigma = 2$ voxels is the standard deviation of the Gaussian function and d is the Euclidian distance between the voxel and the true landmark. We truncate the values of this function below $p = 0.1$ to 0 to limit the number of values and speed up the training process. Given the number of training pairs $\{\tilde{f}_n, p_n\}_{n=1}^N$, random forest regressors are trained to learn a non-linear mapping from the feature space $\{\tilde{f}\}$ to the probability space $\{p\}$.

In the training phase, all the voxels in a $21 \times 21 \times 21$ cuboidal region centered on the manually selected landmark are used as training samples to create seven models (one per landmark). We use 20 regression trees to construct the forest. For each tree, all the training samples are fed to the root node. Given the training samples $\{\tilde{f}_n, p_n\}_{n=1}^{N'}$, a feature f^m and a threshold t are selected at each node to best split the data. This is achieved by minimizing the mean squared error (MSE), i.e.,

$$t, m = \arg_{t,m} \min \text{MSE}(\{p_n : f_n^m < t\}) + \text{MSE}(\{p_n : f_n^m \geq t\}). \quad (3)$$

By design, when the number of samples arriving at leaf nodes is smaller than 5, or the best split threshold cannot be found or a maximum tree depth of 10 is reached, the tree stops growing. By preventing the regression trees from growing when they are deep enough or when the number of samples in one node is small enough, the risk of overfitting is reduced. Each leaf stores the mean probability of all training samples associated with it to be the landmark. When a test sample is fed to the forest, the probabilities of the leaf node to which it is associated in each tree are averaged and used to determine the probability that it is the landmark for which the forest is trained.

In the testing phase, the regression forests are applied to the entire volumes to produce response maps, one for each of the seven landmarks. In these maps, the value of a voxel is its probability to be the landmark of interest. Ideally, each map would have one single maximum that corresponds to the landmark for which the forest is trained. Results will show that this is not the case. Each map typically contains multiple local maxima and the global maximum does not always correspond to the actual landmark position. To select the correct landmark set, i.e., the best landmark in each of the seven response maps, we use the known spatial relation between landmarks as discussed in the next section.

2.3 Selection of the Landmark Set Using Spatial Constraints

To select the correct maximum in each map, we use the relative position of the maxima in all the maps. To do so, we rely on the fact that (1) the landmarks we use surround the inner ear and (2) the shape of the inner ear does not change very substantially across subjects. If two sets of points in two image volumes correspond to the actual landmarks, these two sets should thus be approximately related by a rigid-body transformation. To select

the correct set of landmarks for a particular volume, we first threshold each of the probability maps to eliminate spurious local maxima. The threshold is the maximum probability of the response map divided by a positive number $d (d \geq 1)$ that will be discussed in Sec. 3.3. All local maxima in each of the thresholded maps are then localized, which produces a set of possible positions for each landmark. Here, a local maximum is identified as a voxel with higher probability than any other voxel inside a $17 \times 17 \times 17$ cuboidal region. Next, a rigid-body transformation is computed between each landmark set configuration, i.e., one particular set of landmark candidates, and the landmark positions in a reference volume (the atlas) using a standard least squares method.²⁵ The landmark set that produces the configuration that leads to the smallest fiducial registration error (FRE),²⁶ i.e., the distance after registration between homologous landmarks that have been used to estimate the transformation is selected as the solution.

To find the best configuration, the simplest and most reliable way would be to perform an exhaustive search, i.e., to generate all possible landmark configurations and test each of them. This is, however, computationally prohibitive. Indeed, suppose that for the i 'th landmark, the number of possible positions, i.e., the number of local maxima in the response map that have been kept, is N_i ($i = 1, 2, \dots, L$). In this case, the total number of configurations and thus of point-based registrations that need to be computed is

$$N = \prod_{i=1}^L N_i. \quad (4)$$

Rather than considering all possible configurations, we rely on a heuristic search technique: we begin with identifying the first two landmarks by computing the Euclidean distance between all pairs of landmarks in their respective thresholded response maps. These distances are compared to the distance between the first two landmarks in the atlas. The k landmark configurations with distances closest to the reference distance are kept to form the current solution set. Next, we take the third landmark into consideration. The k configurations in the current solution set are augmented with each candidate for the third landmark. This produces a set of candidate triplet configurations. A rigid-body transformation is computed between each of these and the landmark points in the atlas. The transformation is computed using the algorithm in Ref. 25, which is a commonly used technique to estimate a rotation matrix and a translation vector to minimize the root-mean-squares error between homologous points. The FRE is then computed for each configuration. Because the smaller the FRE, the better the landmarks can be aligned with a rigid body transformation, we use the FRE to rank the solutions and we again keep the k best ones. This procedure is repeated for all subsequent landmarks. When all seven landmarks have been included, the configuration that leads to the smallest FRE is kept as the final solution. Using this approach, the maximum number of configurations we need to evaluate is

