Published in final edited form as:

Proc SPIE Int Soc Opt Eng. 2016 February 27; 9784: . doi:10.1117/12.2216925.

Automatic localization of landmark sets in head CT images with regression forests for image registration initialization

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

Dept. of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN 35235, USA

Abstract

Cochlear Implants (CIs) are electrode arrays that are surgically inserted into the cochlea. Individual contacts stimulate frequency-mapped nerve endings thus replacing the natural electro-mechanical transduction mechanism. CIs are programmed post-operatively by audiologists but this is currently done using behavioral tests without imaging information that permits relating electrode position to inner ear anatomy. We have recently developed a series of image processing steps that permit the segmentation of the inner ear anatomy and the localization of individual contacts. We have proposed a new programming strategy that uses this information and we have shown in a study with 68 participants that 78% of long term recipients preferred the programming parameters determined with this new strategy. A limiting factor to the large scale evaluation and deployment of our technique is the amount of user interaction still required in some of the steps used in our sequence of image processing algorithms. One such step is the rough registration of an atlas to target volumes prior to the use of automated intensity-based algorithms when the target volumes have very different fields of view and orientations. In this paper we propose a solution to this problem. It relies on a random forest-based approach to automatically localize a series of landmarks. Our results obtained from 83 images with 132 registration tasks show that automatic initialization of an intensity-based algorithm proves to be a reliable technique to replace the manual step.

Keywords

Image Registration; Landmark Localization; Regression Forest

1. INTRODUCTION

Cochlear implants are electrode arrays that are surgically inserted into the cochlea. Individual contacts in the array stimulate frequency-mapped nerve endings, thus replacing the natural electro-mechanical transduction mechanism. Post-operatively, the electrode array 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 IGCIP for Image-Guided Cochlear Implant Programming [1]. This technique relies on a series of image processing steps that permit the segmentation of the inner ear structures in pre-operative CT images and the localization of the implant in post-operative CT images. Using this information, we have designed techniques that assist the audiologist in programming the implants. Results we have obtained in a retrospective study performed with 68 recipients show that in 78% of the cases the parameter settings selected with imaging assistance were preferred to those selected clinically with the current clinical methodology [2]. In [3] we show on 21 ears that 85.7% of pediatric recipients also prefer the settings selected with imaging assistance. IGCIP thus has the potential to profoundly change the way CIs are programmed. One barrier to the large scale deployment and evaluation of our methods is the lack of full automation. Indeed, several steps in our sequence of algorithms still require manual intervention. One such step is the registration of an atlas CT image volume that is used for atlas-based segmentation to pre-operative CT images. Although this problem may appear trivial, a few properties of our dataset make it challenging. First, the image sets often have very different fields of view. Instead of covering the whole head, they can be a slab covering the left and right temporal bones, or one half of the head, or even just one temporal bone. Second, the orientation of different scans can vary substantially, exceeding the capture range of traditional registration algorithms. Figure 1 shows some examples of our dataset that exhibit such properties. Our current approach is to initialize a sequence of affine and non-rigid intensity-based registration steps interactively either by manually rotating and translating the images or by selecting homologous points that are used to compute a rigid body transformation.

In this paper we present a method to replace this interactive steps: First, we define a set of landmarks surrounding the ear (the Region of Interest) on the atlas image. Second, by labeling those landmarks on a set of training images, we create a training set that is used to train a random forest regressor that maps the textural features of voxels in the image to their probabilities to be the true landmark. Third, for a test image, we use these trained random forest regressors to produce probability maps, one for each landmark. The maps represent the probability of each voxel in a test image to be a landmark. To eliminate false candidates we rely on the a-priori knowledge on the spatial relationships between landmarks. Results will show that this approach permits us to localize the landmarks with an accuracy that is good enough to compute a standard point-based rigid-body transformation that can be used to initialize an intensity-based affine registration step. We evaluate the performance of the intensity-based registration when no initialization, manual initialization and automatic initialization are used. We show that no initialization can lead to catastrophic failures and that manual and automatic initializations do not and produce results that are statistically equivalent.

2. METHODS

To initialize our intensity-based algorithms we automatically localize 7 distinctive landmarks with a random forest regression approach [4, 5], and use those landmarks to estimate a rigid body transformation [6]. These landmarks are shown in Figure 2. As proposed by Pauly *et al.* [7], 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 3D

cuboidal region R_x^s centered on x and of a similar 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 window sizes of 2, 4, 8, and 16. This process is

repeated $M = 2000$ times (500 times at each scale) to obtain the feature set $\{f^m\}_{m=1}^M$. At each scale, the displacements are obtained by uniformly sampling a cube centered at voxel x . With a truncated Gaussian function, each point is assigned a probability p to be the landmark as follows, $p(x) = e^{-d^2/2\sigma^2}$, where σ is the standard deviation of the Gaussian function and d is the Euclidian distance between the voxel and the true landmark. We truncate this function at $p = 0.1$ to speed up the training process. This limits the range of values and speeds up the

training process. Given a number of training pairs $\{\vec{f}_n, p_n\}_{n=1}^N$, random forest regressors are trained to learn a nonlinear mapping from the feature space $\{\vec{f}\}$ 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 to create 7 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 $\{\vec{f}_n, p_n\}_{n=1}^{N'}$, 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.,

$$t, m = \arg_{t,m} \min MSE(\{p_n: f_n^m < t\}) + MSE(\{p_n: f_n^m \geq t\}) \quad (2)$$

