Published in final edited form as:

Proc SPIE Int Soc Opt Eng. 2017 February 11; 10133: . doi:10.1117/12.2254147.

Multi-atlas Segmentation Enables Robust Multi-contrast MRI Spleen Segmentation for Splenomegaly

Yuankai Huo^a, Jiaqi Liu^b, Zhoubing Xu^a, Robert L. Harrigan^a, Albert Assad^c, Richard G. Abramson^d, and Bennett A. Landman^{a,b,d,e}

^aElectrical Engineering, Vanderbilt University, Nashville, TN, USA 37235

^bComputer Science, Vanderbilt University, Nashville, TN, USA 37235

^cIncyte Corp., Wilmington, DE, USA 19803

^dRadiology and Radiological Science, Vanderbilt University, Nashville, TN, USA 37235

^eBiomedical Engineering, Vanderbilt University, Nashville, TN, USA 37235

Abstract

Non-invasive spleen volume estimation is essential in detecting splenomegaly. Magnetic resonance imaging (MRI) has been used to facilitate splenomegaly diagnosis in vivo. However, achieving accurate spleen volume estimation from MR images is challenging given the great inter-subject variance of human abdomens and wide variety of clinical images/modalities. Multi-atlas segmentation has been shown to be a promising approach to handle heterogeneous data and difficult anatomical scenarios. In this paper, we propose to use multi-atlas segmentation frameworks for MRI spleen segmentation for splenomegaly. To the best of our knowledge, this is the first work that integrates multi-atlas segmentation for splenomegaly as seen on MRI. To address the particular concerns of spleen MRI, automated and novel semi-automated atlas selection approaches are introduced. The automated approach interactively selects a subset of atlases using selective and iterative method for performance level estimation (SIMPLE) approach. To further control the outliers, semi-automated craniocaudal length based SIMPLE atlas selection (L-SIMPLE) is proposed to introduce a spatial prior in a fashion to guide the iterative atlas selection. A dataset from a clinical trial containing 55 MRI volumes (28 T1 weighted and 27 T2 weighted) was used to evaluate different methods. Both automated and semi-automated methods achieved median DSC > 0.9. The outliers were alleviated by the L-SIMPLE (≈ 1 min manual efforts per scan), which achieved 0.9713 Pearson correlation compared with the manual segmentation. The results demonstrated that the multi-atlas segmentation is able to achieve accurate spleen segmentation from the multi-contrast splenomegaly MRI scans.

1. INTRODUCTION

Splenomegaly is an abnormal enlargement of the spleen, which is associated with liver disease, infection and cancer [1]. Accurate non-invasive spleen volumetric size estimation plays an essential role in splenomegaly diagnosis and scientific studies [2]. Ultrasound [3–5]

and computerized tomography (CT) [6–8] have been used as the major imaging techniques in quantifying spleen size [9, 10]. The manual delineation of whole spleen slice-by-slice has been regarded as a gold standard of spleen volume estimation. However, the manual tracing on entire 3D spleen is time and resource consuming. To accelerate manual efforts, the first family of techniques was proposed to enable routine splenic assessment relied on manually derived 1D index measurements (e.g., splenic length, width, thickness, etc.) to replace time consuming 3D whole spleen delineation. Using such measurements, the spleen volume size can be estimated using regression models [7]. A second family of techniques seeks to automate 3D volumetric spleen segmentation, e.g., shape/contour based models [11, 12], intensity based models [13], graph cuts [14, 15], learning based models [16], and atlas-based methods [17, 18].

A major challenge of automated MRI spleen segmentation is that the absolute intensity of MRI is not in a quantitative scale like the Hounsfield Units (HU) in CT. Another challenge is that the relative intensity contrasts of abdominal tissues are in large variation using the different contrast mechanisms (e.g., T1-weighted (T1w), T2-weighted (T2w), proton density (PD), etc.). Such challenges hinder frequently used CT segmentation methods, which depend on absolute intensity scales. In our previous work [19–22], the multi-atlas segmentation (MAS) framework has been introduced to the CT abdominal organ segmentation. Since the state-of-the-art multi-atlas label fusion methods typically search the intensity similarities between patches that are not restricted to the unified intensity scales or contrasts [23–25], the MAS is an appealing method of conducting MRI spleen segmentation.

