Hierarchical Multi–atlas Based Segmentation for Anatomical Structures: Evaluation in the VISCERAL Anatomy Benchmarks

Oscar Alfonso Jiménez del Toro and Henning Müller

University of Applied Sciences Western Switzerland (HES–SO), Switzerland
University and University Hospitals of Geneva, Switzerland

Abstract. Computer–based medical image analysis is often started with the exact location of anatomical structures in clinical scans. Many methods have already been proposed for segmenting single and multiple anatomical structures. However, it is uncommon to compare different approaches with the same testset, particularly a publicly available testset. The comparison of these methods allows to define objectively the advantages and limitations for each method. A hierarchical multi–atlas based segmentation approach is proposed for multiple anatomical structures in computed tomography scans. The method defines an anatomical hierarchy that exploits the inherent spatial and anatomical variability of medical images using image registration techniques. It was submitted and tested in the VISCERAL project Anatomy benchmarks. In this paper, the results are analyzed and compared to the results of the other segmentation methods submitted in the benchmark. Six out of the ten structures obtained the lowest average distance error and four had the best DICE overlap with the proposed method. Although the method was trained with a small trainingset it generated accurate output segmentations for liver, kidneys and other organs.

Keywords: Multi–atlas based segmentation, image registration, VISCERAL

1 Introduction

As part of their daily clinical workload, health providers visually inspect medical images to support a diagnostic hypothesis. This task has time limitations and it can not be scaled for big data repositories [6]. Without an objective interpretation of these medical images, more advanced research is limited to a small subset of patient cases. It is then fundamental to have a first automatic interpretation and selection of the images before performing more in–depth image analysis. Many computer–aided diagnosis techniques thus initially require the identification and segmentation of the anatomical structures.

There are already various methods that have been proposed for segmenting anatomical structures. Some approaches are based on shape modeling or random decision forests, while others combine intensity feature selection with dynamic programming [5, 1]. Even though some of the available approaches have

obtained high segmentation overlap, the results are frequently obtained in private datasets. Many of these methods are also targeted towards a single anatomical structure and their application for other organs is not straightforward. The Visual Concept Extraction Challenge in Radiology (VISCERAL[1]) benchmarks aim at evaluating the available state–of–the–art segmentation methods on a large public dataset [4]. The VISCERAL dataset includes multiple anatomical structures manually annotated by expert radiologists. These annotations were considered the 'gold truth' against which the methods output segmentations are compared. These include multiple organs (e.g. lungs, kidney, liver...) and relevant anatomical structures (e.g. some muscles and bones) in different imaging modalities such as computed tomography (CT) and magnetic resonance (MR).

A fully automatic segmentation approach for multiple anatomical structures is presented in this paper. It is implemented using multi–atlas registration in a hierarchical pipeline that includes *a priori* anatomical localization knowledge. It maximizes the information contained in each patient scan, exploiting the inherent anatomical variability and spatial distribution of the structures. The results from the method are presented and compared to those of the other segmentation methods submitted in the VISCERAL Anatomy Benchmarks.

2   Materials and Methods

The proposed method is divided into three steps: 1. Pre–processing, 2. Hierarchic anatomical registration and 3. Label fusion. Each step is detailed in the following paragraphs. The algorithm has already been described in [10].

2.1   Preprocessing

The test or target volumes were resampled to obtain isometric voxels in all dimensions. The volumes were then downscaled half their size to improve the algorithm speed and reduce the search area for the image registrations. The volumes were upscaled to their original size for the final label fusion step of the method.

