Published in final edited form as:

Proc SPIE Int Soc Opt Eng. 2017 February 11; 10133: . doi:10.1117/12.2254437.

Multi-Atlas Spleen Segmentation on CT Using Adaptive Context Learning

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

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

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

^cIncyte Corp., Wilmington, DE, USA 19803

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

Abstract

Automatic spleen segmentation on CT is challenging due to the complexity of abdominal structures. Multi-atlas segmentation (MAS) has shown to be a promising approach to conduct spleen segmentation. To deal with the substantial registration errors between the heterogeneous abdominal CT images, the context learning method for performance level estimation (CLSIMPLE) method was previously proposed. The context learning method generates a probability map for a target image using a Gaussian mixture model (GMM) as the prior in a Bayesian framework. However, the CLSSIMPLE typically trains a single GMM from the entire heterogeneous training atlas set. Therefore, the estimated spatial prior maps might not represent specific target images accurately. Rather than using all training atlases, we propose an adaptive GMM based context learning technique (AGMMCL) to train the GMM adaptively using subsets of the training data with the subsets tailored for different target images. Training sets are selected adaptively based on the similarity between atlases and the target images using cranio-caudal length, which is derived manually from the target image. To validate the proposed method, a heterogeneous dataset with a large variation of spleen sizes (100 cc to 9000 cc) is used. We designate a metric of size to differentiate each group of spleens, with 0 to 100 cc as small, 200 to 500cc as medium, 500 to 1000 cc as large, 1000 to 2000 cc as XL, and 2000 and above as XXL. From the results, AGMMCL leads to more accurate spleen segmentations by training GMMs adaptively for different target images.

1. INTRODUCTION

Spleen volume is an essential biomarker of liver disease and infection [1–3]. To estimate the spleen volume in vivo, computerized tomography (CT) is one of the most widely accepted imaging techniques [1, 4, 5]. The manual delineation of 3D whole spleen has been regarded as a gold standard of estimating accurate spleen volume from clinical CT image [6, 7]. However, manually tracing an entire spleen is time and resource consuming. Instead, the splenic index metrics (e.g., splenic length, width, etc.) are typically traced to estimate the spleen volume, which reduces the complexity and time of manual work [4]. In the past

decades, automatic spleen estimation methods have been proposed to conduct spleen volume estimation [8–14], but have not been widely applied in splenomegaly.

Among automatic spleen estimation methods, the multi-atlas segmentation (MAS) is one of the most popular families due to its flexibility and accuracy [15–21]. Our prior works used MAS to conduct 3D whole spleen segmentation using context learning based selective and iterative method for performance level estimation (CLSIMPLE) framework [15–18]. The key contribution of such work is to improve the SIMPLE method by integrating spatial priors of abdominal organs, which are calculated by a Gaussian mixture model based context learning (GMMCL) method [18]. Practically, when the sizes of spleen in atlas images and the target image are on similar scales, the GMMCL yields accurate prior probability maps. However, when the target scan has much larger or smaller spleen size compared with atlases, (e.g., splenomegaly spleens shown in Figure 1), GMMCLs typically provide worse spatial priors, because the GMMCL is trained based on the preselected atlases before conducting MAS without taking the target spleen size into account.

To improve the performance of context learning by taking the individual feature of target image into account, a semi-automated adaptive Gaussian mixture model (AGMM) based context learning technique (AGMMCL) is proposed to train the GMM adaptively for different target images. The candidate atlases used to train the GMM are selected by the manually traced splenic index measurement, cranio-caudal length (L) [4]. The proposed AGMMCL introduces the robustness of L estimation to guide the context learning, which derives better performance than conducting the two methods respectively.