$$N_h = N_1 N_2 + k \sum_{i=3}^L N_i. \quad (5)$$

The parameter k permits balancing of the robustness and the computation cost of the algorithm. When k is small, the

algorithm is fast but the correct solution may be discarded in the process. When k is large, the correct solution is more likely to be found but the computational cost will increase. Even for relatively large values of k , this approach is much faster than an exhaustive search.

After the landmark set is localized in the target image, it is used to compute a rigid-body transformation between the atlas and the target image.

2.4 Timing

The random forest regressors are trained in Python using scikit-learn.²⁷ The testing has been first prototyped in MATLAB and Python and then implemented in C++. While it takes roughly 30 min to test each image in the prototype code (most of the time is spent on feature extraction), our C++ multithreaded implementation takes an average of 13.08 s per case on a standard PC [Intel (R), Xeon (R) CPU X5570, 2.93 GHz, 48 GB RAM, 16 cores]. In the 13.08 s, the computation of the feature vectors for all the voxels in one image requires 4.43 s. Creation of probability maps requires 6.03 s. Identifying the true landmarks among the possible candidates with our heuristic search algorithm requires 20 ms. Loading image and random forest models, resampling images, and writing results to the output file require 2.6 s. We also calculated the average time t_p to compute one rigid body transformation and found it to be 2.31×10^{-6} s. With an average number of candidates for each landmark equal to 43, the number of possible combinations is $43^7 = 2.72 \times 10^{11}$. The time to perform an exhaustive search rather than using our heuristic approach to find the best set of landmarks would thus be 6.28×10^5 s = 174.4 h, which is prohibitive.

2.5 Evaluation

To test the proposed method, we measure landmark localization error by both comparing our results with manual localization and assessing the effect that differences between manual and automatic initializations of the intensity-based algorithms have on the final results. To do the latter, we register images using three different schemes illustrated in Fig. 6. In scheme I, the process is initialized with the identity transformation, i.e., we use the volumes as they are acquired. This is followed by an intensity-based step applied to the entire volume to estimate a global affine transformation. In the last step, we crop a region of interest (ROI) that contains the inner ear as shown in Fig. 7, and we apply the same intensity-based algorithm we used in step

Fig. 7 The red bounding box shows the ROI in an axial cross section.

2 to this ROI to compute a local affine transformation. Scheme II is similar to scheme I but the process is initialized manually (this is the procedure we currently use when scheme I fails). In scheme III, we use our automatic method to initialize the process. Note that because the automatic method produces a transformation that registers the ears, there is no need to estimate a global transformation. Scheme III thus only requires a local intensity-based registration step.

With the transformations computed with each of the schemes, landmark points are projected from the atlas onto the test volumes, and the position of the projected landmarks is compared to the manually selected positions for each of the landmarks in each of our test volumes. The mean Euclidean distance between projected and manually selected landmark positions is used as the quality measurement.

3 Results

3.1 Landmark Set Localization

Figure 8 shows the response maps for the seven landmarks for one representative volume, and it illustrates the multiple maxima problem. Table 3 reports the landmark localization errors we have observed. In this table and in the remainder of this paper, a failure is a case with localization error larger than 10 mm, and failed cases are not included when computing error statistics. In our first set of experiments with random forests trained on images acquired with a conventional scanner, we did not

Fig. 6 A flowchart of the scheme used to evaluate the proposed automatic initialization method.

Fig. 8 The axial view of the probability maps of one test CT, illustrating the multiple-maxima problem.

Table 3 Statistics of the mean landmark set localization error when different regressors are used for both preop and postop CT images.