2.2   Hierarchic Anatomical Registration

Multi–atlas Based Registration Multi–atlas based segmentation is an approach that estimates the unknown location of a structure in a target image using multiple reference atlases. Each atlas includes a patient volume and a label volume, created by manual annotation. The label volume had annotated the location of one or more structures in the patient volume. Parameterized image registration based on a cost function was used to increase the spatial relationship between the target and atlas volume. The cost function was optimized in each step using the adaptive stochastic gradient descent optimizer proposed in [2]

---

[1] http://www.visceral.eu/, as of 29 June 2014

with a multi–resolution approach. Normalized cross–correlation was used as a similarity metric for the cost function. The label volumes were then transformed taking the coordinate transformation obtained from the registration. The image registration was performed using the implementation and method of Elastix software[2] [3].

Both affine and non–rigid registrations were carried out for the final label estimation. Affine registration allowed scaling, rotation and translation. It was used as an initial alignment for both the global and local transformations. The non–rigid B–spline registrations defined the final output label and were used only in the local anatomical regions–of–interest (ROI) created for each structure.

Anatomical Hierarchy In [9] it was shown how some structures influence the distribution of the multiple organs in the thorax and abdomen. This anatomical property was exploited to create a hierarchical segmentation pipeline. The advantage of such a hierarchy is that the smaller and harder to segment structures can benefit from a previous initialization made with a bigger surrounding anatomical structure. It also maximizes the information contained in the training set, since each structure is registered within their own anatomical ROI. Therefore, the registrations are less affected by global inconsistencies between the atlas and target volumes.

A global affine registration is followed by individual affine registrations using local binary masks to enforce the spatial correlation of each anatomical structure separately. These masks are obtained from the morphological dilation of the output labels of the different atlases registered in the previous step. After each anatomical structure has its own independent ROI mask, the volumes are registered again but using a non–rigid B–spline transformation model. The new transformed label volumes for each structure constitute the individual votes that will be used for the label fusion step. The information regarding the hierarchical anatomical pipeline was previously described in [10]

2.3 Label Fusion

After the multi–atlas registration there is a label estimation for each of the atlases in the training set. The spatial information contained in each of the output estimations can be fused to obtain a more accurate output segmentation. Although there are many label fusion methods available, setting a majority voting threshold has proven to be an effective straightforward solution [7]. In this approach we selected five different thresholds for the allowed configurations in the VISCERAL Anatomy 1 Benchmark. The thresholds are implemented on a per–voxel basis and were optimized in the training set.

---

[2] Elastix: http://elastix.isi.uu.nl, 2014.[Online; accesed 27–April–2014].
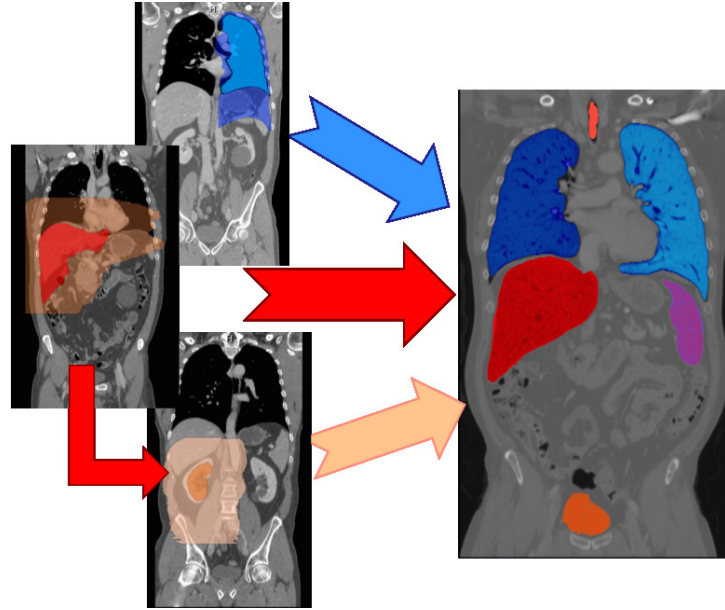
Fig. 1. Hierarchic anatomical registrations. The registrations of the bigger structures are used as a initialization for the smaller structures, which are harder to segment. Most of the registrations of structures like liver, lungs and urinary bladder will be reused in the method which makes it faster and more robust for the segmentation of structures like gallbladder and trachea.

