VISCERAL@ISBI 2014

VISCERAL Organ Segmentation and Landmark Detection Challenge

at IEEE International Symposium on Biomedical Imaging 2014

Beijing, China, May 1, 2014

Proceedings

Orcun Goksel (Ed.)

Editor's address:
Prof. Dr. Orcun Goksel
Swiss Federal Institute of Technology (ETH) Zürich
Computer Vision Laboratory
Sternwartstrasse 7
8092 Zürich, Switzerland

ogoksel@ethz.ch

Preface

VISCERAL (Visual Concept Extraction Challenge in Radiology) aims to organize series of benchmarks on the processing of large-scale 3D radiology images, by using an innovative cloud-based evaluation approach.

While a growing number of benchmark studies compare the performance of algorithms for automated organ segmentation in images with restricted field of views, emphasis on anatomical segmentation and landmark localization in images with wide field-of-view (e.g. showing entire abdomen, trunk, or the whole body) has been limited. VISCERAL Anatomy^2 benchmark series, namely *Organ Segmentation and Landmark Detection Benchmarks*, aim to address this need. This ISBI VISCERAL Challenge, a part of Anatomy^2 series, has been organized on May 1^{st} 2014, within the IEEE International Symposium on Biomedical Imaging (ISBI) in Beijing, China.

The challenge participants have submitted segmentation and localization results two weeks before the challenge session, that were evaluated against test data by the organizers with results presented during the challenge session. Each participant presented his method in a 15 minute oral session during the challenge session. Participants also submitted short papers summarizing their specific methodologies that were used to generate their results.

This volume contains two parts. The first part consist of one paper authored by the organizers of the challenge, and the second part presents a compilation of the submissions by the challenge participants. We thank the authors for their submissions and the program committee for their hard work.

Orcun Goksel
On behalf of *VISCERAL Consortium*

Session Chairs

Orçun Göksel, Swiss Federal Institute of Technology (ETH) Zürich, Switzerland
Bjoern Menze, Munich University of Technology (TUM), Germany

VISCERAL Consortium

Allan Hanbury, Vienna University of Technology, Austria (coordinator)
Henning Müller, University of Applied Sciences Western Switzerland, Switzerland
Georg Langs, Medical University of Vienna, Austria
Orçun Göksel, ETH Zürich, Switzerland
Marc-André Weber, University of Heidelberg, Germany
Tomàs Salas Fernandez, Catalan Agency for Health Information, Assessment and Quality, Spain

Contributing VISCERAL Team Members

Ivan Eggel, University of Applied Sciences Western Switzerland, Switzerland
Katharina Grünberg, University of Heidelberg, Germany
Markus Holzer, Medical University of Vienna, Austria
András Jakab, Medical University of Vienna, Austria
Oscar Jiménez, University of Applied Sciences Western Switzerland, Switzerland
Georgios Kontokotsios, Vienna University of Technology, Austria
Markus Krenn, Medical University of Vienna, Austria
Roger Schaer, University of Applied Sciences Western Switzerland, Switzerland
Abdel Aziz Taha, Vienna University of Technology, Austria
Marianne Winterstein, University of Heidelberg, Germany

Contents

PART I: Organization and Evaluation

VISCERAL – VISual Concept Extraction challenge in RAdioLogy: ISBI 2014 Challenge Organization
Oscar Alfonso Jiménez del Toro, Orcun Goksel, Bjoern Menze, Henning Müller, Georg Langs, Marc-André Weber, Ivan Eggel, Katharina Gruenberg, Markus Holzer, András Jakab, Georgios Kontokotsios, Markus Krenn, Tomàs Salas Fernandez, Roger Schaer, Abdel Aziz Taha, Marianne Winterstein, Allan Hanbury **6**

PART II: Participant Submissions

Rule-Based Ventral Cavity Multi-Organ Automatic Segmentation in CT Scans
Assaf B. Spanier, Leo Joskowicz **16**

Automatic Liver Segmentation Using Multiple Prior Knowledge Models and Free-Form Deformation
Cheng Huang, Xuhui Li, Fucang Jia **22**

Automatic Multi-Organ Segmentation Using Fast Model Based Level Set Method and Hierarchical Shape Priors
Chunliang Wang, Örjan Smedby **25**

Hierarchical Multi-structure Segmentation Guided by Anatomical Correlations
Oscar Alfonso Jiménez del Toro, Henning Müller **32**

Segmentation and Landmark Localization Based on Multiple Atlases
Orcun Goksel, Tobias Gass, Gabor Szekely **37**

VISCERAL — VISual Concept Extraction challenge in RAdioLogy: ISBI 2014 Challenge Organization

Oscar Alfonso Jiménez del Toro[1], Orcun Goksel[2], Bjoern Menze[2], Henning Müller[1]
Georg Langs[3], Marc–André Weber[4], Ivan Eggel[1], Katharina Gruenberg[4],
Markus Holzer[3], András Jakab[3], Georgios Kontokotsios[5], Markus Krenn[3],
Tomàs Salas Fernandez[6], Roger Schaer[1], Abdel Aziz Taha[5],
Marianne Winterstein[4], Allan Hanbury[5]

University of Applied Sciences Western Switzerland, Switzerland[1]
Swiss Federal Institute of Technology (ETH) Zürich, