Strengths and Pitfalls of Whole-heart Atlas-based Segmentation in Congenital Heart Disease Patients

Maria A. Zuluaga[1], Benedetta Biffi[2,3], Andrew M. Taylor[2], Silvia Schievano[2], Tom Vercauteren[1], and Sébastien Ourselin[1]

[1] Translational Imaging Group, Centre for Medical Image Computing, University College London, UK
[2] Department of Medical Physics and Biomedical Engineering, University College London, UK
[3] Centre for Cardiovascular Imaging, UCL Institute of Cardiovascular Science & Great Ormond Street Hospital for Children, UK

Abstract. Atlas-based whole-heart segmentation is a well-established technique for the extraction of key cardiac structures of the adult heart. Despite its relative success in this domain, its implementation in whole-heart segmentation of paediatric patients suffering from a form of congenital heart disease is not straightforward. The aim of this work is to evaluate the current strengths and limitations of whole-heart atlas based segmentation techniques within the context of the Whole-Heart and Great Vessel Segmentation from 3D Cardiovascular MRI in Congenital Heart Disease Challenge (HVSMR). Obtained results suggest that there are no significant differences in the accuracies of state-of-the-art methods, reporting maximum Dice scores of 0.73 for the myocardium and 0.90 for the blood pool.

1 Introduction

Congenital heart disease (CHD) has a reported incidence that varies between 4 - 10 per 1000 births [8]. In the last decades, the survival of CHD patients has increased thanks to improvements in diagnosis, medical treatment and surgical repair [13]. This has led to new challenges for follow-up, reinvestigation and reoperation [13].

Imaging plays a key role in the diagnosis of CHD and it is required at all stages of patient care [9]. It outlines anatomy and physiology, assists the intervention planning, helps to refine management, evaluates the consequences of interventions and facilitates prognosis [9]. At all stages of the CHD assessment pipeline, delineation and extraction of the whole heart, which includes the four chambers and eventually the great vessels, are crucial to achieve a successful outcome. To date, the process requires a significant amount of manual labour involving many hours of work [5] and it is prone to inter- and intra-observer variations.

Cardiovascular magnetic resonance (CMR) has emerged as a major alternative to echocardiography (the by default modality of choice in paediatric patients) as an imaging tool within the CHD assessment pipeline [9, 11], due to its non-invasiveness, lack of ionizing radiation, and its higher anatomical resolution and extra cardiac information. Moreover, the vast literature in CMR analysis tools [14] proposes a large set of semi- and fully-automated methods for adult whole-heart segmentation that could be adapted and applied to paediatric CMR images to ease and improve the delineation and extraction of the heart.

Among fully-automated segmentation methods, atlas-based techniques are largely popular due to their robustness and high reported accuracies [6, 15, 17]. However, when applied in the context of CHD patients, they are not as successful. This might be explained by the substantial changes in heart topology and high anatomical variability in CHD that lead to poor segmentations [2, 10, 16]. In an ideal scenario, this could be solved by using as many atlas databases as pathological conditions. This is, however, difficult to achieve in practice. If atlas-based techniques were to be used for CHD assessment, there is a need to develop algorithms which are robust enough to topological and anatomical variations without the requirement of pathology-specific databases.

In the context of the Whole-Heart and Great Vessel Segmentation from 3D Cardiovascular MRI in Congenital Heart Disease Challenge (HVSMR) [10], this work focuses on atlas-based segmentation techniques and aims to assess the strengths and limitations of these methods in CHD patients. In the remaining of this paper, Section 2 provides a more detailed overview of atlas-based approaches, Section 3 describes the materials and methods used and Section 4 presents the obtained results. Finally, section 5 discusses the results and presents the conclusions.

2 Background: Atlas-based Segmentation

An atlas, in its simplest form, is made up by an intensity image and its corresponding annotated image. The final goal of atlas-based segmentation is to use the relationship between the labels of the annotated image and the intensities of the corresponding CMR image to assign labels to the voxels of an unseen (or target) image. Typically, this is achieved by registering the atlas into the unseen image space and then applying a technique to convert the labelled images into a final segmentation.