3   Experimental Setup

3.1   Dataset

Seven volumes were provided for both unenhanced and contrast–enhanced computed tomography (ceCT) in the VISCERAL Anatomy[1] benchmark. For each scan there were up to 15 anatomical structures manually annotated by radiologists. These structures include organs like liver, lungs and kidneys, and bones or muscles like the first lumbar vertebra. The complete dataset structure list and detailed description of the medical images is presented in [8].

The VISCERAL Anatomy[1] benchmark allowed participants to select the tasks and structures in which they wish to participate. The proposed method was trained only with the 7 ceCT scans of the trunk for the first benchmark. Ten structures fully included in the field–of–view of these volumes were selected. The liver, right and left kidneys, right and left lungs, urinary bladder, spleen, trachea, first lumbar vertebra and gallbladder were segmented in ceCT scans with the submitted approach in Anatomy[1] benchmark.

3.2 Evaluation

The participants received a virtual cloud–computing 8–core CPU instance with 16 GB RAM. Both the executable and required libraries were installed by the participants in the provided virtual machines (VM). The organizers took over the VMs and ran the executables for the volumes in the testset for a limited time period. Once the evaluation phase had finished, the results were published in the VISCERAL project website (Anatomy[1] Benchmark Results [3]). The goal of this framework is to generate an objective and un–biased evaluation of the different algorithms with the same testset and computing capabilities for all the participants.

The testset from the Anatomy[1] benchmark included 12 ceCT volumes of the trunk. The results from the proposed method are presented and compared to those of other organ segmentation approaches submitted in the benchmark. For the Anatomy[1] benchmark, five other methods were submitted that segmented anatomical structures in ceCT scans. The methods proposed by Gass et al., Huang et al., Kechichian et al., Spanier et al. and Wang et al. segmented at least one anatomical structure in ceCT.

The DICE coefficient, adjusted rand index, interclass correlation and average distance error were computed and published for each anatomical structure contained in the testset.

Participant	Modality	Region	left kidney	right kidney	spleen	liver	left lung	right lung	urinary bladder	muscle body of left rectus abdominis	muscle body of right rectus abdominis	lumbar Vertebra 1	thyroid	pancreas	left psoas major muscle	right psoas major muscle	gallbladder	sternum	aorta	trachea	left adrenal gland	right adrenal gland
			29663	29662	86	58	1326	1302	237	40358	40357	29193	7578	170	32249	32248	187	2473	480	1247	30325	30324
Measure			DICE coefficient																			
SJ	Ctce	ThAb	0.631	0.663	0.690	0.747	0.848	0.975												0.785		
HJ	Ctce	ThAb				0.891																
W	Ctce	ThAb	0.804	0.872	0.873	0.898	0.965	0.969	0.805						0.792	0.811		0.713				
JM	Ctce	ThAb	*0.921*	*0.913*	0.852	*0.918*	0.955	0.965	0.700			0.522					*0.566*			0.836		
K	Ctce	ThAb	0.747	0.632	0.768	0.806	0.856	0.892	0.718	0.130	0.171	0.447	0.004	0.155	0.706	0.633	0.281	0.454	0.505	0.696	0.000	0.007
GG	Ctce	ThAb	0.903	0.877	0.802	0.899	0.961	0.968	0.676			0.604	0.252	0.465	0.811		0.334	0.595	0.785	0.847	0.204	0.164

Table 1. Best average DICE coefficient in ceCT testset of the VISCERAL Anatomy[1] benchmark.The proposed hierarchical multi–atlas based method (JM Jimenez del Toro et al. in grey) submitted output segmentations for ten anatomical structures. Highlighted in white are the best overlap obtained for 4 structures: right and left kidney, liver, gallbladder. The left lung and right lungs obtained also high overlap among the submitted methods. SJ Spanier et al., HJ Huang et al., W Wang et al., Kechichian et al. and GG Gass et al. (Benchmark Anatomy1 Results, http://www.visceral.eu/closed-benchmarks/benchmark-1/benchmark-1-results/, as of 29 June 2014).

