

Whole Heart Segmentation of Cardiac MRI Using Multiple Path Propagation Strategy

X. Zhuang^{1,*}, K. Leung^{1,2}, K. Rhode³, R. Razavi³, D. Hawkes¹,
and S. Ourselin^{1,2}

¹ CMIC, Medical Physics Department, University College London
x.zhuang@ucl.ac.uk

<http://www.cs.ucl.ac.uk/staff/x.zhuang/>

² Dementia Research Centre, Institute of Neurology, University College London

³ Division of Imaging Science, King's College London

Abstract. Automatic segmentation of cardiac MRI is an important but challenging task in clinical study of cardiac morphology. Recently, fusing segmentations from multiple classifiers has been shown to achieve more accurate results than a single classifier. In this work, we propose a new strategy, MULTiple Path Propagation and Segmentation (MUPPS), in contrast with the currently widely used multi-atlas propagation and segmentation (MAPS) scheme. We showed that MUPPS outperformed the standard MAPS in the experiment using twenty-one *in vivo* cardiac MR images. Furthermore, we studied and compared different path selection strategies for the MUPPS, to pursue an efficient implementation of the segmentation framework. We showed that the path ranking scheme using the image similarity after an affine registration converged faster and only needed eleven classifiers from the atlas repository. The fusion of eleven propagation results using the proposed path ranking scheme achieved a mean Dice score of 0.911 in the whole heart segmentation and the highest gain of accuracy was obtained from myocardium segmentation.

1 Introduction

Magnetic resonance imaging (MRI) has become a routine modality for the determination of patient cardiac morphology. The extraction of this morphological information can be important for the development of new clinical applications as well as the planning and guidance of cardiac interventional procedures. Manual delineation is labor intensive and subject to inter- and intra-observer variability. Therefore, it is highly desirable to develop an automatic technique for whole heart segmentation of cardiac magnetic resonance (MR) images. However, automating this process is complicated by the limited quality of acquired images and large shape variation of the heart between subjects.

Many works have shown the applicability of registration-based atlas propagation for the automatic segmentation, particularly in brain MR segmentation,

* This work has been funded by EPSRC grant EP/H02025X/1.

where the multi-atlas propagation and segmentation (MAPS) has gained its wide popularity in recent years [1,2,3,4,5]. In MAPS, each atlas is a combination of an intensity image and a label image. The intensity image is used for the registration process to map the space of the atlas to the unseen image, while the label image contains the segmentation information for propagation to the unseen image. The label image can also have prior knowledge such as the shape of the heart to assist the registration process [6].

In MAPS, good quality images, such as MR data with high signal-to-noise ratios and minimal artifacts, are selected as atlases for propagation to maximize the registration accuracy [2,3,4,5]. However, clinical cardiac MR images often contain strong artifacts, in particular the high resolution volumetric data, due to the effects from complex heart motions and long acquisition. Therefore, the registration of these images may have large errors or even fail. Furthermore, constructing an atlas with good image quality may need a large amount of training data and may be practically expensive in terms of manual image processing and data acquisition. Therefore, in this work we propose to build a good quality atlas from limited quality images and use the Multiple Path Propagation and Segmentation (MUPPS) strategy to achieve a result of multiple classifiers.

In MUPPS, the propagation of an atlas to unseen image can be done through a number of different paths to resemble the multiple classifier strategy. A set of propagation results are then available for fusion using the available methods. However, the number of available paths from the training subjects can be large, leading to the problem of scale. Three path selection strategies are hence investigated and an efficient method is proposed.

The main contribution of this work includes: (1) propose MUPPS and demonstrate its superiority over standard MAPS; (2) investigate propagation path selection and the effect on accuracy and efficiency.

2 Data and Method

2.1 Data

The cardiac MRI sequence used in the experiment was the balanced steady state free precession (b-SSFP) for whole heart imaging, at the end diastolic phase. The sequence was implemented on a 1.5T clinical scanner (Philips Healthcare, Best, The Netherlands) equipped with 32 independent receive channels. A fat saturation and T2 preparation pulses were used to null fat and to increase the contrast between blood and cardiac muscle. A free breathing scan was realized by enabling one navigator beam before data acquisition for each cardiac phase.

A *test dataset* of twenty-one cardiac MR volumes, all with voxel size $2 \times 2 \times 2$ (mm), was used in our experiments. The blood cavities of the four chambers and the myocardium of the left ventricle of all the images in the test set were separately segmented as gold standard for accuracy assessment. In addition, the pulmonary artery, ascending aorta and aortic arch, and descending aorta were also separately delineated for the atlas construction registration. They were done by fitting a deformable mesh model [7] with manual corrections. The manual



Fig. 1. Left: an MR image. Middle: the corresponding label image of the MR image. Right: an average intensity image.

