Multi-atlas Propagation Whole Heart Segmentation from MRI and CTA Using a Local Normalised Correlation Coefficient Criterion

Maria A. Zuluaga, M. Jorge Cardoso, Marc Modat, and Sébastien Ourselin

Centre for Medical Image Computing,
University College of London, UK

Abstract. Accurate segmentation of the whole heart from 3D image sequences is an important step in the developement of clinical applications. As manual delineation is a tedious task that is prone to errors and dependant on the expertise of the observer, fully automated segmentation methods are highly desirable. In this work, we present a fully automated method for the segmentation of the whole heart and the great vessels from 3D images. The method is based on a muti-atlas propagation segmentation scheme, that has been proven to be succesful in brain segmentation. Based on a cross correlation metric, our method selects the best atlases for propagation allowing the refinement of the segmentation at each iteration of the propagation. We show that our method allows segmentation from multiple image modalities by validating it on computed tomography angiography (CTA) and magnetic resonance images (MRI). Our results are comparable to state-of-the-art methods on CTA and MRI with average Dice scores of 90.9% and 89.0% for the whole heart when evaluated on a 23 and 8 cases, respectively.

1 Introduction

Recent advances in imaging techniques have positioned 3D CTA and MRI as potential alternatives of conventional 2D approaches (2D x-ray angiography or cine MR) for cardiac examination. The 3D imaging of the heart opens a wide set of applications that range from accurate ventricular measurements, computational simulation to image guided interventions. For such applications, it is necessary to have an accurate segmentation of the whole heart which includes the four chambers and eventually also the great vessels. Manual delineation of the cardiac anatomical structures is a labour intensive and tedious task, which is prone to inter- and intraobserver variations. Therefore, it is highly desirable to develop techniques for automatic segmentation.

Although most of the heart segmentation algorithms found in the literature have been focused in the left ventricle [1], recently new methods that tackle whole heart segmentation in 3D data have been developed. These segmentation methods have used either model-based approaches [2, 3] or atlas propagation techniques [4–7]. Among these, atlas propagation techniques are well-known and

proven succesful for brain segmentation and, recently have gained popularity for heart segmentation.

In this work, we present a multi-atlas propagation segmentation approach for segmenting the whole heart. Our approach differentiates from previous atlas propagation methods in three aspects: 1) By using multiple atlases instead of a single one [6], 2) by providing a segmentation capable of labelling the differen structures of the heart, instead of providing a single binary mask of the whole heart [4, 5], and 3) by using a multi-label fusion strategy based on the STAPLE algorithm [8] that incorporates a locally normalised cross correlation (LNCC) based ranking combined with a consensus based region-of-interest selection and an iterative Markov Random Field into the multi-label STAPLE algorithm. Previously, Kirişli *et al.* have used a majority voting scheme that does not allow to evaluate the local similarity among the images.

In the remaining sections of the paper, we first describe the proposed multi atlas propagation segmentation approach, followed by the experimental setup and the obtained results. To prove the robustness of our method to different image modalities, quantitative evaluation is given on MRI and CTA data. Finally, we discuss on the results and present conclusions.

2   Methodology

2.1   Multi-atlas Propagation Segmentation Framework

Let Y_u be an unseen image to be segmented and $A_i = (Y_i; Li)$, $i = \{1, ..., n\}$, be the set of n atlases composed of an intensity image Y_i and a label image L_i. The final segmentation is obtained by transforming the set of n atlases into the image space of Y_u and then applying a fusion criterion to combine the label images L_i from each atlas into a consensus segmentation L_u.

Registration of cardiac images is challenging due to the structures surrounding the heart (e.g. ribs, liver) that tend to bias the registration. To avoid such problem, we divide the registration in two-stages. At the first stage, we seek to define a region of interest (ROI) that encloses the heart in the Y_u and remove the surrounding structures. The unseen image Y_u is affinely registered [9] to the intensity images Y_i of the atlases (Fig. 1), step 1). The obtained transformations are applied to the binarized L_i images, which are then fused using a simple majority voting criterion. As the different labels from L_i are not considered, a binary mask M_u roughly englobing the heart is obtained as an output (Fig. 1), step 2-3).