2. METHODS

2.1 Dataset

We use a set of 69 clinically acquired CT images in this project (50 from MICCAI 2015 MeDiCAL challenges, https://miccai2015.org/frontend/index.php?page_id=589, and 19 from a clinically acquired splenomegaly cohort). The scans are of different volume sizes ($512 \times 512 \times 47 \sim 512 \times 512 \times 195$), and the in-plane resolution ranges from $0.59 \times 0.59 \text{ mm}^2$ to $0.98 \times 0.98 \text{ mm}^2$. The scans in MeDiCAL dataset typically have regular spleen size while the 19 splenomegaly dataset consists of larger spleens. The spleens were manually traced by an experienced rater using MIPAV software[22]. All images and labels were cropped along the cranio-caudal axis following [23].

2.2 Multi-atlas Segmentation Framework

Figure 2 illustrates the AGMMCL based MAS approach. The target image is firstly localized and cropped using random forest bounding box localization [24]. Then, all atlas images and labels are registered to the cropped target image using affine and non-rigid registrations by NiftyReg [25]. Meanwhile, a prior probability map for the target image is derived adaptively from the AGMM approach (described in following section). Then the most likely successful atlases are selected using the iterative context learning and finally fused to the final spleen segmentation using joint label fusion (JLF) [26].

2.3 Adaptive Gaussian Mixed Model based Context Learning

First, the GMM model is trained adaptively based on a selected subset from all training atlases. The cranio-caudal length (L) is traced and calculated for the target image. Five atlases which have the closest L compared the test image are selected. Then, selected atlases are used to train the GMM model in the similar way as [18]. In this work, four types of features from the manual segmentations are used to train and test the GMM model: (1) intensity, (2) relative locations to lung landmarks, (3) gradients, and (4) second central moment. Briefly, intensity is selected because the intensities of tissues in CT are in the unified Hounsfield scale. The intensity information is used to tell the GMM model which intensity ranges that spleen should be in. However, other organs (e.g., kidney and liver) are typically share the similar intensity range as spleen. Therefore, other features are included to further distinguish spleen from others.

The organ location feature is derived by capturing the relative location between organ voxels to the automated derived lung landmarks following [18]. The spleen boundary feature is derived by the L2-Norm of 3D gradient, and the L2-Norm of second central moment features are also calculated. The four types of features are obtained respectively for five tissues (spleen, left kidney, fat, muscle, and other tissues in body) from the atlas labels. For a target image, the GMM model is applied to all voxels to probabilities of five tissues. Then the spleen probability map is derived by dividing the probabilities of spleen by the sum of the probabilities of all tissues. Finally, the spleen probability map is used as a prior in the context learning following [16, 27].

2.4 Experimental Design

In the first experiment, the performance of AGMMCL model on dataset with large variation of spleen sizes is evaluated using 30 scans from MeDiCAL and 19 scans from the clinically acquired splenomegaly cohort. We empirically use 500 cubic centimeter (cc) as a threshold to categorize the 49 scans into two groups, “large” (19 scans with spleen size >500 cc) and “small” (30 scans with spleen size $\leq$ 500 cc). Then, the 49 images are separated into 25 training scans and 24 testing scans.

The training set consists of 15 randomly selected scans from MeDiCAL dataset and 10 randomly selected scans from the clinically acquired splenomegaly cohort. The testing set is formed by the remaining scans. All training scans containing both large and small spleens are used to train a GMM model (GMM L&S) and the AGMM model. Meanwhile, 10 large spleen training scans are used to train a GMM-L while the 15 regular spleen training scans are used to train a GMM-S model. Lastly, the performance of direct atlas selection by 5 closest L (Selected by L) into fusion without context learning is also evaluated. All statistical significance tests are made using a Wilcoxon signed rank test ($p < 0.05$).

In the second experiment, the performance of AGMMCL model on dataset with small variation on regular spleen size is evaluated by using the remaining 20 scans from MeDiCAL dataset (not used in the first experiment). We train GMM (GMM-30) and AGMM models (AGMM-30) using the 30 MeDiCAL scans in the first experiment and test their performance using the remaining 20 images from MeDiCAL. Then, the 19 clinically

acquired splenomegaly cohort scans are added to the training set (totally 49 atlases) to verify the performance of AGMM model (AGMM-49).