Regression forest	Model trained using conventional CTs		Robust model trained using conventional + Xoran CTs			
	Conventional preop (145)	Xoran preop (30)	Conventional preop (145)	Xoran preop (30)	Conventional postop (16)	Xoran postop (80)
Dataset (# of volumes)						
Mean (mm)	3.11	3.55	3.01	4.17	4.52	4.58
SD (mm)	0.91	0.5	0.8	1.01	1.74	1.24
Failure rate (%)	0	10	0	0	0	2.50

observe failures when the testing set contains only preoperative images of this type. It, however, rises to 10% when tested on preoperative images acquired with an xCAT[®] scanner. A likely reason for this discrepancy is the difference in intensity characteristics between the two types of scanner. As discussed in Sec. 2, this was addressed by modifying the training set to include images acquired both with conventional and xCAT[®] scanners. With this training set and the same testing set, the failure rate drops to zero for both the conventional and the xCAT[®] images. The same regressors were used to localize landmarks in postimplantation volumes acquired both with a conventional scanner and with an xCAT[®] scanner. We did not observe any failure for the former. Our approach failed for only two xCAT[®] volumes (2.5%), despite the presence of artifacts caused by the electrode array.

3.2 Effect of Landmark Localization Error on Final Registration

3.2.1 Registration between atlas and preoperative/postoperative images

In our IGCIP technique, a registration is typically done between the atlas and a preoperative CT. When a preoperative CT is not available, registration between the atlas and the postoperative CT is necessary. In both scenarios, the same atlas is used and the manually labeled landmark set on the atlas and the automatically localized landmark set on the target image are utilized to compute the initial rigid-body transformation. Table 4 reports differences between manually localized landmarks in each of the test volumes and the position of the atlas landmarks projected onto each of the test volume using the three strategies described in Sec. 2. Without any initialization, we observe a high failure rate indicating that differences in orientation

between the atlas, and the acquired images exceed the capture range of our global intensity-based algorithm. There is no statistically significant difference between manual initialization (scheme II) and our automatic technique (scheme III) except for the conventional postop CT test set for which the automatic error is lower ($p < 0.05$ for a two-sided Wilcoxon signed rank test).

3.2.2 Registration between preoperative and postoperative images

As discussed earlier, in the normal IGCIP process, preoperative and postoperative images are available. The ear anatomy is segmented in the preoperative images, the electrode array is localized in the postoperative images, and both images are registered to show the position of the array with respect to the anatomy. Registering pre- and postoperative images involves a rigid-body transformation, the computation of which may appear to be trivial. But Table 4 shows that because of differences in coverage and head orientation between pre- and postoperative scans, relying only on intensity-based algorithms leads to a high failure rate. Intensity-based algorithms thus need to be initialized either manually or using the set of landmarks localized in the preoperative and postoperative images. Here, we treat the pre- and postoperative images differently to reflect what is currently done in our processing pipeline. In this pipeline, the affine intensity-based registration step described above to register the atlas and the preoperative images is followed by a nonrigid intensity-based registration step²⁵ that accurately registers the atlas and the preoperative images. When evaluating our method on the pre- to postoperative registration task, the position of the landmarks is obtained by projecting them from the atlas to the preoperative volume using the composite (rigid, affine, and nonrigid) transformation that registers these

Table 4 Statistics of the registration error when no initialization (scheme I), manual initialization (scheme II), and automatic initialization (scheme III) is used for the intensity-based registration.

Registration from atlas to preop/postop CTs						
Dataset (# of volumes)	Conventional preop CT (145)			Xoran preop CT (30)		
Schemes	Mean (mm)	SD (mm)	Failure rate (%)	Mean (mm)	SD (mm)	Failure rate (%)
I	1.6	0.85	15.2	2.75	0.74	16.70
II	1.47	0.45	0.0	2.76	0.83	0.0
III	1.49	0.49	0.0	2.78	0.84	0.0
Dataset (# of volumes)	Conventional postop CT (16)			Xoran postop CT (80)		
Schemes	Mean (mm)	SD (mm)	Failure rate (%)	Mean (mm)	SD (mm)	Failure rate (%)
I	3.1	0.59	31.3	2.17	0.51	61.30
II	3.58	0.98	0.0	2.24	0.74	0.0
III	3.04	0.69	0.0	2.46	1.09	0.0
Registration from preop CT to postop CT						
Dataset (# of volumes)	Conventional CT (7)			Xoran CT (76)		
Schemes	Mean (mm)	SD (mm)	Failure rate (%)	Mean (mm)	SD (mm)	Failure rate (%)
I	1.34	0.63	28.6	1.36	1.25	49.70
II	1.65	1.02	0	1.11	0.37	0
III	1.71	1.01	0	1.12	0.36	2.60

volumes. We assume that through this process landmarks are localized accurately in the preoperative images, and this is confirmed by doing a visual check. In the postoperative images, we localize the landmarks using the proposed method. We then compare the same three approaches to compute the preop-to-postop intensity-based rigid-body transformation that were used to compute the atlas-to-volume affine transformations, i.e., no initialization, manual initialization, and landmark-based initialization. The bottom section of Table 4 reports the results that were obtained; in this experiment, all preoperative images were conventional CTs. The proposed approach failed for two postoperative xCAT[®] volumes.