In whole-heart segmentation, atlas-based methods generally follow a two-stage registration approach to transform the atlas into the target image space. At the first stage, the atlas is registered using an affine transformation to the unseen image to achieve a rough alignment. After the affine alignment step, a non-rigid deformation registration is applied to align the atlas with the unseen image. The resulting transformation is finally used to resample the atlas' labels into the unseen image. Differences in methods making use of this two-stage framework can be found in: 1) the atlas database that is used, 2) the labels and 3) the algorithm used to convert the transformed atlas labels into the un-

seen image segmentation. In the following, a brief description of each is provided.

Atlas database. In its simplest form, an atlas consists of a single pair of images. This *simple atlas* [15] can be composed of a CMR image and its associated label image or, more often, by a mean intensity image, which is obtained from the registration and averaging of several CMR images, and a label image. The associated label image can be formed by binary labels [15] or by a probabilistic image [7], reflecting the chance of a voxel to belong to a specific label/class. Alternatively, if a database of intensity images is available, it is possible to have multiple annotated images and propagate them to the unseen image [4, 6, 17]. This approach is referred to as multi-atlas segmentation.

Labels. The most common approach in cardiac segmentation seeks to not only identify the heart, but also its different structures [7, 6, 15, 17]. Only a few works have focused on the heart as a unit by providing a single binary mask of the whole heart [4]. When a simple atlas is used, the use of multiple labels is straightforward. In multi-atlas segmentation, there is the need for specific methods (fusion algorithms) which can handle potential overlapping of different labels once the images are transformed into the target image space.

Fusion algorithm. Methods using a simple atlas obtain the final segmentation by directly transforming the labelled image into the unseen image space [15]. When probabilistic images are used, it is common to have a post-processing stage to refine the propagations [7]. When multiple atlases and/or labels are used, it is mandatory to implement a merging technique which can combine the labels from multiple images into a final consensus segmentation. This is commonly denoted a fusion algorithm. Common combination rules applied to cardiac segmentation include majority voting [6] and weighted decision functions [4, 17].

3 Materials and Methods

In this section, we describe the images and the experimental setup used to evaluate the strengths and limitations of atlas-based methods (Section 2) when aiming at highly accurate whole-heart segmentations of CHD patients.

3.1 Materials

CMR images provided by the challenge organisers[4] were acquired during clinical practice at Boston Children's Hospital (Boston, USA). Cases include a variety of congenital heart defects. Some subjects have undergone interventions. Imaging was done in on a 1.5T scanner (Phillips Achieva) without contrast agent using a steady-state free precession (SSFP) pulse sequence. Image dimension and image spacing varied across subjects, and average $390 \times 390 \times 165$ and $0.9 \times 0.9 \times 0.85$ mm, respectively, in the full-volume training dataset.

[4] http://segchd.csail.mit.edu/index.html

Manual segmentation of the blood pool and ventricular myocardium was performed by a trained rater, and validated by two clinical experts. The annotated data is composed of two labels: blood pools and myocardium. The blood pool class includes the left and right atria, left and right ventricles, aorta, pulmonary veins, pulmonary arteries, and the superior and inferior vena cava. The myocardium class includes the thick muscle surrounding the two ventricles and the septum between them. Coronaries are not included in the blood pool class, and are labelled as myocardium if they travel within the ventricular myocardium.

3.2 Methods

As described in Section 2, state-of-the-art whole-heart segmentation frameworks mainly differ in the type of atlas, the number of labels and the type of fusion scheme. In this section, we describe the different strategies that were considered for the evaluation of these elements.

Although there exist several tools and schemes to implement the two-stage registration pipeline typically used in whole-heart segmentation we have selected the registration scheme described in [16] for its performance and because sit contains the necessary information to be reproduced. The effect of the registration algorithm in whole-heart atlas-based segmentation is out of the scope of this work.