After the affine alignment step, a nonrigid free form deformation registration [10, 11] using normalised mutual information is applied to align the atlases with the the unseen image (Fig. 1) step 4). To avoid the bias that surrounding structures can produce in the registration, the unseen image is masked using the binary image M_u. After non-rigidly transforming the atlases to the unseen image space, the final segmentation is obtained through combination of the label images. In the following section, we further describe the fusion criterion (Fig. 1) step 5).
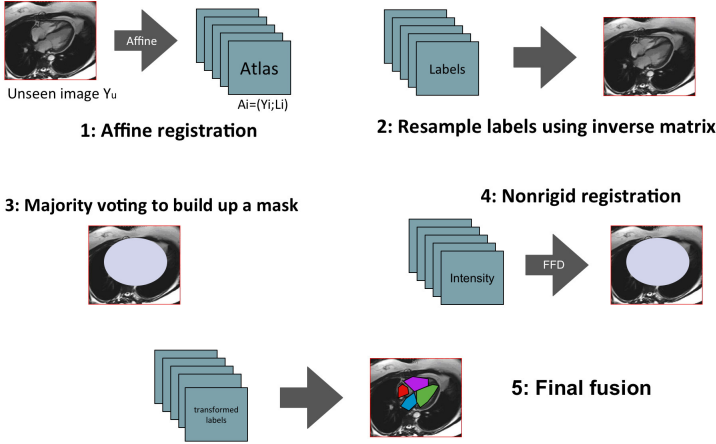
Fig. 1. In the proposed method, the atlases A_i are first roughly aligned with the unseen image, Y_u, through an affine registration (step 1). The obtained transformations are used to map the label images into the unseen image space (step 2). The binarized label images are fused using a majority voting scheme to obtain a mask of the region of interest (step 3). Using the obtained mask as an additional input, the atlases are nonrigidly aligned to the unseen image and subsequently transformed (step 4). Finally, the transformed labels are fused using a cross-correlation metric.

2.2   Locally Normalised Cross Correlation (LNCC)

The final segmentation L_u is obtained by using the multi-STEPS algorithm [12] in combination with a LNCC-based ranking strategy in order to determine which are the most suitable registered atlases to use in the fusion process.

The goal of the LNCC is to rank the registered atlases according to the local quality of the match between the unseen image Y_u and the warped atlases. The metric requires the computation of the mean and standard deviation of a pair of images on a local Gaussian window using a convolution method. Denoting for the sake of simplicity the unseen image as Y and an atlas intensity image as A, the LNCC at a voxel position p is defined as:

$$LNCC_p = \frac{\langle Y, A \rangle_p}{\sigma_p(Y)\sigma_p(A)} \tag{1}$$

where

$$\langle Y, A \rangle_p = \overline{Y \cdot A}_p - \overline{Y}_p \cdot \overline{A}_p$$
$$\overline{Y}_p = G_{\sigma_G} * Y$$
$$\sigma_p(Y) = \overline{Y^2}_p - \overline{Y}_p^2$$

with $*$ denoting the convolution operator and G_{σ_G} a Gaussian kernel with standard deviation σ_G.

Due to the local nature of the metric, the similarity between the unseen image and the atlases can be described on a voxel by voxel basis, enabling a voxel by voxel ranking. The multi-STEPS algorithm provides a variable X that allows to control the number of atlases to use locally according to the LNCC. If the atlas A_i is among the top X ranked at a particular voxel position, then it will be considered for the fusion process at that specific voxel.

3   Experiments and Results

Our method was evaluated on 22 3D SSFP MRI scans acquired at King's College London [13] and 8 3D CTA cases acquired in the Erasmus MC (Rotterdam, The Netherlands) [7]. MRI images had a size of $256 \times 256 \times 140$ and the CTA mean size is $512 \times 512 \times 274$. For each scan, a manual segmentation of the four main chambers, the myocardium and the aorta. Additionally, the CTA cases contained manual annotations of the pericardium and the MRI scans contained annotations of the pumonary artery. We chose not to consider these two for the evaluation, as they were not present in both datasets.

A leave-one-out strategy was applied, where one scan is segmented using the remaining cases as atlas database. It should be noted that MRI and CTA scans were evaluated separately. The quality of the automatic segmentations was assessed by comparing them to the manual segmentations through the Dice score, $Dice = \frac{2|V_{seg} \cap V_{man}|}{|V_{seg} + V_{man}|}$, where V_{seg} and V_{man} denote the automated segmentation and the manual annotation, respectively. Table 1 presents the obtained results for each cardiac structure, whereas Fig. 2 illustrates some of the obtained results in MRI and CTA.