3.3 Sensitivity to Parameter Values

The performance of the heuristic search algorithm we use to localize the set of landmarks is influenced by two main parameters: (1) the value that is used to threshold the probability map to identify potential candidates, which we call T and (2) the number of solutions we keep in the solution set at each iteration, which we call k . The value of the threshold T is defined as follows:

$$T = \frac{M}{d}, \quad (d \geq 1). \quad (6)$$

Fig. 9 Localization failure rate as a function of d . **Fig. 10** Localization failure rate as a function of k .

In this expression, M is the maximum value in the probability map and d is a parameter. When $d = 1$, only one landmark candidate is kept, i.e., the point at which the probability map is maximum. As d increases, so does the number of candidates. The values of d and k were set heuristically to 8 and 32, respectively, for all the results presented in this study.

To study the sensitivity of the results to these parameter values, we retrospectively repeated the localization experiments with a range of values. Figures 9 and 10 show the ratio of localization failures among all cases we have observed as a function of the parameter values. These figures show that the performance of our approach reaches an asymptote for values of d larger than 4 and values of k larger than 16.

4 Conclusions and Discussions

In this paper, we propose a method to automatically localize a set of landmarks in head CT images to assist image registration initialization. Random forest regression is used to produce a number of candidates, which are then pruned using *a priori* information about the spatial relationship between landmarks. For the pruning process, we developed a fast and reliable heuristic technique to find the right configuration. Our technique is validated on a large heterogeneous dataset, made of both conventional and xCAT[®] preoperative CTs, as well as conventional and xCAT[®] postoperative CTs. After the localization is done, we have shown on this dataset that the automatically localized landmark set can be used to estimate an initial transformation for registrations that could replace the manual method we currently use. We evaluate this technique by doing registrations from atlas to preoperative CTs, atlas to postoperative CTs, and preoperative CTs to postoperative CTs. This is an important step toward the full automation of our processing pipeline, which is required for the large-scale evaluation of our IGCIP technique.

Disclosures

This work was performed under an IRB protocol for human data approved by Vanderbilt University. No conflicts of interest, financial or otherwise, are declared by the authors.

Acknowledgments

This research was supported by NIH Grant Nos. R01DC014037, R01DC008408, and R01DC014462 from the National Institute of Deafness and Other Communications Disorders. The experiments have been conducted with the support of the Advanced Computing Center for Research and Education (ACCRE) at Vanderbilt University.

References

1. J. H. Noble et al., "Image-guidance enables new methods for customizing cochlear implant stimulation strategies," *IEEE Trans. Neural Syst. Rehabil. Eng.* **21**(5), 820–829 (2013).
2. J. H. Noble et al., "Clinical evaluation of an image-guided cochlear implant programming strategy," *Audiol. Neurotol.* **19**(6), 400–411 (2014).
3. J. H. Noble et al., "Initial results with image-guided cochlear implant programming in children," *Otol. Neurotol.* **37**(2), e63–e69 (2016).
4. J. H. Noble et al., "Automatic segmentation of intracochlear anatomy in conventional CT," *IEEE Trans. Biomed. Eng.* **58**(9), 2625–2632 (2011).
5. J. H. Noble and B. M. Dawant, "Automatic graph-based localization of cochlear implant electrodes in CT," in *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham, pp. 152–159 (2015).