Simple atlas vs. Multi-atlas. In order to evaluate the performance of a simple atlas w.r.t a multi-atlas approach, mean intensity atlases were constructed using the CMR images of the training data through co-registration to a reference image within the training set. Binary labels, rather than probabilistic ones, were obtained via majority voting and manual correction of the results. For multi-atlas evaluation, no additional processing of the data was required: the training data provided was used as the atlas.

Single vs. Multi-label. An additional atlas using a single label was constructed by combining the two original labels, blood pool and myocardium, into a single one, denoted whole-heart. When evaluating these two schemes, it was the accuracy in segmenting the whole heart rather than each individual class what was assessed. For this purpose, in the multi-label case all the labels were used during the propagation and fusion process, but the resulting labels in the final segmentation were considered as one when comparing to the single label approach.

Fusion schemes. We evaluated three different fusion schemes for the merging of the labels when using multiple labels or atlases. These are: majority voting (MV), STEPS [1] and the STAPLE algorithm [12]. Both MV and STEPS have been successfully applied in whole-heart segmentation [6, 17]. STAPLE has not been applied in this domain but, given its large success in brain image segmentation, we included it in the fusion schemes to consider.

Post-processing. The label fusion result does not necessarily represent the final segmentation; sometimes it is fed to another algorithm to estimate the output labels [3]. In whole-heart segmentation, this kind of post-processing techniques have not been that popular and are rarely used [7]. However, we have included this step within the evaluated framework as it could be used to improve the

segmentation obtained from the fusion, which can be prone to errors due to the existing pathologies. In this work we evaluated the use of the label fusion result as initialisation for parameter estimation of a Gaussian mixture model using expectation maximisation (EM) algorithm.

4 Experimental Results

Evaluation of the different settings involved in the implementation of an atlas-based segmentation pipeline was performed using the Dice score coefficient (DSC), as suggested by the challenge organisers.

4.1 Training Data

Table 1 summarises the results obtained when using a simple atlas. Results are reported using multiple labels and a single one. In the latter, when no post-processing is applied, only the whole heart DSC is reported. Table 2 presents the results for the multi-atlas based segmentation using different labels, fusion schemes and post-processing. A comparison of the results from both tables shows that multi-atlas segmentation is superior to that one using a single atlas.

4.2 Testing Data

The best performing pipeline, as reported in Table 2 was submitted to the HVSMR Challenge. For this matter, 10 additional cases were segmented. The full training data set was used as atlas. Average DSC of 0.73 and 0.90 were obtained for the myocardium and blood pool, respectively. These values are very similar to those obtained when performing cross-validation over the training set. The full set of results are reported in the HSVMR website[5].

Table 1. Mean DSC ($\pm$ std. deviation) using a simple atlas to segment the whole heart (single label) and the multiple labels. Results are also reported with no post-processing and post-processing through EM refinement.

	Whole-heart	Myocardium	Blood pool
No post-processing	0.61 ± 0.08	0.43 ± 0.08	0.54 ± 0.08
EM refinement	-	0.65 ± 0.08	0.75 ± 0.08

5 Discussion and Conclusions

In this work we have evaluated the use of standard atlas-based segmentation techniques for the extraction of the blood pool and the myocardium in CHD patients. The results demonstrate that, as it had been suggested in previous works,

[5] https://challenge.kitware.com/#submission/57dc2c02cad3a51cc66c8b12

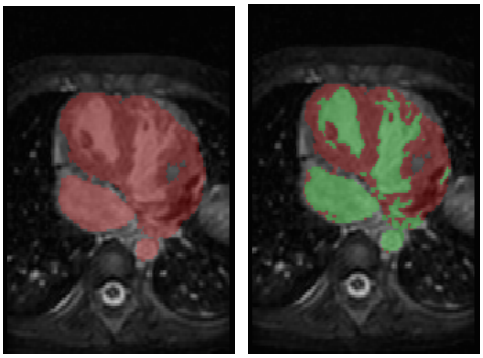

Fig. 1. Poor segmentation (case 4) using a multi-atlas, single label and STEPS fusion approach with post-processing (right) and without post-processing (left). A large section of the myocardium is missed originally and this cannot be recovered by the post-processing.

atlas-based methods show lower accuracies in the presence of high anatomical variations that can not be captured by the registration algorithm. Nevertheless, the results demonstrate the potential of atlas-based methods within the CHD assessment pipeline.