In this paper, we propose to use MAS for MRI spleen segmentation across imaging modalities. To address the particular concerns of (1) large intensity/contrast variation of spleen MRI and (2) large variation of spleen size and shape in splenomegaly scans (Figure 1), automated and novel semi-automated atlas selection approaches were developed. The automated approach interactively selects a subset of atlases using selective and iterative method for performance level estimation (SIMPLE) approach [26]. In the semi-automated method, the craniocaudal length (L) based SIMPLE atlas selection (L-SIMPLE) is used to introduce a spatial prior in a Bayesian fashion during the atlas selection. For comparison, the segmentation results using all atlases and using a single L-based atlas selection pipelines are demonstrated.

2. METHODS

2.1 Data and Experimental Setup

55 abdominal MRI scans were acquired from splenomegaly patients (27 T1w/ 28 T2w) from a clinical trial. The spleens were manually traced by an experienced rater using the Medical Image Processing Analysis and Visualization (MIPAV) software [27]. In the manual spleen segmentation, the minimum spleen size is 368 cubic centimeter (cc) while the maximum spleen size is 5670 cc. The mean spleen volume is 1881 cc with a standard deviation 1219 cc.

This heterogeneous dataset is used to evaluate the performance of the MAS on detecting splenomegaly from multi-contrast MRI scans using a leave-one-out strategy. The Pearson

correlation, Dice similarity coefficient (DSC), means surface distance (MSD) and Hausdorff distance (HD) are employed as the metrics to evaluate the performance of different segmentation methods in comparison of manual segmentations. All statistical significance tests are made using a Wilcoxon signed rank test ($p < 0.05$).

2.2 General Multi-atlas Segmentation Framework

The canonical MAS framework included registration, atlas selection, label propagation and label fusion [28]. The N4 bias field correction was applied to on the target image, which was then resampled to 1.5 mm isotropic voxel size. Then, the affine registration followed by non-rigid registration was conducted to register each atlas image and label to the target image using DEnsE Displacement Sampling (DEEDS) registration method [29]. The atlas selection approaches were used to select the best ten atlases from all registered atlases. Finally, the selected atlases were fused to achieve the final spleen segmentation using joint label fusion (JLF) [30]. To address substantial registration failures and the large inhomogeneity in the splenomegaly dataset, four different atlas selection strategies (pipelines) were evaluated with the general MAS framework.

2.3 Automated and Semi-automated Pipelines

Two automated MAS pipelines are shown in Figure 2. In the first pipeline (Pipeline 1), the atlas selection step was not included during the general MAS framework, which means all available atlases were used in the label fusion. In the second pipeline (Pipeline 2), the SIMPLE method [26] was used to select the ten best atlases during an iterative procedure.

Two semi-automated MAS pipelines are illustrated in Figure 2. In Pipeline 3, the atlases were selected by comparing the spleen volume sizes between the target image and all atlases. We hypothesized that the atlases with similar spleen size as the target image would lead to more accurate fusion results. Instead of conducting time-consuming whole spleen manual tracing for a target image, a unidimensional splenic index measurement craniocaudal spleen length (L) (Figure 2) was used to estimate the spleen volume size [7]. L was derived by multiplying the number of visible axial slices of spleen by the slice thickness [7]. Then, the spleen volume (Vol) was estimated by $L = (0.0126 \times \text{Vol}) + 5.8006$. Since L represents the variations of the spleen shape in this splenomegaly dataset, the atlas selection was conducted by selecting ten atlases with the closest L values compared with the target image.

By combining the advantages in Pipeline 2 and Pipeline 3, novel L-SIMPLE approach (Pipeline 4) was proposed. First, a probabilistic map was derived by averaging the registered spleen labels selected by L. Then the best atlases were selected by integrating the spatial prior with the SIMPLE framework in a Bayesian manner following the similar idea as [19, 20]. The inputs of the L-SIMPLE methods were (1) m registered spleen labels atlases $\mathbf{A} = \{A_1, A_2, \dots, A_m\}$, where each A_j is an atlas. (2) The measurement L of the target images. The output was n selected atlases $\mathbf{A}'$ where $n \leq m$. The complete algorithm is:

Step 1) Initialize $\mathbf{A}$ as all atlases. The spleen spatial probabilistic prior $f(T=1)$ is derived by averaging the registered label atlases selected by Pipeline 3. $f(T=1)$ is the probability map of the spleen (spleen label is 1) while $f(T=0)$ is the probability map of non-spleen tissues and background.