4 Results

The results for the 10 anatomical structures had a total average DICE coefficient of 0.815 in ceCT. The DICE coefficients and average distance error tables for the ceCT testset are shown in table 1 and table 2 respectively. The ranking of the different methods participating in each of anatomical structure are presented in table 3. The hierarchical multi–atlas based segmentation approach obtained the best overlap in four clinically important anatomical structures: liver, right and left kidneys and gallbladder. Almost all the other segmented structures were in the top–three ranking among the submitted methods. The method also computed the smallest total average distance error with 1.11 average for the 10 structures. Seven of the submitted structures were in top position for this evaluation metric in VISCERAL Anatomy[1].

Participant	Modality	Region	left kidney	right kidney	spleen	liver	left lung	right lung	urinary bladder	muscle body of left rectus abdominis	muscle body of right rectus abdominis	lumbar Vertebra 1	thyroid	pancreas	left psoas major muscle	right psoas major muscle	gallbladder	sternum	aorta	trachea	left adrenal gland	right adrenal gland
			29663	29662	86	58	1326	1302	237	40358	40357	29193	7578	170	32249	32248	187	2473	480	1247	30325	30324
Measure			Average distance error																			
SJ	Ctce	ThAb	12.98	11.53	17.13	4.39	11.72	0.05												2.60		
HJ	Ctce	ThAb				0.74																
W	Ctce	ThAb	1.61	0.65	0.51	0.42	0.16	0.08	0.85													
JM	Ctce	ThAb	*0.15*	*0.24*	0.66	*0.30*	*0.09*	0.06	1.61			5.36					*2.36*		*0.30*			
K	Ctce	ThAb	2.82	6.13	1.95	1.96	1.18	0.96	1.74	17.77	16.10	5.50	19.69	24.19	2.51	4.38	9.49	6.63	5.88	2.34	28.97	8.45
GG	Ctce	ThAb	0.28	0.75	1.12	0.58	0.22	0.43	1.65			2.47	5.74	4.64	0.67		8.36	1.40	0.92	0.57	5.52	6.08

Table 2. Minimum average distance error in ceCT testset of VISCERAL Anatomy[1] benchmark. Highlighted in white are the lowest averaged distance errors obtained with the proposed algorithm (JM Jimenez del Toro et al. in grey). Six of the ten submitted structures (right and left kidney, liver, left lung, gallbladder and trachea) got the lowest distance error with a very low error also for right lung. SJ Spanier et al., HJ Huang et al., W Wang et al., Kechichian et al. and GG Gass et al. (Benchmark Anatomy1 Results, http://www.visceral.eu/closed-benchmarks/benchmark-1/benchmark-1-results/, as of 29 June 2014)

5 Discussion and Conclusions

The main contributions of the hierarchical multi–atlas based segmentation approach are:

- A robust automatic segmentation approach for multiple anatomical structures. The method was benchmarked against different segmentation approaches for single or multiple anatomical structures. It obtained the best

Anatomical structure	Rank Anatomy 1	Other participating methods
left kidney	*1st*	SJ, W, K, GG
right kidney	*1st*	SJ, W, K, GG
liver	*1st*	SJ,HJ, W, K, GG
gallbladder	*1st*	K, GG
trachea	2nd	SJ, K, GG
spleen	2nd	SJ, W, K, GG
1st lumbar vertebra	2nd	K, GG
left lung	3rd	SJ, W, K, GG
urinary bladder	3rd	W, K, GG
right lung	4th	SJ, W, K, GG

Table 3. Segmentation ranking in ceCT testset of VISCERAL Benchmark 1 Anatomy. According to the DICE overlap this is the ranking for the ten submitted structures. The other methods that had output segmentations for these structures are mentioned in the far–left tab: SJ Spanier et al., HJ Huang et al., W Wang et al., Kechichian et al. and GG Gass et al. (Benchmark Anatomy1 Results, http://www.visceral.eu/closed-benchmarks/benchmark-1/benchmark-1-results/, as of 29 June 2014)

overlap and smallest average distance error for most of the structures it segmented in the publicly available VISCERAL testset for ceCT scans.