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. 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 test phase feature vectors are computed for all the voxels in the image and a response map is generated that represents the probability of each voxel to be the landmark.

For each test volume, 7 maps (one for each landmark) are created. Ideally, each map would have a clear maximum indicating the position of the landmark. This is currently not the case. Every map could contain multiple local maxima and the global maximum does not always correspond to the actual landmark position. 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. All local maxima in each of the thresholded map 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 cuboidal region (the size of the cuboidal region we set is $17 \times 17 \times 17$) centered at it. Next, a rigid body transformation is computed between each landmark configuration, i.e., one particular set of landmark candidates, and the landmark positions in a reference volume (the atlas). The landmark positions that pertain to the configuration that leads to the least fiducial registration error are selected as the solution.

To find the best configuration, the simplest and most reliable way would be to perform an exhaustive search, i.e., 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 maxima in the response map that have been kept, is N_i ($i = 1, 2, \dots, 7$). In this case, the total number of configurations and thus of point-based registrations that need to be computed is

$$N = \prod_{i=1}^7 N_i \quad (3)$$

Rather than considering all possible configurations we rely on a heuristic search technique: We begin with identifying the first two landmarks by first computing the Euclidean distance between all pairs of landmarks in their respective thresholded response maps. These distances are compared with the distance between the first 2 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 triplet candidate configurations. A rigid body transformation is computed between each of these and the landmark points in the atlas using a standard point-based registration method and the fiducial registration error (FRE) [9] is computed for each configuration. 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. The parameter k permits to balance the robustness and the cost of the algorithm. When k the correct solution may be discarded in the process. When k is big, the correct solution is more likely to be found but the computation cost increases. Even with larger values of k this approach is substantially faster than an exhaustive search.

After the landmark set is localized in a new image, a rigid-body transformation between the atlas and this image can be estimated. To evaluate our approach, we use 3 registration schemes as shown in Figure 3. Scheme I: we use the identity transformation to initialize the process; we then compute a global, i.e., we use the entire volume, affine transformation using

an intensity-based registration method; we crop a region of interest (ROI) that contains the inner ear and is shown in Figure 4; and we compute a second, local, transformation with the same intensity-based method applied to this ROI. Scheme II: it is similar to Scheme I except that the process is initialized manually (this is the procedure we currently use when Scheme I fails). Scheme III: since the landmarks are all included in the ROI drawn in the atlas, we register the atlas landmarks with the target landmarks. We can thus skip the global intensity-based registration step and only rely on the local affine transformation step to compute the final transformation. With the transformations computed with each of the schemes, landmark points are projected from the atlas onto the test volume and the position of the projected landmarks is compared to “ground truth” positions selected manually for each of the landmarks in each of our target images. The mean Euclidean distance between projected and manually selected landmark positions is used as the quality measure.

3. RESULTS

We have acquired pre-operative head CT volumes for 166 patients who underwent a cochlear implantation surgery. These images have been obtained using scanners from different manufacturers with various resolutions. Half of them are used to train the random forest regressors and another half to test. In the test phase, since the CT scans could cover either both temporal bones or just one of them, 132 localization and registration tasks are done in total. Because the models are trained on landmarks localized for the left ear, when detecting landmarks on the right ear, we mirror the image with respect to the middle y-z plane, which is roughly the mid-sagittal plane of the head.

Prior to processing, all volumes are down-sampled to isotropic dimensions ($2.25 \times 2.25 \times 2.25\text{mm}^3$). This is coarse but, as the results will show, sufficient for our main purpose, i.e., initialization of intensity-based algorithms. Figure 5 shows the response maps for the seven landmarks for one representative volume and it illustrates the multiple maxima problem. We evaluate the performance of our heuristic search algorithm when $k = 1, 2, 4, 8$. Figure 6 shows the results of this experiment. As can be seen, when $k = 1$, 9 localization failures indicated by large errors occurring, which means that erroneous local maxima have been identified as the correct landmarks in the process. With $k = 2$, this number decreases to 1. No localization failure happens when $k = 4$. To be conservative, we have selected $k = 8$. Figure 7 shows the localization error, i.e., the mean Euclidean distance between the manual landmark localization and the localization obtained with each of the three schemes we have compared. Table 1 presents values for the overall localization error for Scheme II and III. As can be seen, Scheme I leads to catastrophic registration failure in 12 cases, while both Scheme II & III do not produce any registration failure. Values presented in Table 1 show that Scheme II & III lead to comparable results, meaning that the proposed automatic initialization method is as good as the manual initialization method we currently use. A paired t-test between the mean localization errors using automatic initialization and those using manual initialization indicates that the difference is not statistically significant ($p=0.505$). The errors reported may not appear to be small but in our current processing pipeline, the intensity-based affine registration step is followed by a non-rigid intensity-based step and the accuracy reported in Table 1 is sufficient, in our experience, to initialize the non-rigid registration algorithm.