Step 2) At each iteration k , $\mathbf{A}^k$ is the set of the remaining n atlases. For each voxel i , the likelihood function of spleen is defined by

$$f(\mathbf{A}_i^k | T_i=1) = \frac{1}{n} \sum_{j=1,2,\dots,n} A_{ji}^k \quad (1)$$

$$f(\mathbf{A}_i^k | T_i=0) = 1 - \frac{1}{n} \sum_{j=1,2,\dots,n} A_{ji}^k$$

Step 3) From a Bayesian approach, the posterior probability of spleen at voxel i is derived as

$$f(T_i=1 | \mathbf{A}_i^k) = \frac{f(\mathbf{A}_i^k | T_i=1) f(T_i=1)}{f(\mathbf{A}_i^k)} \quad (2)$$

$$f(T_i=0 | \mathbf{A}_i^k) = \frac{f(\mathbf{A}_i^k | T_i=0) f(T_i=0)}{f(\mathbf{A}_i^k)}$$

Step 4) The label of the intermediate spleen segmentation S at voxel i is derived as

$$S_i=1, \text{ if } f(T_i=1 | \mathbf{A}_i^k) \geq f(T_i=0 | \mathbf{A}_i^k) = 0, \text{ if } f(T_i=1 | \mathbf{A}_i^k) < f(T_i=0 | \mathbf{A}_i^k) \quad (3)$$

Step 5) The one-dimensional weight vector w is derived by calculating DSC between each A_j^k and S .

$$w_j = \text{DSC}(A_j^k, S) \quad (4)$$

Step 6) For the next iteration, the $\mathbf{A}^{k+1}$ is a subset of $\mathbf{A}^k$ by comparing w_j with mean ($\bar{w}$) and standard deviation (σ_w) of w .

$$\mathbf{A}^{k+1} = \{A_j^k\}, \text{ for } j: w_j > (\bar{w} - \sigma_w) \quad (5)$$

Step 7) If the size of $\mathbf{A}^{k+1}$ is less than the number of atlases that we want or k hits the maximum iteration, the iteration is terminated and $\mathbf{A}' = \mathbf{A}^k$ is returned as selected atlases. Otherwise, go to Step 2.

2.4 Multi-contrast MRI Spleen MAS Segmentation

In MRI, users are free to design and use particular sequences based on different clinical and scientific purposes. This provides the opportunity to achieve contrast which is optimized for certain tissue types. However, such flexibility makes it challenging to design a single spleen

MRI segmentation framework that is compatible with multiple contrast mechanisms. To evaluate the robustness of MAS on multi-model data, the T1w and T2w scans are both employed as atlases and target images. Different scenarios are considered in this study: (1) Using both T1w and T2w images as atlases and target images (Figure 4). (2) T1w images are employed as target images when using both T1w and T2w atlases, only T1w atlases, and only T2w atlases (Figure 5a). (3) T2w images are employed as target images when using both T1w and T2w atlases, only T1w atlases, and only T2w atlases (Figure 5b).

3. RESULTS

A leave-one-out validation strategy was used to evaluate the performance of the four spleen MAS pipelines. Figure 3 presents the qualitative results of four pipelines compared with manual segmentations. Three scans were shown with axial and coronal slices to represent the best, median and worst performance in Pipeline 4. Figure 4a indicates that the pipelines 2 and 4 achieved better statistical performance than pipelines 1 and 3 on DSC, MSD and HD. Figure 4b shows that the strongest Pearson correlation value was achieved by Pipeline 4 when comparing the estimated spleen volume to the manual segmentation volume. Table 1 reported the metrics as well as the number of failures of four pipelines with the overall best performance in bold. The “failure” was defined as the numbers of scans whose DSC values were smaller than 0.8, 0.75 or 0.7 (Table 1). Figure 5 evaluated the performance of using multi-contest MRI sequences by considering T1w and T2w as target images separately.