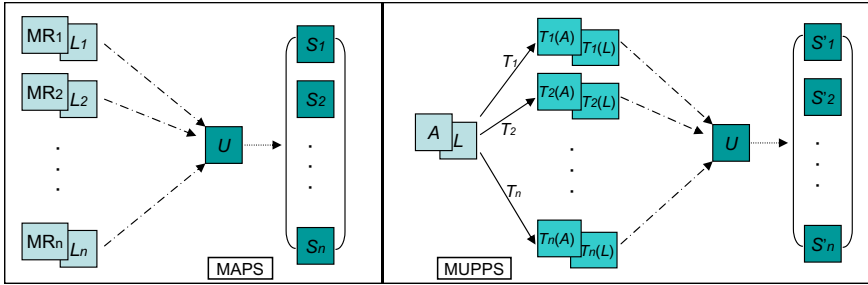


Fig. 2. The two propagation and segmentation schemes: $\{MR_i\}$ are the MR images of each subject and $\{L_i\}$ are the corresponding label images, U indicates the unseen image, and $\{S_i\}$ are the resultant propagation segmentations; A and L are the atlas intensity image and corresponding label image, $\{P_i\}$ and $\{T_i\}$ are the propagation paths and corresponding transformations.

segmentation was completed with the agreement by at least one cardiologist. Fig. 1 (left and middle) shows an MR image and its segmentation label image.

2.2 MULTIPLE Path Propagation and Segmentation (MUPPS)

In MAPS, different segmentations are obtained by registering the MR images, $\{MR_i\}$, in the atlas repository to the unseen image and propagating the corresponding labels, $\{L_i\}$, as Fig. 2 (left) shows. The clinical cardiac MR images often contain strong artifacts which may lead to large errors or even failures in the *propagation registration*. Therefore, we propose to construct a common atlas, consisting of an *average intensity image* A and the corresponding label image L , and register it to the unseen image through multiple paths to obtain multiple segmentations for label fusion, as Fig. 2 (right) illustrates.

The average intensity image A is computed from the set of training data, $\{MR_i\}$, by transforming all of them into a user-defined common space and averaging the intensity values [6]. These transformations, $\{T_i\}$, are computed from the registration of their corresponding binary label images $\{L_i\}$ and L , using the following three registration steps [6]: global affine registration, locally affine registration method (LARM) [8], and the fluid registration. The resultant average intensity image has better image quality, such as high signal and contrast-to-noise ratio and minimal artifacts, than the original atlas images in MAPS, and

the propagation registration using A is expected to achieve better results for segmentation. Fig. 1 (right) shows an example of A .

The paths in MUPPS are defined by $\{T_i\}$. For each path, we replace the original atlas images in MAPS, MR_i and L_i , by the images $T_i(A)$ and $T_i(L)$, transformed from A and L using T_i , for the propagation registration. This is equivalent to register A to the unseen image via the different paths in the shape manifold [9] defined by the spaces of $\{MR_i\}$, and hence the multiple path propagation and segmentation.

2.3 Registration and Path Ranking

Two registration procedures are used in MUPPS: propagation registration and selection registration.

Propagation registration is used to register $T_i(A)$ in MUPPS or MR_i in MAPS to the unseen image to achieve the segmentation. This process includes three steps [6]: the global affine registration, LARM using the local regions defined by the labels in the segmentation, and the FFD registration.

Selection registration is used to register the unseen image to the atlas intensity image and then compute the similarity between them to rank the path or atlas for propagation selection. In the multiple classifier strategy, the number of available atlases is normally greater than that of atlases needed to achieve the best result [3]. Therefore, the selection registration is also required to be computationally efficient compared to the propagation registration. Three registration schemes will be investigated in this work:

- the affine registration, referred to as *After affine*;
- the affine plus LARM registration, referred to as *After LARM*;
- the registration using global affine, LARM, and FFD registration, referred to as *After nonrigid*.

Path or atlas ranks are then calculated using the similarity values of the registered atlas intensity image and unseen image. We assess the similarity using the normalized mutual information (NMI) [10] with a mask on the interested region, similar to the work for brain MR MAPS [3]. The mask region is the endo- and epi-cardial surfaces of the transformed atlas label image and with an operation of morphological dilation of 5 mm.

