

Efficient Segmentation Pipeline Using Diffeomorphic Image Registration: A Validation Study

Shalin Parikh¹, Anupama Goparaju¹, Riddhish Bhalodia¹, Bosten Loveless², Alan, Morris², Joshua Cates², Evgueni Kholmovski^{2,3}, Nassir Marrouche², Shireen Elhabian¹

¹ Scientific Computing and Imaging Institute, University of Utah, Salt Lake City, Utah, USA

² Comprehensive Arrhythmia Research and Management Center, Division of Cardiovascular Medicine, School of Medicine, University of Utah, Salt Lake City, Utah, USA

³ Department of Radiology and Imaging Sciences, School of Medicine, University of Utah, Salt Lake City, Utah, USA

Abstract

Functional measurements of the left atrium (LA) in atrial fibrillation (AF) patients is limited to a single CINE slice midway through the LA. Nonetheless, a full 3D characterization of atrial functional measurements would provide more insights into LA function. But this improved modeling capacity comes at a price of requiring LA segmentation of each 3D time point, a time-consuming and expensive task that requires anatomy-specific expertise. We propose an efficient pipeline which requires ground truth segmentation of a single (or limited) CINE time point to accurately propagate it throughout the sequence. This method significantly saves human effort and enable better characterization of LA anatomy. From a gated cardiac CINE MRI sequence we select a single CINE time point with ground truth segmentation, and assuming cyclic motion, we register other images corresponding to all time points using diffeomorphic registration in ANTs. The diffeomorphic registration fields allow us to map a given anatomical shape (segmentation) to each CINE time point, facilitating the construction of a 4D shape model.

1. Introduction

The heart is an organ in constant motion, and its functional output is highly correlated to the shape changes of its chambers as they contract and expand over different cardiac phases. Thus, population models of cardiac structure over the entire cardiac cycle will be important tools for understanding how structural changes contribute to cardiac function and failure. The goal of catheter ablation for AF is to create areas of scarring in the LA wall, but the effects of this scarring on the function of the atria are not well understood and can contribute to heart failure. AF is characterized. AF is defined by the uncoordinated electrical

activity and contractions in the LA and is associated with significant increased mortality and morbidity. AF is ultimately the disease of the underlying cardiac tissue. Heart failure is a significant complication of AF, but its origins are not well understood. The development of an image-based measure for shape remodeling of the cardiac cycle will enable quantitative measurements of cardiac function and pathology that are not currently possible using other tools. The effects of catheter ablation on the function of the atria is also not well understood. Ablation intentionally introduces scarring of the LA wall[1], in order to block electrical conductivity. An inevitable side-effect of scarring in cardiac tissue is a decrease in the ability of that tissue to contract. Noninvasive descriptors of LA function, like those to be developed in this project, are needed in order to understand how different amounts of scarring and different scar patterns (e.g. pulmonary vein isolation versus ablation of atrial wall targets like rotors, additional lines or fibrosis debulking) affect LA function.[2] This understanding will lead to better-informed ablations and may significantly reduce complications such as heart failure and stroke risk after ablation. The cost of getting such a keen understanding requires segmentation of each 3D time point which is both expensive and time-consuming. There are many pitfalls of manual segmentation. It suffers from inter-subject variability. It also requires significant investment of time and manual effort. Expert clinicians also require hours to segment a single time point accurately. This amount of investment can be better directed towards other areas like CINE time point acquisitions. Thus we propose an efficient pipeline of operations that requires a single or limited CINE ground truth segmentation to accurately propagate throughout the time sequence. This provides a significant boost in saving human effort and time. We choose a single (or multiple) CINE time point from a gated sequence. The assumption of cyclic motion enables us to register other

images corresponding to the other time points using diffeomorphic registration in ANTs. This helps in creating a time point between shapes and allows us to propagate segmentation. We also present a criteria of choosing the best CINE time point(s) for manual segmentation.