4. CONCLUSION

We present the first work using MAS on MRI for splenomegaly. To address the substantial abdominal registration errors and the particular concerns of the large variations in MRI intensities, four different atlas selection based pipelines were presented and compared in this work. Among the two automated pipelines, the SIMPLE based method (Pipeline 2) performed better than the method that did not use atlas selection (Pipeline 1) (Figure 4). We demonstrated that SIMPLE based atlas selection was able to improve the label fusion performance on MRI for splenomegaly.

To further improve the performance for the outliers in Pipeline 2, the semi-automated L-SIMPLE framework was introduced by using the unidimensional splenic index measurement L. The L-SIMPLE generated a spatial prior to guide the interactive atlas selection in Bayesian framework. The results demonstrated that the L-SIMPLE (Pipeline 4) performed better than only using craniocaudal length L of spleen (Pipeline 3) (Figure 4) and had a smaller number of failures than SIMPLE (Pipeline 2) (Table 1).

Figure 5 demonstrates that using both T1w and T2w atlases provided more candidate atlases during atlas selection, which yielded statistically better DSC than only using atlases from a single modality.

5. DISCUSSION

From the results, the automated Pipeline 3 achieved >0.9 median DSC relative to manual segmentation. However, there were cases where the majority of registrations failed in such a

similar fashion that the SIMPLE atlas selection incorrectly chose atlases which were not similar to the target. To address such cases, the craniocaudal length (L) of spleen was used to derive a prior probabilistic map to guide the atlas selection in Bayesian fashion. In this study, the L-SIMPLE (Pipeline 4) using 1D measurement L (≈ 1 min manual work per scan) was much less time consuming than the whole spleen 3D delineation (≈ 20 min manual work per scan) while achieving the 0.9713 Pearson correlation on the multi-contrast splenomegaly dataset. The continuing investigation of this work would be the automated L derivation using random forest [31], deep convolutional neural network [32], or other approaches.

In Table 1, the number of failures is defined by counting the number of segmented images whose spleen DSC is smaller than particular values. In the future, it would be interesting to define “failures” in a quantitative manner that is well accepted in the community. The results (Figure 5) indicate that the MAS framework is not sensitive to the modality difference between atlases and target images. It means all available atlases (with atlas selection), even across modalities, should be included to deal with large body variability. To further validate this claim, a validation dataset for which both T1w and T2w scans are included for every subject would be better. It is not claimed that the registration method, number of atlases and the label fusion method used in this work are optimal.

Acknowledgments

This research was supported by NSF CAREER 1452485, NIH grants 5R21EY024036, 1R21NS064534, 2R01EB006136 (Dawant), 1R03EB012461 (Landman) and R01NS095291 (Dawant). InCyte Corporation (Abramson/Landman). This research was conducted with the support from Intramural Research Program, National Institute on Aging, NIH. This study was in part using the resources of the Advanced Computing Center for Research and Education (ACCRE) at Vanderbilt University, Nashville, TN. This project was supported in part by ViSE/VICTR VR3029 and the National Center for Research Resources, Grant UL1 RR024975-01, and is now at the National Center for Advancing Translational Sciences, Grant 2 UL1 TR000445-06. We are grateful for the assistance of Kunal Nabar in helping to prepare this manuscript.