2.4 Fusion and Accuracy Assessment

We evaluate three fusion strategies in MUPPS: the ‘vote rule’ [2], referred to as VOTE, shape-based average (SBA) [11], and the simultaneous truth and performance level estimation (STAPLE) [12]. The accuracy of segmentation results are assessed using Dice coefficient ($\frac{2|V_s \cap V_g|}{|V_s| + |V_g|}$) [13] between the segmentation and gold standard. The Dice score is calculated on the blood pool of the left ventricle (LV), left atrium (LA), right ventricle (RV), right atrium (RA), and myocardium (MYO). The mean of the volume size adjusted Dice of the five local regions is computed as the error of whole heart (Whole):

$$\text{MeanDice} = \frac{2 \sum_{i=1}^5 |V_{s_i} \cap V_{g_i}|}{\sum_{i=1}^5 (|V_{s_i}| + |V_{g_i}|)}. \quad (1)$$

We compare the mean Dice scores of MUPPS and MAPS using the two tailed, paired t-test.

3 Results and Discussions

MAPS vs MUPPS. Fig. 3 and Table 1 present the results of the standard MAPS, using the MR image of each subject as the atlas intensity image, and the proposed MUPPS, using the transformed mean atlas intensity image through different paths. In both schemes, we employed the leave-one-out to test the segmentation propagation, resulting in 420 cases in total for each scheme. Fig. 3 (left) presents the mean Dice scores of the two segmentation schemes. The mean Dice score of MUPPS was 0.059 ($P < 0.001$ and 95% confidence interval [0.0564, 0.0630]) greater than that of MAPS. Fig. 3 (right) gives the mean Dice scores of the VOTE fusion results. Both MAPS and MUPPS improved the accuracy after using the fusion. However, the best accuracy of MAPS was lower than that of MUPPS. The results of each local regions are provided in Table 1.

Path selection. Fig. 4 plots the mean Dice scores of the VOTE, STAPLE, SBA fusion results of MUPPS using the three different registration schemes for path ranking: the *After affine*, *After LARM*, and *After nonrigid*. All the three fusion methods agreed that the path selection strategy using the *After affine* was better than the other two. The best results of the *After affine* scheme were achieved when using the best 11-13 ranked paths for propagation and segmentation, while the other two both needed more, about 20, cases. Hence, we needed N_{sr} , which was 20 in our experiment, affine registration for path selection and ranking and 11-13 propagation registration to achieve the best result using the *After affine*. Also, affine registration was the fast one among the three selection registration schemes. It could be achieved within one minute while the nonrigid registration, LARM plus FFDs, might need over one hour in our experiment.

It may seem counterintuitive that the *After affine* performed better than the *After nonrigid*, as the image similarity after nonrigid registration should be more accurate in evaluating the segmentation accuracy and thus in ranking the paths. This could be an inconclusive result from the specific test dataset, as the difference between the *After nonrigid* and the other two on 11-13 of Fig. 3 (right) was small. This could also be the indication that the “good” segmentations ranked by *After nonrigid* might tend to have similar bias, resulting in the gain of accuracy by fusing them not being significant, compared to the fusion results of “bad” segmentations which nevertheless have unrelated random errors.

Accuracy of MUPPS. Fig. 5 and Table 1 provide the mean Dice scores of local regions and the whole heart. The $VOTE_U$, $STAPLE_U$, and SBA_U in Table 1 are the fusion results of MUPPS using the *After affine* registration to select eleven highest ranked paths for propagation. The accuracy improvement using fusion was evident, as all the three fusion techniques gained at least 0.016,

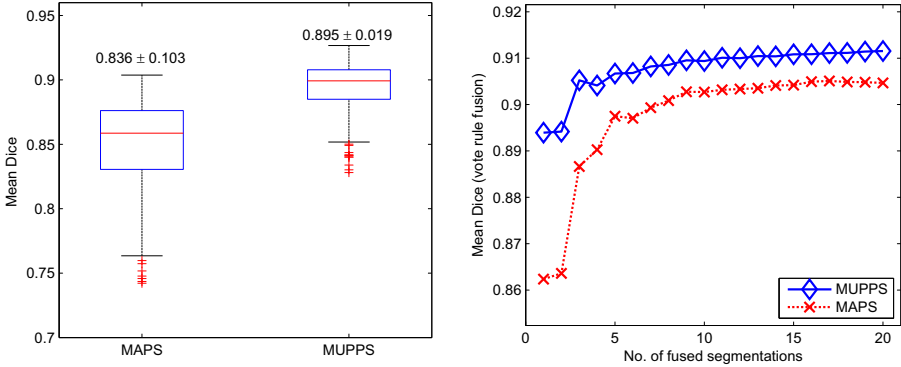


Fig. 3. Left: Box-and-Whisker diagram of the Dice scores using MAPS and MUPPS. Right: the Dice scores from the VOTE fusion results of them. Note that the difference of MAPS and MUPPS in local regions and the fusion results using STAPLE and SBA are almost identical to the results presented here.