Overall the presented results suggest to prefer the use of multiple atlases over a simple atlas. However, when in a multi-atlas based scheme the results have suggested that there is not a specific pipeline configuration that performs significantly better than the others. In the same way that no differences were encountered when evaluating different multi-atlas schemes, the authors believe that no significant differences should arise when using different registration algorithm implementations (under the assumption it has been properly tuned). However, it should be noted that registration is out of the scope of this work. Based on these findings, we believe there is room for improvement by exploiting the way the information of the atlases is structured (*e.g* single vs multiple atlas or labels). To date, most of the efforts have been directed towards the development of novel registration and fusion algorithms rather than towards the development of novel atlas generation strategies, e.g. data augmentation to simulate CHD atlases, with enriched information and features.

The obtained results also suggest that the use of a post-processing scheme can improve the results of the atlas-based segmentation. For instance, the best performing pipeline was obtained through the use of the EM algorithm. However, it should be noted that when the initial segmentation is not good enough (Fig. 1), there is not much that can be improved through the post-processing. Finally, if a single technique had to be recommended to be used currently in the CHD pipeline assessment, then we would recommend to prefer multiple atlases and the use of a single label to locate the heart, in combination with the EM algorithm.

Table 2. Mean DSC (± std. deviation) using a multi-atlas approach. Results are reported for all the possible configurations. (SL) denotes the use of a single label for fusion and the results from the best performing pipeline are highlighted in bold.

Fusion	Segmentation	No post-processing	EM refinement
Majority voting	Whole heart (SL)	0.89 ± 0.05	-
	Whole heart	0.88 ± 0.04	-
	Myocardium (SL)	-	0.74 ± 0.09
	Myocardium	0.72 ± 0.13	0.74 ± 0.09
	Blood pool (SL)	-	0.90 ± 0.05
	Blood pool	0.88 ± 0.04	0.90 ± 0.05
STEPS	Whole heart (SL)	**0.90 ± 0.03**	-
	Whole heart	0.90 ± 0.04	-
	Myocardium (SL)	-	**0.74 ± 0.09**
	Myocardium	0.73 ± 0.07	0.73 ± 0.08
	Blood pool (SL)	-	**0.90 ± 0.03**
	Blood pool	0.89 ± 0.03	**0.90 ± 0.03**
STAPLE	Whole heart (SL)	0.87 ± 0.09	-
	Whole heart	0.85 ± 0.09	-
	Myocardium (SL)	-	0.68 ± 0.14
	Myocardium	0.70 ± 0.09	
	Blood pool (SL)	-	0.87 ± 0.08
	Blood pool	0.86 ± 0.09	0.88 ± 0.06

Acknowledgments. This work was supported through an Innovative Engineering for Health award by the Wellcome Trust [WT101957]; Engineering and Physical Sciences Research Council (EPSRC) [NS/A000027/1]. BB is funded by UCL EPSRC Centre for Doctoral Training in Medical Imaging Scholarship Award. AMT and SS receive funding from Heart Research UK, the British Heart Foundation and the National Institute for Health Research Biomedical Research Centre at GOSH and UCL. SO receives funding from the National Institute for Health Research University College London Hospitals Biomedical Research Centre (NIHR BRC UCLH/UCL High Impact Initiative BW.mn.BRC10269) and the EPSRC (EP/K005278/1).