Additionally, we evaluated the method performance as a function of the atlas database size (Fig. 3 left), n and the number of top atlases X considered in the final fusion (Fig. 3 right). For the first experiment, we varied n from 3 to the total available atlases ($n = 22$ in MRI and $n = 7$ in CTA), while considering $X = n/3$ for the final fusion. The atlases were ranked using a global metric [14] to define the inclusion order To evaluate the effect of X in the segmentation results, we varied it in the range $X = \{2, \ldots, n/2\}$ while using the whole set of available atlases for each modality. The quality of the segmentation was assessed through the whole heart Dice score.

Table 1. Mean ($\pm$ std) Dice score for the evaluated datasets.

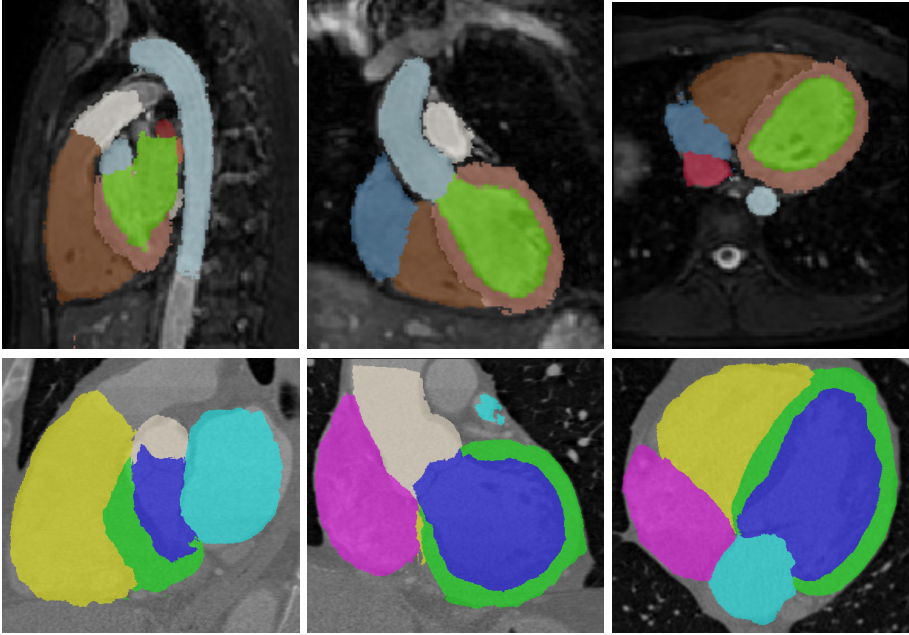|	MR	CTA
Myocardium	0.87 ± 0.06	0.86 ± 0.07
Left ventricle	0.95 ± 0.02	0.93 ± 0.05
Right ventricle.	0.92 ± 0.03	0.89 ± 0.03
Left atrium	0.92 ± 0.03	0.90 ± 0.04
Right atrium	0.89 ± 0.05	0.84 ± 0.05
Aorta	0.86 ± 0.09	0.87 ± 0.06

Fig. 2. Coronal, sagital and axial views of two segmented volumes. On top an MRI scan (whole heart Dice score 0.903) and, on the bottom a CT scan (whole heart Dice score 0.894).
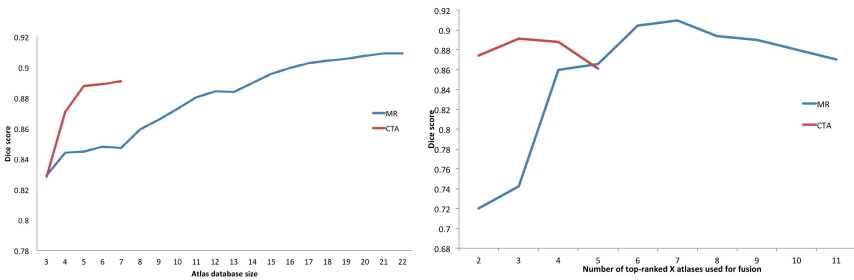

Fig. 3. *Left.* Evolution of whole heart Dice score as a function of the atlas database size in MRI and CTA scans. Best results are achieved with $n = 22$ in MRI ($Dice = 0.909$) and in $n = 7$ CTA ($Dice = 0.890$).Note that the curve for CTA is shorter as less atlases are available. *Right.* Algorithm performance as a function of the number X of top ranked atlases considered for the fusion in MRI scans and CTA. A database of $n = 22$ atlases was considered for MRI and $n = 6$ for CTA. Best results are achieved for $X = 7$ in MRI ($Dice = 0.909$) and $X = 3$ in CTA ($Dice = 0.890$).