– A hierarchy for anatomical structure segmentation was defined based on the organ size and tissue contrast. This hierarchy can be extended for more anatomical structures and implemented also with other medical imaging modalities such as MR.

– An efficient exploitation of a small training set based on the inherent anatomical variability of anatomical structureswa achieved. This allows enough flexibility to be quickly adapted for new images coming from scanners with different parameter tuning.

Although the overlap coefficient are consistent for most of the evaluated structures, the smaller and harder to segment structures (e.g. gallbladder) still needs to be improved to be used in a real clinical scenario. It is also not possible to define where and how significant the main errors in the output segmentations are. This information is kept by the benchmark organizers and the provided feedback is an overall interpretation of the results.

For future work the method will be extended to include all of the anatomical structures in the VISCERAL dataset. An evaluation of the method for the other modalities (MR and contrast–enhanced MR) is also foreseen with a much bigger testset.

6 Acknowledgments

References

1. Criminisi, A., Robertson, D., Konukoglu, E., Shotton, J., Pathak, S., White, S., Siddiqui, K.: Regression forests for efficient anatomy detection and localization in computed tomography scans. Medical Image Analysis 17(8), 1293–1303 (2013)
2. Klein, S., Pluim, J.P., Staring, M., Viergever, M.A.: Adaptive stochastic gradient descent optimisation for image registration. International Journal of Computer Vision 81(3), 227–239 (2009)
3. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.: Elastix: a toolbox for intensity–based medical image registration. IEEE Transactions on medical imaging 29(1), 196–205 (2010)
4. Langs, G., Müller, H., Menze, B.H., Hanbury, A.: Visceral: Towards large data in medical imaging – challenges and directions. Lecture Notes in Computer Science 7723, 92–98 (2013)
5. Linguraru, M.G., Sandberg, J.K., Li, Z., Shah, F., Summers, R.M.: Automated segmentation and quantification of liver and spleen from CT images using normalized probabilistic atlases and enhancement estimation. Medical Physics 37(2), 771–783 (2010)
6. Müller, H., Deserno, T.M.: Content-based retrieval. Springer (2010)
7. Rohlfing, T., Brandt, R., Menzel, R., Maurer Jr., C.R.: Evaluation of atlas selection strategies for atlas–based image segmentation with application to confocal microscopy images of bee brains. Neuroimage 23(8), 983–994 (April 2004)
8. Jiménez del Toro, O.A., Goksel, O., Menze, B., Müller, H., Langs, G., Weber, M.A., Eggel, I., Gruenberg, K., Holzer, M., Kotsios-Kontokotsios, G., Krenn, M., Schaer, R., Taha, A.A., Winterstein, M., Hanbury, A.: VISCERAL – VISual Concept Extraction challenge in RAdioLogy: ISBI 2014 challenge organization. In: Goksel, O. (ed.) Proceedings of the VISCERAL Challenge at ISBI. pp. 6–15. No. 1194 in CEUR Workshop Proceedings, Beijing, China (May 2014)
9. Jiménez del Toro, O.A., Müller, H.: Multi–structure atlas–based segmentation using anatomical regions of interest. In: MICCAI workshop on Medical Computer Vision. Lecture Notes in Computer Science, Springer (2013)
10. Jiménez del Toro, O.A., Müller, H.: Hierarchical multi–structure segmentation guided by anatomical correlations. In: Goksel, O. (ed.) Proceedings of the VISCERAL Challenge at ISBI. pp. 32–36. CEUR Workshop Proceedings, Beijing, China (May 2014)