2. Methods

In order to measure the dynamic atrial shape of a patient heartbeat, we are acquiring a gated cardiac CINE MRI (sequence). Cine MRI has relatively high in-plane resolution that is suitable for imaging and tracking the atrial wall deformations from frame to frame. However, cine MRI slices are relatively thick (5-8mm), thus in order to build this warp we first re-sample all the CINE volumes to have the same inter voxel spacing between slices. This helps us in maintaining consistency across scans and segmentation. Then we proceed onto choosing a time point which would serve as the reference time point for registering the other time points. The question lies within the fact that which or how many time points to choose. There were couple of options the first being choosing a time point which has been chosen by clinicians which is representative of the MRA acquisition. A single time point does give promising results but it falters when registering with time points which are farther away than the reference time points in the cardiac cycle. As the structural difference between the singly chosen reference time point and the other time points changes the warp that we obtain is not powerful enough to capture all the shape variability amongst time points. This is evident from the fact that the DICE coefficients are much higher of time points closer to the singly chosen reference time point. This same behavior is also reflected in the distance map overlay of the segmentation. The segmentation closer to the reference time point are more accurate and resemble ground truth segmentation correctly.

Then we proceed onto choosing reference time points using uniform sampling. This allows us to space out the reference time points more evenly across the cardiac cycle in an attempt to capture the shape variability. A similar pattern of results are observed as expected. The time points closer to the reference time points chosen in the cardiac cycle have better segmentation produced. This is partly because the structural change between time points is less and thus can predict more accurate segmentation.

Uniform sampling is a reasonable approach in choosing reference time points in the cardiac cycle but till this point we have not taken into the considerations the biological changes that the heart goes through anatomically throughout the cardiac cycle. The diastolic period of the cycle where the chambers of the heart are filled in with blood is much slower and the structural changes are not that significant. While in the diastolic period of the cardiac

cycle where the heart chambers contract and pump out the blood is much faster. This results in considerable structural changes between time points in the cardiac cycle. Building on the above mentioned theme we divide time points into smaller subsets grouping time points having similar structure together. Then we choose the representative time point of that subset as the reference time point for that subset. This technique helps us in using the inherent characteristics of the time points in the cardiac cycle in choosing the most appropriate reference time point.

The choice of reference time points and the number of them is completely dependent of the budget of the study being conducted. As we all know manual segmentation is both expensive and time consuming. The number of time points can be chosen on basis of available manpower. The more number of time points guarantees more accurate segmentation. This approach also allows for reasonably accurate segmentation and can be employed to conserve resources which could be better suited for other tasks.

Once the time points itself and the number are chosen we register all of them together. We use diffeomorphic registration to register the time points.[3] Diffeomorphic mapping allows us to relate various sources of information which are inherently a function of spatial position. It is often the underlying technology used for mapping and understanding the data present in the human coordinate system which have been captured by medical imaging in our case a gated cardiac CINE MRI sequence. To obtain the diffeomorphic mapping we use ANTs [4] [5] as the choice of registration algorithm. It being the best in class and also provides considerable speedy results. This allows us to form a warp from our chosen time point(s) to all the time points. We also take advantage of the cyclic nature of the cardiac cycle and obtain the warping from the reference time point to time points in both direction till we encounter the next reference time point. This gives us a rich collection of diffeomorphic mappings between time points. Then we only have to segment the area of interest in the scans of the chosen reference time points. Once we have those segmentation we can use the diffeomorphic mapping to propagate the segmentation from the reference time point to all the other time points.

3. Results and Analysis

In this section we state the results obtained by the methods described in section 2 and also state the analysis of the same. We had an expert clinician segment the left atrium of all time points of all patients which constituted the ground truth segmentation. We selected time points using both (i) Uniform Sampling and (ii) Grouping time points together based on similar properties. We proceeded with a choice of one, three, five and seven reference time points for both the methods.[6, 7] We compute the DICE coefficients and the

per vertex Hausdorff distances between the predicted volumetric segmentation and the ground truth segmentation as shown in figure 1. The more the number of reference segmentation the closer the predicted segmentation are to the ground truth as shown in table 1,2,3 and 4.