4   Discussion

The results presented in Table 1 show that our method has a good performance in the segmentation of the different cardiac structures under different image modalities. Average whole heart segmentation on MRI was 0.909 ± 0.03 and 0.890 ± 0.04 on CTA images. The latter results are comparable to those reported by Zhuang *et al.* [6] using a single atlas on MRI images (Dice score of 0.84 ± 0.05) and by Kirişli *et al* [7] using a majority voting scheme to fuse 7 atlases on CTA images (Dice score of 0.93).

   We consider that the overall better performance of our method on MRI images than on CTA can be explained mainly by the fact that our MRI atlas database is larger than that one of CTA. From Fig. 3, it can be seen that the Dice score of the MRI segmentations has a rapid increase but, afterwards it tends to stabilise (for $n >= 17$ the Dice score varies from 0.902 to 0.909). If such behavior is extrapolated to the CTA curve, we should expect to have a better performance on CTA with a larger database. Although the LNCC-based ranking implies using only a percentage of the n for the fusion, having a larger atlas database increases the probability of having better matches between the atlases and the unseen image. Our experiments on MRI suggest that using $X = n/3$ top atlases provides the best results (Fig. 3). When this rule was applied to the CTA database, only $X = 2$ atlases should be used, which leads to consensus problems. We consider that the limited number of CTA atlases does not allow to exploit the advantages of our atlas ranking scheme. An alternative to increase the atlas database could be to use all the available atlases images, regardless of their modality, in the fusion process. However, after initial experimentations we found out that the inter-modality registration did not provide satisfactory results. The different imaging protocols result in images with different field of views (Fig. 4) that introduce large registration errors. A more robust registration strategy that can cope with such inter modality differences is part of the future work.

   Despite our algorithm performing better on MRI than on CTA, the Dice score obtained for the aorta in CTA is higher than on MR. This can be explained by the quality of the aorta labels in the atlases. The annotations for the aorta on the CTA are more homogenious and follow an annotation protocol [7]. On the opposite, the MRI aortic annotations do not follow a particular protocol which results on different extensions of the ascending aorta being labeled. This behavior points out one of the disadvantages of multi-atlas segmentation approaches where the quality of the final segmentation strongly depends on the quality of the label images.

   The runtime of our method depends more on the size of the images that are to be segmented than on the number of atlases that are used. Despite having a larger number of atlases, the MRI database (22 compared to 7 from CTA), it takes approximately 30 minutes to segment one patient on a PC with a quad-core processor (2.13GHz), whereas the segmentation of a CTA scan takes approximately one hour.
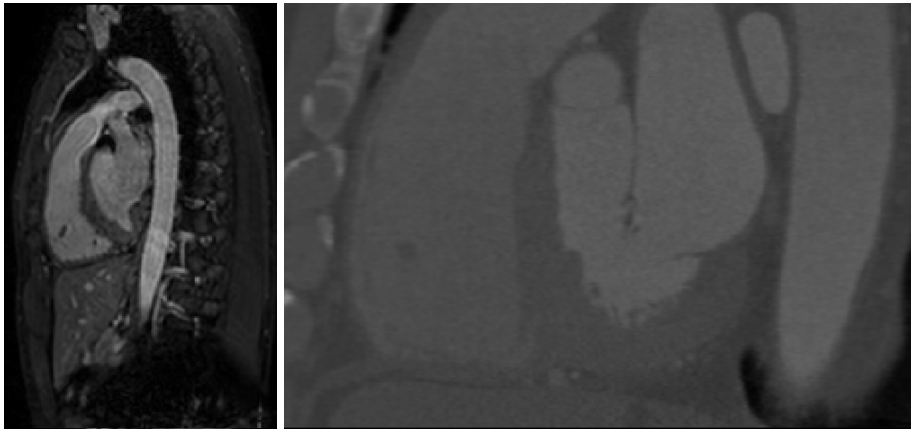
Fig. 4. Sagittal view of the heart at approximately the same position in an MRI (left) and CTA (right) dataset. The field of views of each case are completely different and thus the information contained.