The random forest regressors are trained in Python using scikit-learn [8]. 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 computation is spent on feature extraction) our C++ implementation takes an average time of 27 seconds in a sequential implementation and 7.8 seconds in a multithreaded implementation to run each case on a standard PC (Intel (R), Xeon (R) CPU X5570, 2.93GHz, 48GB RAM, 16 cores).

4. CONCLUSIONS

In this work, we propose a method to automatically localize a set of landmarks in head CT images. 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. In the pruning process, to find the right configuration, we developed an efficient and reliable heuristic technique. We have shown on a large-scale dataset that the automatically localized landmark set can be used to estimate an initialization transformation that could replace the manual method we currently use to support our clinical studies. Its robustness is demonstrated on image sets from different scanners, having very different fields of view and substantially varying orientations. 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.

ACKNOWLEDGMENTS

This research has been supported by grants NIH R01 DC014037, R01DC008408, and R01DC014462 from the National Institute of Deafness and Other Communications Disorders. The content is solely the responsibility of the authors and does not necessarily represent the official views of this institute. This work was conducted in part using the resources of the Advanced Computing Center for Research and Education at Vanderbilt University, Nashville, TN.

REFERENCES

- [1]. Noble, Jack H., et al. Image-guidance enables new methods for customizing cochlear implant stimulation strategies. *Neural Systems and Rehabilitation Engineering*, IEEE Transactions on. 2013; 21.5:820–829.
- [2]. Noble, Jack H., et al. Clinical Evaluation of an Image-Guided Cochlear Implant Programming Strategy. *Audiology and Neurotology*. 2014; 19.6:400–411. [PubMed: 25402603]
- [3]. Noble JH, Hedley-Williams AJ, Sunderhaus LW, Dawant BM, Labadie RF, Gifford RH. Initial results with image-guided cochlear implant programming in children. *Otology & Neurotology*. 2015 in press.
- [4]. Criminisi, Antonio, et al. Regression forests for efficient anatomy detection and localization in computed tomography scans. *Medical image analysis*. 2013; 17.8:1293–1303. [PubMed: 23410511]
- [5]. Liu Y, Dawant BM. Automatic detection of the Anterior Commissure, Posterior Commissure, and midsagittal plane on MRI scans using regression forests. *IEEE Journal of Biomedical and Health Informatics*. Jul; 2015 19(no. 4):1362–1374. [PubMed: 25955855]
- [6]. Farrell JL, Stuelpnagel JC. Problem 65–1: A least squares estimate of satellite attitude. *SIAM Rev*. 1966; 8:384–386.
- [7]. Pauly, Olivier, et al. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011*. Springer; Berlin Heidelberg: 2011. Fast multiple organ detection and localization in whole-body MR Dixon sequences; p. 239-247.

- [8]. Pedregosa, Fabian, et al. Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*. 2011; 12:2825–2830.
- [9]. Fitzpatrick, J. Michael, Jay B., West, Calvin R., Maurer, Jr.. Predicting error in rigid-body point-based registration. *Medical Imaging, IEEE Transactions on*. 1998; 17.5:694–702.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.

Examples of images included in our dataset that illustrate the range of orientation and coverage that can occur. Top to bottom: axial, coronal, and sagittal views.

Figure 2.

The landmark set: For each landmark, they are axial (top), coronal (middle) and sagittal (bottom) views. Landmarks are labeled with yellow crosses.

Figure 3.
Flowchart of the validation procedure

Figure 4.
A slice of the atlas in axial view and the Region of Interest where we do the local registration

Figure 5. Probability maps (landmark #1~#7) of one test image, showing multiple local maxima problem (axial view)

Figure 6.
Number of localization failures when different k are used in our heuristic search technique

Figure 7.
 Mean landmark localization error of each case when Scheme I (up), Scheme II (middle) and Scheme III (bottom) are used

Table 1

Mean, standard deviation and maximum of landmark localization error of all test images

	Scheme II	Scheme III
Mean(mm)	1.40	1.42
SD(mm)	0.46	0.52
Max(mm)	3.21	4.24