The common trend that we can observe from the tables is that more the number of reference points the better the segmentation is. The DICE scores increase as and when we increase the number of reference time point and the per vertex Hausdorff distance decreases with more reference time points. The next trend that we observe is the difference between the method of choosing the reference time points. The average DICE scores and the average Hausdorff distances are higher when we choose reference time points using uniform sampling. This is in synchronous with our assumption of using the inherent properties of the CINE time points as a metric of choosing the reference time points. As the cardiac cycle is not symmetric in nature with respect to the spatial representation we need to leverage the intrinsic properties in choosing the reference time points.

The graphs show the actual distribution of the individual DICE scores and Hausdorff distance for each time points for each patient. The DICE scores at the peaks are the reference time point which shows a perfect match. The same can be observed for the Hausdorff distance as it produces the minimum value at the reference time points. The other figure portrays the distance map overlay on the segmentation, the red area shows the maximum difference between the ground truth segmentation and the predicted segmentation. As we can see the maximum difference can be observed for the region around the border of the left atrium and ventricle. This segmentation region is chosen by the person performing the segmentation and can vary for each time point.

4. Conclusion

In this we paper we propose an efficient pipeline that uses limited ground truth segmentation of CINE time points to accurately propagate it throughout the sequence. This method significantly reduces human effort and resources which could better directed towards other tasks. We analyze this method by choosing the reference time points and the number of them. The common inference is that more the number of time points chosen as reference the better is the predicted segmentation. The second conclusion is based on the method of choosing the reference time points. We can sample the time points, equally distributing them over the entire cardiac cycle or use the inherent characteristics of the volumes. Our results show that grouping the time points by their common characteristics overall performs better than uniform sampling.

This pipeline can considerably reduce human effort and

also save resources and time. This method aims at introducing a certain level of homogeneity in the segmentation and not suffer from inter person differences.

Acknowledgments

This work was supported by NIH [grant numbers R01-AR-076120-01, R01- HL135568-02, and P41-GM103545-19].

References

- [1] Oakes R, Badger T, Kholmovski E, Akoum N, Burgon N, Fish E, Blauer J, Rao S, DiBella E, Segerson N, et al. Detection and quantification of left atrial structural remodeling with delayed-enhancement magnetic resonance imaging in patients with atrial fibrillation. *Circulation* 119 1758;1767.
- [2] Akoum N, Daccarett M, McGann C, Segerson N, Vergara G, Kuppahally S, Badger T, Burgon N, Haslam T, Kholmovski E, et al. Atrial fibrosis helps select the appropriate patient and strategy in catheter ablation of atrial fibrillation: A demri guided approach. *Journal of cardiovascular electrophysiology* 2011;22(1):16–22.
- [3] Klein A, Andersson J, Ardekani BA, Ashburner J, Avants B, Chiang MC, Christensen GE, Collins DL, Gee J, Hellier P, et al. Evaluation of 14 nonlinear deformation algorithms applied to human brain mri registration. *Neuroimage* 2009; 46(3):786–802.
- [4] Avants BB, Tustison N, Song G. Advanced normalization tools (ants). *Insight j* 2009;2:1–35.
- [5] Avants BB, Tustison NJ, Song G, Gee JC. Ants: Open-source tools for normalization and neuroanatomy. *HeanetIe* 2009; 10:1–11.
- [6] Taha AA, Hanbury A. Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC medical imaging* 2015;15(1):29.
- [7] Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Medical Imaging* August 2015;15:29.

Address for correspondence:

Shalin Parikh
72 Central Campus Drive, 3rd floor desk, Salt Lake City, Utah
84112, USA
shalinprkh@sci.utah.edu

Patient	1 reference time point	3 reference time points	5 reference time points	7 reference time points
(i)	0.8645 ±0.091	0.8734 ±0.127	0.8801 ±0.086	0.8982 ±0.102
(ii)	0.7661 ±0.125	0.790 ±0.128	0.8521 ±0.119	0.8528 ±0.136
(iii)	0.7945 ±0.068	0.8124 ±0.128	0.8521 ±0.119	0.8528 ±0.136
(iv)	0.8426 ±0.085	0.8714 ±0.109	0.8977 ±0.098	0.8981 ±0.093

Table 1. DICE Scores for all patients across all time points using Uniform Sampling.