References

1. Cardoso, M.J., Leung, K., Modat, M., Keihaninejad, S., Cash, D., Barnes, J., Fox, N.C., Ourselin, S.: Steps: Similarity and truth estimation for propagated segmentations and its application to hippocampal segmentation and brain parcelation. Medical Image Analysis 17(6), 671 – 684 (2013)
2. Gilbert, K., Cowan, B., Suinesiaputra, A., Occleshaw, C., Young, A.: Rapid D-Affine Biventricular Cardiac Function with Polar Prediction. In: Medical Image Computing and Computer Assisted Interventions (MICCAI). LNCS, vol. 1702, pp. 546–553 (2014)
3. Iglesias, J.E., Sabuncu, M.R.: Multi-atlas segmentation of biomedical images: A survey. Medical Image Analysis 24(1), 205–219 (2015)

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 Transactions on Medical Imaging 28(7), 100–1010 (2009)
5. Jacobs, S., Grunert, R., Mohr, F.W., Falk, V.: 3d-imaging of cardiac structures using 3d heart models for planning in heart surgery: a preliminary study. Interactive Cardiovascular and Thoracic Surgery (2008)
6. Kirisli, H.A., Schaap, M., Klein, S., Papadopoulou, S., Bonardi, M., Chen, C., Weustink, A., Mollet, N., Vonken, E.P.A., van der Geest, R., van Walsum, T., Niessen, W.: 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)
7. Lorenzo-Valdes, M., Sanchez-Ortiz, G., Elkington, A., Mohiaddin, R., Rueckert, D.: Segmentation of 4D cardiac MR images using a probabilistic atlas and the EM algorithm. Medical Image Analysis 8, 255–265 (2004)
8. Marelli, A.J., Mackie, A.S., Ionescu-Ittu, R., Rahme, E., Pilote, L.: Congenital heart disease in the general population changing prevalence and age distribution. Circulation 115, 163–172 (2007)
9. Ntsinjana, H.N., Hughes, M.L., Taylor, A.M.: The role of cardiovascular magnetic resonance in pediatric congenital heart disease. Journal of Cardiovascular Magnetic Resonance 13(51), – (2011)
10. Pace, D.F., Dalca, A.V., Geva, T., Powell, A.J., Moghari, M.H., Golland, P.: Interactive whole-heart segmentation in congenital heart disease. In: Medical Image Computing and Computer Assisted Interventions (MICCAI). LNCS, vol. 9351, pp. 80–88 (2015)
11. Prakash, A., Powell, A.J., Geva, T.: Multimodality noninvasive imaging for assessment of congenital heart disease. Circulation 3, 112–125 (2010)
12. Warfield, S., Zou, K., Wells, W.: Simultaneous truth and performance level estimation (staple): an algorithm for the validation of image segmentation. IEEE Transactions on Medical Imaging 23(7), 903–921 (2004)
13. Wren, C., O'Sullivan, J.: Survival with congenital heart disease and need for follow up in adult life. Heart 85, 438–443 (2001)
14. Zhuang, X.: Challenges and methodologies of fully automatic whole heart segmentation: A review. Journal of Healthcare Engineering 4, 371–407 (2013)
15. Zhuang, X., Rhode, K., Razavi, R., Hawkes, D., Ourselin, S.: A Registration-Based Propagation Framework for Automatic Whole Heart Segmentation of Cardiac MRI. IEEE Transactions on Medical Imaging 29(9), 1612–1625 (2010)
16. Zuluaga, M.A., Burgos, N., Mendelson, A., Taylor, A., Ourselin, S.: Voxelwise atlas rating for computer assisted diagnosis: Application to congenital heart diseases of the great arteries. Medical Image Analysis 26(1), 185–194 (2015)
17. Zuluaga, M.A., Cardoso, M., Modat, M., Ourselin, S.: Multi-Atlas Propagation Whole Heart Segmentation from MRI and CTA Using a Local Normalised Correlation Coefficient Criterion. In: Functional Imaging and Modeling of the Heart (FIMH). LNCS, vol. 7945, pp. 174–181 (2013)