6. Y. Zhao et al., "Automatic localization of cochlear implant electrodes in CT," *Med. Image Comput. Comput. Assist. Interv.* **17**, 331–338 (2014).
7. Y. Zhao et al., "Automatic localization of cochlear implant electrodes in CTs with a limited intensity range," *Proc. SPIE* **10133**, 101330T (2017).
8. F. A. Reda et al., "An artifact-robust, shape library-based algorithm for automatic segmentation of inner ear anatomy in post-cochlear-implantation CT," *Proc. SPIE* **9034**, 90342V (2014).
9. F. Maes et al., "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imaging* **16**(2), 187–198 (1997).
10. G. K. Rohde et al., "The adaptive bases algorithm for intensity-based nonrigid image registration," *IEEE Trans. Med. Imaging* **22**(11), 1470–1479 (2003).
11. K. Rohr, *Landmark-based Image Analysis: Using Geometric and Intensity Models*, Vol. **21**, Kluwer Academic Publishers (2001).
12. V. Potesil et al., "Improved anatomical landmark localization in medical images using dense matching of graphical models," in *Proc. of the British Machine Vision Conf. (BMVC)*, Vol. 4, No. 6 (2010).
13. R. Donner et al., "Global localization of 3D anatomical structures by pre-filtered Hough Forests and discrete optimization," *Med. Image Anal.* **17**(8), 1304–1314 (2013).
14. Y. Zheng et al., "Automatic aorta segmentation and valve landmark detection in C-arm CT for transcatheter aortic valve implantation," *IEEE Trans. Med. Imaging* **31**(12), 2307–2321 (2012).
15. Y. Liu and B. M. Dawant, "Multi-modal learning-based pre-operative targeting in deep brain stimulation procedures," in *2016 IEEE-EMBS Int. Conf. on Biomedical and Health Informatics (BHI)*, IEEE, pp. 17–20 (2016).
16. Y. Liu and B. M. Dawant, "Automatic localization of the anterior commissure, posterior commissure, and midsagittal plane in MRI scans using regression forests," *IEEE J. Biomed. Health Inf.* **19**(4), 1362–1374 (2015).
17. L. Breiman, "Random forests," *Mach. Learning* **45**(1), 5–32 (2001).
18. A. Criminisi et al., "Regression forests for efficient anatomy detection and localization in computed tomography scans," *Med. Image Anal.* **17**(8), 1293–1303 (2013).
19. M. A. Dabbah et al., "Detection and location of 127 anatomical landmarks in diverse CT datasets," *Proc. SPIE* **9034**, 903415 (2014).
20. D. Han et al., "Robust anatomical landmark detection with application to MR brain image registration," *Comput. Med. Imaging Graphics* **46**, 277–290 (2015).
21. T. Ebner et al., "Towards automatic bone age estimation from MRI: localization of 3D anatomical landmarks," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2014)*, pp. 421–428 (2014).
22. C. Lindner et al., "Fully automatic system for accurate localisation and analysis of cephalometric landmarks in lateral cephalograms," *Sci. Rep.* **6**, 33581 (2016).
23. F. A. Reda et al., "Automatic segmentation of intra-cochlear anatomy in post-implantation CT of unilateral cochlear implant recipients," *Med. Image Anal.* **18**(3), 605–615 (2014).
24. O. Pauly et al., "Fast multiple organ detection and localization in whole-body MR Dixon sequences," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2011)*, Springer Berlin Heidelberg, pp. 239–247 (2011).
25. J. L. Farrell et al., "Problem 65-1: a least squares estimate of satellite attitude," *SIAM Rev.* **8**, 384–386 (1966).
26. J. M. Fitzpatrick et al., "Predicting error in rigid-body point-based registration," *IEEE Trans. Med. Imaging* **17**(5), 694–702 (1998).
27. F. Pedregosa et al., "Scikit-learn: machine learning in python," *J. Mach. Learning Res.* **12**, 2825–2830 (2011).

Dongqing Zhang received his BS degree in electrical and information engineering from the University of Science and Technology of China, Hefei, China, in 2013. He is currently a PhD student in the Department of Electrical Engineering and Computer Science, Vanderbilt University. His research interest is image-guided surgery, medical image processing, and machine learning.

Yuan Liu received her BEng degree in computer science from the Northwestern Polytechnical University, Xi'an, China, in 2010,

and her PhD in electrical engineering and computer science from Vanderbilt University, Nashville, Tennessee, USA, in 2016. She is currently a software engineer at Google Inc. Her research interests include image processing, computer vision, machine learning, and big data analytics.

Jack H. Noble received his BE, MS, and PhD degrees in electrical engineering from Vanderbilt University, Nashville, Tennessee, USA, in 2007, 2008, and 2011, respectively. He is currently an assistant professor in the Department of Electrical Engineering and Computer Science, Vanderbilt University. His primary research

interests include medical image processing, image segmentation, registration, statistical modeling, and image-guided surgery techniques.

Benoit M. Dawant received his MSEE degree from the University of Louvain, Leuven, Belgium, in 1983, and his PhD from the University of Houston, Houston, Texas, in 1988. He is a Cornelius Vanderbilt professor and a professor of electrical engineering at Vanderbilt University. He is also the director of Vanderbilt Institute for Surgery and Engineering and a fellow of the IEEE. His main research interests include medical image processing and analysis.