Table 1. The mean Dice score of segmentation using standard MAPS and its best fusion results $VOTE_A$, and using MUPPS and the best fusion results, $VOTE_U$, $STAPLE_U$, and SBA_U . Note that the fusion results of MAPS using STAPLE and SBA are almost identical to $VOTE_A$.

Dice	Whole	MYO	LV	LA	RV	RA
MAPS	$.836 \pm .103$	$.722 \pm .116$	$.879 \pm .114$	$.788 \pm .084$	$.852 \pm .075$	$.783 \pm .105$
MUPPS	$.895 \pm .019$	$.836 \pm .027$	$.940 \pm .016$	$.869 \pm .028$	$.915 \pm .024$	$.871 \pm .050$
$VOTE_A$	$.905 \pm .018$	$.858 \pm .025$	$.946 \pm .016$	$.881 \pm .034$	$.921 \pm .021$	$.879 \pm .041$
$VOTE_U$	$.911 \pm .017$	$.866 \pm .026$	$.949 \pm .014$	$.883 \pm .025$	$.930 \pm .023$	$.892 \pm .041$
$STAPLE_U$	$.911 \pm .017$	$.866 \pm .025$	$.949 \pm .014$	$.880 \pm .026$	$.930 \pm .021$	$.888 \pm .044$
SBA_U	$.912 \pm .016$	$.887 \pm .025$	$.949 \pm .014$	$.883 \pm .026$	$.930 \pm .021$	$.891 \pm .038$

about one standard deviation, in Whole category. The gains in local regions were different. The myocardium, which resembles a thin cup-shaped sheet and normally reports a worst segmentation result in the five regions, achieved the most significant improvement, about 0.03 after using the fusion. By contrast, the left ventricle had the least improvement, less 0.01, and other categories had less than 0.02. We therefore conclude that it can be more beneficial to use MUPPS and fusion techniques in the segmentation of thin, sheet-shaped objects where one propagation path tends to have certain random errors in the result. Note that the conclusion from the surface distance as error measure was similar, as the mean root mean squared surface distance of MUPPS without fusion, and the fusion results using VOTE, STAPLE, and SBA were 1.29 ± 0.319 , 1.13 ± 0.275 , 1.25 ± 0.346 , and 1.13 ± 0.248 (mm), respectively. Only that STAPLE fusion seemed to be disfavored in this error measure.

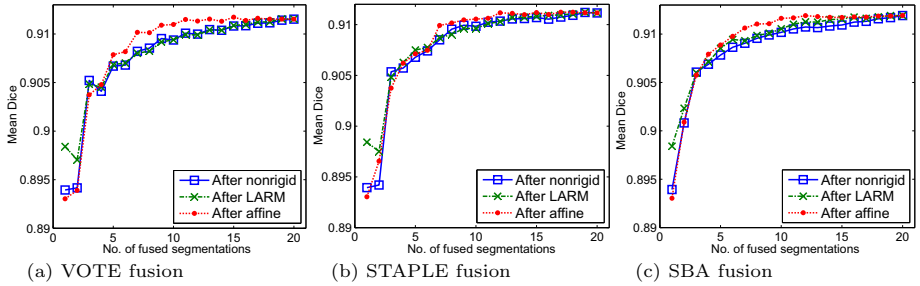


Fig. 4. The mean Dice scores using different path selection strategies: *After affine*, *After LARM*, and *After nonrigid*

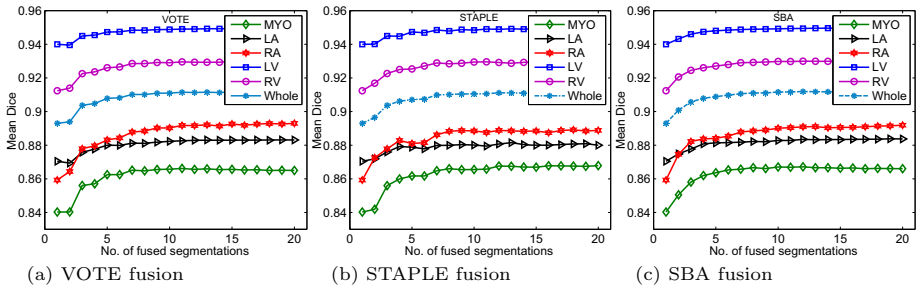


Fig. 5. The mean Dice scores of MUPPS for whole heart and each local region segmentation using the *After affine* path selection