Patient	1 reference time point	3 reference time points	5 reference time points	7 reference time points
(i)	0.8645 ±0.091	0.8753 ±0.086	0.8862 ±0.093	0.8947 ±0.089
(ii)	0.7661 ±0.125	0.806 ±0.135	0.8431 ±0.120	0.8749 ±0.120
(iii)	0.7945 ±0.068	0.8008 ±0.107	0.8149 ±0.120	0.8528 ±0.105
(iv)	0.8426 ±0.085	0.9028 ±0.074	0.9368 ±0.066	0.9438 ±0.130

Table 2. DICE Scores for all patients across all time points by choosing reference points using K medoids.

Patient	1 reference time point	3 reference time points	5 reference time points	7 reference time points
(i)	0.4704 ±0.519	0.4099 ±0.97	0.3079 ±0.3094	0.3597 ±0.364
(ii)	1.0674 ±0.977	0.7990 ±0.978	0.7181 ±1.020	0.8631 ±1.208
(iii)	0.7503 ±0.305	0.7990 ±0.978	0.7181 ±1.020	0.8631 ±1.208
(iv)	0.7308 ±0.734	0.5847 ±0.851	0.4992 ±0.845	0.4694 ±0.673

Table 3. Hausdorff distance for all patients across all time points by choosing reference points using Uniform Sampling.

Patient	1 reference time point	3 reference time points	5 reference time points	7 reference time points
(i)	0.4704 ±0.519	0.3968 ±0.393	0.3802 ±0.469	0.3344 ±0.417
(ii)	1.0674 ±0.977	1.0508 ±1.429	0.6420 ±0.768	0.5888 ±0.868
(iii)	0.7503 ±0.305	0.7122 ±0.508	0.7025 ±0.647	0.4967 ±0.449
(iv)	0.7308 ±0.7341	0.4027 ±0.581	0.2851 ±0.492	0.1952 ±0.696

Table 4. Hausdorff Distance for all patients across all time points by choosing reference points using K medoids.

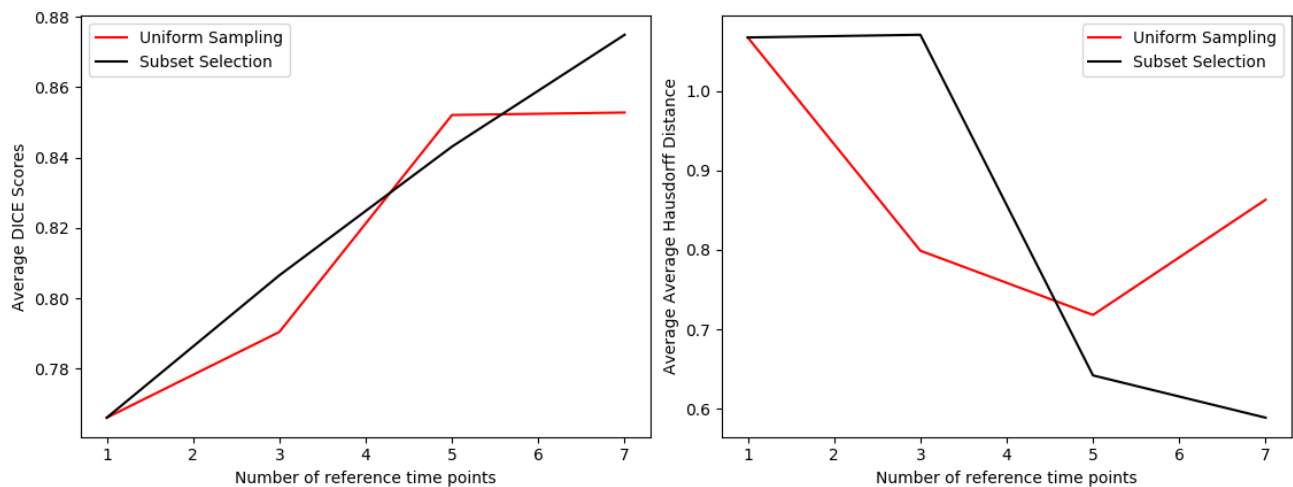


Figure 1. Comparison between DICE scores and Hausdorff Distances across time points.