REFERENCES

1. Eichner ER. Splenic function: normal, too much and too little. *Am J Med.* 1979; 66(2):311–320. [PubMed: 371397]
2. Paley, MR., Ros, PR. *Imaging of spleen disorders.* Springer; 2002.
3. De Odorico I, Spaulding KA, Pretorius DH. Normal splenic volumes estimated using three-dimensional ultrasonography. *Journal of ultrasound in medicine.* 1999; 18(3):231–236. [PubMed: 10082358]
4. Rodrigues AJ, Rodrigues CJ, Germano MA. Sonographic assessment of normal spleen volume. *Clinical Anatomy.* 1995; 8(4):252–255. [PubMed: 7552962]
5. Spielmann AL, DeLong DM, Kliever MA. Sonographic evaluation of spleen size in tall healthy athletes. *American Journal of Roentgenology.* 2005; 184(1):45–49. [PubMed: 15615949]
6. Prassopoulos P, Daskalogiannaki M, Raissaki M. Determination of normal splenic volume on computed tomography in relation to age, gender and body habitus. *Eur Radiol.* 1997; 7(2):246–248.
7. Bezerra AS, D'Ippolito G, Faintuch S. Determination of splenomegaly by CT: is there a place for a single measurement? *AJR Am J Roentgenol.* 2005; 184(5):1510–1513. [PubMed: 15855107]
8. Linguraru MG, Sandberg JK, Jones EC. Assessing splenomegaly: automated volumetric analysis of the spleen. *Acad Radiol.* 2013; 20(6):675–684. [PubMed: 23535191]
9. Yetter EM, Acosta KB, Olson MC. Estimating splenic volume: sonographic measurements correlated with helical CT determination. *American Journal of Roentgenology.* 2003; 181(6):1615–1620. [PubMed: 14627584]

10. Lamb P, Lund A, Kanagasabay R. Spleen size: how well do linear ultrasound measurements correlate with three-dimensional CT volume assessments? *The British journal of radiology*. 2002; 75(895):573–577. [PubMed: 12145129]
11. Campadelli P, Casiraghi E, Pratissoli S. Fully automatic segmentation of abdominal organs from CT images using fast marching methods. :554–559.
12. Campadelli P, Casiraghi E, Pratissoli S. A segmentation framework for abdominal organs from CT scans. *Artif Intell Med*. 2010; 50(1):3–11. [PubMed: 20542673]
13. Campadelli P, Pratissoli S, Casiraghi E. Automatic abdominal organ segmentation from CT images. *ELCVIA: electronic letters on computer vision and image analysis*. 2009; 8(1):001–014.
14. Chen X, Bagci U. 3D automatic anatomy segmentation based on iterative graph-cut-ASM. *Med Phys*. 2011; 38(8):4610–4622. [PubMed: 21928634]
15. Chen X, Udupa JK, Bagci U. Medical image segmentation by combining graph cuts and oriented active appearance models. *IEEE Trans Image Process*. 2012; 21(4):2035–2046. [PubMed: 22311862]
16. Behrad A, Masoumi H. Automatic spleen segmentation in MRI images using a combined neural network and recursive watershed transform. :63–67.
17. Linguraru MG, Sandberg JK, Li Z, et al. Atlas-based automated segmentation of spleen and liver using adaptive enhancement estimation. :1001–1008.
18. Linguraru MG, Sandberg JK, Li Z. Automated segmentation and quantification of liver and spleen from CT images using normalized probabilistic atlases and enhancement estimation. *Med Phys*. 2010; 37(2):771–783. [PubMed: 20229887]
19. Xu Z, Asman AJ, Shanahan PL. SIMPLE is a good idea (and better with context learning). *Med Image Comput Comput Assist Interv*. 2014; 17(Pt 1):364–371. [PubMed: 25333139]
20. Xu Z, Burke RP, Lee CP. Efficient multi-atlas abdominal segmentation on clinically acquired CT with SIMPLE context learning. *Med Image Anal*. 2015; 24(1):18–27. [PubMed: 26046403]
21. Xu Z, Li B, Panda S. Shape-Constrained Multi-Atlas Segmentation of Spleen in CT. *Proc SPIE Int Soc Opt Eng*. 2014; 9034:903446. [PubMed: 24817808]
22. Burke RP, Xu Z, Lee CP. Multi-Atlas Segmentation for Abdominal Organs with Gaussian Mixture Models. *Proc SPIE Int Soc Opt Eng*. 2015; 9417
23. Heinrich MP, Wilms M, Handels H. Multi-atlas Segmentation Using Patch-Based Joint Label Fusion with Non-Negative Least Squares Regression. :146–153.
24. Schreiber E, Marcus DM, Fox T. Multiatlas segmentation of thoracic and abdominal anatomy with level set-based local search. *J Appl Clin Med Phys*. 2014; 15(4):4468. [PubMed: 25207393]
25. Zhou Y, Bai J. Multiple abdominal organ segmentation: an atlas-based fuzzy connectedness approach. *IEEE Trans Inf Technol Biomed*. 2007; 11(3):348–352. [PubMed: 17521085]
26. Langerak TR, van der Heide UA, Kotte AN. Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (SIMPLE). *IEEE Trans Med Imaging*. 2010; 29(12):2000–2008. [PubMed: 20667809]
27. McAuliffe MJ, Lalonde FM, McGarry D, et al. Medical image processing, analysis and visualization in clinical research. :381–386.
28. Iglesias JE, Sabuncu MR. Multi-atlas segmentation of biomedical images: A survey. *Med Image Anal*. 2015; 24(1):205–219. [PubMed: 26201875]
29. Heinrich MP, Jenkinson M, Brady M. MRF-based deformable registration and ventilation estimation of lung CT. *IEEE Trans Med Imaging*. 2013; 32(7):1239–1248. [PubMed: 23475350]
30. Wang HZ, Suh JW, Das SR. Multi-Atlas Segmentation with Joint Label Fusion. *Ieee Transactions on Pattern Analysis and Machine Intelligence*. 2013; 35(3):611–623. [PubMed: 22732662]
31. Liaw A, Wiener M. Classification and regression by randomForest. *R news*. 2002; 2(3):18–22.
32. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. :1097–1105.