4 Conclusion

In this paper, we propose a new multi-classifier method, the Multiple Path Propagation and Segmentation (MUPPS), for the whole heart segmentation of cardiac MRI. We showed that MUPPS achieved better mean Dice scores than the standard MAPS, 0.895 VS 0.836 before the fusion and 0.911 VS 0.905 after the fusion. To minimize the number of propagation paths in MUPPS for a segmentation case, we studied three path ranking schemes. The best scheme was to rank the path using the image similarity of the transformed atlas and unseen image after an affine registration. The segmentation framework of MUPPS achieved 0.911 ± 0.016 mean Dice score, a 0.015 gain of accuracy, using eleven highest ranked paths and the fusion techniques. In particular for the segmentation of myocardium, a gain of 0.03 was reported. The whole heart segmentation error of RMS surface distance was 1.13 ± 0.248 (mm).

Note that the shape manifold in MUPPS is different from the linear space formed from the principal components in the statistical shape model [14]. Also, it only needs about 10 images to construct A with good quality [6]. Hence, for a new image MR_{n+1} , only one off-line registration is needed to compute the new path T_{n+1} and no reconstruction of A is required.

A potential advantage of MUPPS is to further discretize the shape manifold to increase the number of searching paths such as these between two existing paths T_i and T_j . Furthermore, we used a mask on the whole heart for the path ranking. It may be more accurate to rank the path based on the similarity within certain local region such as for the myocardium. The segmentation of the whole heart can then be considered as several segmentations of a number of local regions and different rankings can be used for different regions [5]. Finally, the results of MUPPS contain a set of deformation fields which are the transformations between the atlas and unseen image plus certain random errors. By averaging these fields, we should obtain a better estimation of the registration ground truth for the applications where the spatial mapping is of the interest. We will address these in future work.

References

1. Rohlfing, T., Maurer, J.C.R.: Multi-classifier framework for atlas-based image segmentation. *Pattern Recognition Letters* 26, 2070–2079 (2005)
2. Heckemann, R.A., Hajnal, J.V., Aljabar, P., Rueckert, D., Hammers, A.: Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage* 33(1), 115–126 (2006)
3. Aljabar, P., Heckemann, R., Hammers, A., Hajnal, J., Rueckert, D.: Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *NeuroImage* 46(3), 726–738 (2009)
4. Leung, K., Barnes, J., Ridgway, G., Bartlett, J., Clarkson, M., Macdonald, K., Schuff, N., Fox, N., Ourselin, S.: Automated cross-sectional and longitudinal hippocampal volume measurement in mild cognitive impairment and Alzheimer’s disease. *Neuroimage* 51(4), 1345–1359 (2010)
5. van Rikxoort, E., Isgum, I., Arzhaeva, Y., Staring, M., Klein, S., Viergever, M., Pluim, J., van Ginneken, B.: Adaptive local multi-atlas segmentation application to the heart and the caudate nucleus. *Medical Image Analysis* (14), 39–49 (2010)
6. Zhuang, X., Rhode, K., Razavi, R., Hawkes, D.J., Ourselin, S.: A registration-based propagation framework for automatic whole heart segmentation of cardiac MRI. *IEEE Transactions on Medical Imaging* (2010), doi:10.1109/TMI.2010.2047112
7. 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, 70–84 (2009)
8. Zhuang, X., Rhode, K., Arridge, S., Razavi, R., Hill, D., Hawkes, D., Ourselin, S.: An atlas-based segmentation propagation framework using locally affine registration – application to automatic whole heart segmentation. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) *MICCAI 2008, Part II. LNCS*, vol. 5242, pp. 425–433. Springer, Heidelberg (2008)
9. Hamm, J., Davatzikos, C., Verma, R.: Efficient large deformation registration via geodesics on a learned manifold of images. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009. LNCS*, vol. 5761, pp. 680–687. Springer, Heidelberg (2009)
10. Studholme, C., Hill, D.L.G., Hawkes, D.J.: An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition* 32(1), 71–86 (1999)

11. Rohlfing, T., Maurer, C.R.: Shape-based averaging. *IEEE Trans. Image Processing* 16(1), 153–161 (2007)
12. Warfield, S.K., Zou, K.H., Wells III, W.M.: Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation. *IEEE Transaction on Medical Imaging* 23(7), 903–921 (2004)
13. Dice, L.R.: Measures of the amount of ecologic association between species. *Ecology* 26(3), 297–302 (1945)
14. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models: Their training and application. *Computer Vision and Image Understanding* 61(1), 38–59 (1995)