3. RESULTS

Figure 3 shows the performance between the AGMM model with three GMM models as well as the simple atlas selection by L without context learning strategy. For each segmentation, Dice similarity coefficient (DSC), mean surface distance (MSD) and Hausdorff distance (HD) with manual segmentations are calculated to show the quantitative results. An example of qualitative result is shown in Figure 4. Figure 5 compared the performance between AGMM models with GMM model. The DSC and surface distance are shown in Figure 5 to compare the effect of AGMM method and traditional GMM method.

4. CONCLUSION AND DISCUSSION

This paper presented the semi-automated AGMMCL method to obtain spleen probabilistic maps adaptively for different target images. The cranio-caudal length, whose manual labelling costs only about 1 minute of rater time, much less than manual segmentation time (about 20 minutes per image), is used to select the atlases for the AGMM training. The performance of AGMM model was superior to the GMM model on the testing data with both regular and larger spleens (Figure 3), which makes AGMM capable of dealing with dataset that has great variations of spleen sizes.

It also achieved better performance than the GMM model on testing data with the regular size spleens (Figure 5). The future research interest on this topic could be to automate cranio-caudal length labelling via machine learning or other approaches, to further reduce the overall manual costs.

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.

References

1. Prassopoulos P, Daskalogiannaki M, Raissaki M, et al. Determination of normal splenic volume on computed tomography in relation to age, gender and body habitus. *Eur Radiol.* 1997; 7(2):246–8.
2. Henderson JM, Heymsfield SB, Horowitz J, et al. Measurement of liver and spleen volume by computed tomography. Assessment of reproducibility and changes found following a selective distal splenorenal shunt. *Radiology.* 1981; 141(2):525–527. [PubMed: 6974875]
3. Bolognesi M, Sacerdoti D, Bombonato G, et al. Change in portal flow after liver transplantation: effect on hepatic arterial resistance indices and role of spleen size. *Hepatology.* 2002; 35(3):601–8. [PubMed: 11870373]
4. Bezerra AS, D'Ippolito G, Faintuch S, et al. Determination of splenomegaly by CT: is there a place for a single measurement? *AJR Am J Roentgenol.* 2005; 184(5):1510–3. [PubMed: 15855107]