Figure 1.

Datasets in this study exhibited a range of spleen volumes and multiple imaging contrasts (T1w and T2w).

Figure 2.

Flow chart of the four different MAS pipelines. Pipeline 1 conducts the label fusion without atlas selection. Pipeline 2 conducts the atlas selection using the SIMPLE statistical selection method. Pipeline 3 selects atlases which have the closer L to the target image. Pipeline 4 is the L-SIMPLE method, which combines the advantages in Pipelines 2 and 3. A spatial prior is derived using L to guide the statistical atlas selection procedure.

Figure 3.

Qualitative spleen segmentations results of four pipelines are compared with manual segmentations. The “Largest DSC”, “Median DSC” and “Smallest DSC” panes illustrate the scans that have largest, median and smallest DSC values respectively using pipeline 4.

Figure 4.

Quantitative spleen segmentations results of four pipelines using all datasets in a leave-one-out-approach. (a) The Dice similarity coefficient (DSC), the mean surface distance (MSD) and Hausdorff distance (HD) are shown. “Ref.” indicates the reference method, which is statistically compared with other methods using Wilcoxon signed rank tests. The “*” indicates the $p < 0.05$ while the “NS” means not significant. (b) The scatter plots and the Pearson correlation values between four pipelines and the manual segmentation are indicated.

Figure 5.

The performance of Pipeline 4 using subsets of the scans by MRI contrast are shown. (a) The DSC of spleen segmentation for 27 T1w target images are demonstrated. (b) The DSC of 28 T2w target images are shown. The “T1&T2” means use all T1w and T2w atlases. The “T1 only” means only the T1w atlases are used to segment T1w target images while the “T2 only” only the T2w atlases are used to segment T2w target images. The “*” indicates the differences are significant.

Table 1

Performance of four pipelines using all 55 volumes in a leave-one-out-approach

		Automated		Semi-automated	
		Pipeline 1	Pipeline 2	Pipeline 3	Pipeline 4
Pearson Correlation		0.8481	0.8507	0.9689	0.9713
Dice similarity coef. (DSC)	median	0.8599	0.9032	0.8828	0.9006
	mean $\pm$ std	0.82 $\pm$ 0.11	0.87 $\pm$ 0.11	0.87 $\pm$ 0.05	0.88 $\pm$ 0.06
Mean surface distance (MSD) (mm)	median	5.6563	3.2446	4.1007	3.5097
	mean $\pm$ std	7.2 $\pm$ 4.8	4.82 $\pm$ 5.90	4.77 $\pm$ 2.06	3.88 $\pm$ 2.14
Hausdorff distance (HD) (mm)	median	50.0212	30.9947	48.9910	32.1552
	mean $\pm$ std	61.32 $\pm$ 29.61	40.30 $\pm$ 28.76	60.36 $\pm$ 34.7	37.08 $\pm$ 15.13
Number of Failures (Use different DSC as thresholds)	DSC<0.8	19	6	7	6
	DSC<0.75	11	4	3	2
	DSC<0.7	9	3	0	0