5 Conclusions

In this paper we proposed the use of a multi-atlas propagation segmentation framework for whole heart segmentation in multi-modal 3D images. Our approach performs a two-stage registration step that aims to first localize the heart and globally align the images in order to use such alignment as a robust initialisation of a nonrigid transformation, where surrounding structures are masked out. A key feature of our approach is that it makes use of a multi-label ranking criterion, based on the local normalised cross correlation, to select the best atlases for label fusion. This is the first time that the multi-STEPS is applied to whole heart segmentation and has proven its robustenss by being successfully applied to different image modalities.

References

1. Petitjean, C., Dacher, J.-N.: A review of segmentation methods in short axis cardiac MR images. Med. Image Anal. 15(2), 169–184 (2011)
2. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Four-Chamber Heart Modeling and Automatic Segmentation for 3-D Cardiac CT Volumes Using Marginal Space Learning and Steerable Features. IEEE Transactions on Medical Imaging 27(11), 1668–1681 (2008)

3. Peters, J., Ecabert, O., Meyer, C., Kneser, R., Weese, J.: Optimizing boundary detection via Simulated Search with applications to multi-modal heart segmentation. Medical Image Analysis 14(1), 70–84 (2010)
4. Išgum, I., Staring, M., Rutten, A., Prokop, M., Viergever, M., van Ginneken, B.: Multi-Atlas-Based Segmentation With Local Decision Fusion-Application to Cardiac and Aortic Segmentation in CT Scans. IEEE Trans. Med. Imag. 28(7), 1000–1010 (2009)
5. Rikxoort, E.M., Išgum, I., Arzhaeva, Y., Staring, M., Klein, S., Viergever, M., Pluim, J.P.W., van Ginneken, B.: Adaptive local multi-atlas segmentation: Application to the heart and the caudate nucleus. Medical Image Analysis (2010)
6. Zhuang, X., Rhode, K.S., Razavi, R.S., Hawkes, D.J., Ourselin, S.: A registration-based propagation framework for automatic whole heart segmentation of cardiac MRI. IEEE Trans. Med. Imag. 29(9), 1612–1625 (2010)
7. Kirişli, H.A., Schaap, M., Klein, S., Papadopoulou, S.L., Bonardi, M., Chen, C.H., Weustink, A.C., Mollet, N.R.A., Vonken, E.P.A., van der Geest, R.J., van Walsum, T., Niessen, W.J.: Evaluation of a multi-atlas based method for segmentation of cardiac CTA data: a large-scale, multicenter, and multivendor study. Medical Physics 37(12), 6279–6292 (2010)
8. Warfield, S.K., Zou, K.H., Wells, W.M.: Simultaneous Truth and Performance Level Estimation (STAPLE): An Algorithm for the Validation of Image Segmentation. IEEE Trans. Med. Imag. 23(70), 903–921 (2004)
9. Ourselin, S., Roche, S., Subsol, G., Pennec, X., Ayache, N.: Reconstructing a 3D structure from serial histological sections. Image and Vis. Comp. 19(1-2), 25–31 (2001)
10. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L.G., Leach, M.O., Hawkes, D.J.: Nonrigid registration using freeform deformations: Application to breast MR images. IEEE Trans. Med. Imag. 18(8), 712–721 (1999)
11. Modat, M., Ridgway, G.R., Taylor, Z.A., Lehmann, M., Barnes, J., Fox, N.C., Hawkes, D.J., Ourselin, S.: Fast free-form deformation using graphics processing units. Comput. Methods Programs Biomed. 98(3), 278–284 (2010)
12. Cardoso, M.J., Modat, M., Keihaninejad, S., Cash, D., Ourselin, S.: Multi-STEPS: multi-label similarity and truth estimation for propagated segmentations. In: IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA), pp. 153–158 (2012)
13. Uribe, S., Tangchaoren, T., Parish, I., Wolf, R., Razavi, R., Greil, G., Schaeffter, T.: Volumetric cardiac quantification by using 3-D dual phase whole-heart MR imaging. Radiology 248, 606–614 (2008)
14. Leung, K.K., Barnes, J., Ridgway, G.R., Bartlett, J.W., Clarkson, M.J., Macdonald, K., Schu, N., Fox, N.C., Ourselin, S.: ADNI: Automated cross-sectional and longitudinal hippocampal volume measurement in mild cognitive impairment and Alzheimer's disease. NeuroImage 51(4), 1345–1359 (2010)