5. Linguraru MG, Sandberg JK, Jones EC, et al. Assessing splenomegaly: automated volumetric analysis of the spleen. *Acad Radiol.* 2013; 20(6):675–84. [PubMed: 23535191]
6. Yetter EM, Acosta KB, Olson MC, et al. Estimating splenic volume: sonographic measurements correlated with helical CT determination. *American Journal of Roentgenology.* 2003; 181(6):1615–1620. [PubMed: 14627584]
7. Lamb P, Lund A, Kanagasabay R, et al. 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]
8. Campadelli P, Casiraghi E, Pratisoli S. Fully automatic segmentation of abdominal organs from CT images using fast marching methods. :554–559.
9. Campadelli P, Casiraghi E, Pratisoli S. A segmentation framework for abdominal organs from CT scans. *Artif Intell Med.* 2010; 50(1):3–11. [PubMed: 20542673]
10. Campadelli P, Pratisoli S, Casiraghi E, et al. Automatic abdominal organ segmentation from CT images. *ELCVIA: electronic letters on computer vision and image analysis.* 2009; 8(1):001–14.
11. Chen X, Bagci U. 3D automatic anatomy segmentation based on iterative graph-cut-ASM. *Med Phys.* 2011; 38(8):4610–22. [PubMed: 21928634]
12. Chen X, Udupa JK, Bagci U, et al. Medical image segmentation by combining graph cuts and oriented active appearance models. *IEEE Trans Image Process.* 2012; 21(4):2035–46. [PubMed: 22311862]
13. Linguraru MG, Sandberg JK, Li Z, et al. Atlas-based automated segmentation of spleen and liver using adaptive enhancement estimation. :1001–1008.
14. Linguraru MG, Sandberg JK, Li Z, et al. 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]
15. Xu Z, Asman AJ, Shanahan PL, et al. SIMPLE is a good idea (and better with context learning). *Med Image Comput Comput Assist Interv.* 2014; 17(Pt 1):364–71. [PubMed: 25333139]
16. Xu Z, Burke RP, Lee CP, et al. Efficient multi-atlas abdominal segmentation on clinically acquired CT with SIMPLE context learning. *Med Image Anal.* 2015; 24(1):18–27. [PubMed: 26046403]
17. Xu Z, Li B, Panda S, et al. Shape-Constrained Multi-Atlas Segmentation of Spleen in CT. *Proc SPIE Int Soc Opt Eng.* 2014; 9034:903446. [PubMed: 24817808]
18. Burke RP, Xu Z, Lee CP, et al. Multi-Atlas Segmentation for Abdominal Organs with Gaussian Mixture Models. *Proc SPIE Int Soc Opt Eng.* 2015:9417.
19. Heinrich MP, Wilms M, Handels H. Multi-atlas Segmentation Using Patch-Based Joint Label Fusion with Non-Negative Least Squares Regression. :146–153.
20. 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]
21. Zhou Y, Bai J. Multiple abdominal organ segmentation: an atlas-based fuzzy connectedness approach. *IEEE Trans Inf Technol Biomed.* 2007; 11(3):348–52. [PubMed: 17521085]
22. McAuliffe MJ, Lalonde FM, McGarry D, et al. Medical image processing, analysis and visualization in clinical research. :381–386.
23. Wolz R, Chu C, Misawa K, et al. Automated abdominal multi-organ segmentation with subject-specific atlas generation. *IEEE Trans Med Imaging.* 2013; 32(9):1723–30. [PubMed: 23744670]
24. Li B, Panda S, Xu Z, et al. Regression forest region recognition enhances multi-atlas spleen labeling. :82–92.
25. Modat M, Ridgway GR, Taylor ZA, et al. Fast free-form deformation using graphics processing units. *Comput Methods Programs Biomed.* 2010; 98(3):278–284. [PubMed: 19818524]
26. Wang HZ, Suh JW, Das SR, et al. Multi-Atlas Segmentation with Joint Label Fusion. *Ieee Transactions on Pattern Analysis and Machine Intelligence.* 2013; 35(3):611–623. [PubMed: 22732662]
27. Xu Z, Burke RP, Lee CP, et al. Efficient Abdominal Segmentation on Clinically Acquired CT with SIMPLE Context Learning. *Proc SPIE Int Soc Opt Eng.* 2015:9413.

Figure 1.

Spleen size in this dataset varies from 100 cc to 9000 cc. (a) The spleens of relatively small sizes, further divided into S (size < 100 cc) and M (200 to 500 cc). (b) The spleens of relatively large sizes, further categorized into L (500 to 1000 cc), XL (1000 to 2000 cc) and XXL (size > 2000 cc).

Figure 2.

This figure shows the flow chart of the MAS pipeline with AGMM. The L value is first calculated for both atlas images and the target image, and then a subset of atlas images with closest L value is selected in AGMM training. Next the trained AGMM is employed in context learning to generate the learned prior map. After that, certain registered atlases are selected according to the map, and are combined with joint label fusion.

Figure 3.

This figure shows the quantitative evaluation of each segmentation under different GMM training methods, including GMM-L&S, GMM-L, GMM-S, and AGMM. (a) The Dice similarity coefficient (DSC). (b) The mean surface distance (MSD). (c) Hausdorff distance (HD). The “Ref.” indicates the AGMM 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.

Figure 4.

This figure shows a qualitative evaluation of segmentations using different GMM training methods, including “GMM-L&S”, “GMM-L”, “GMM-S”, “AGMM”, and the accurate manual segmentation.

Figure 5.

This figure show the performance of AGMM method on two different atlas datasets, compared to the performance of a published work under traditional GMM training method over the entire atlas set. (a) The DSC of GMM-30, AGMM-30 and AGMM-49. (b) The MSD of GMM-30, AGMM-30 and AGMM-49. (c) The Hausdorff Distance of GMM-30, AGMM-30 and AGMM-49. The “*” indicates the $p < 0.05$.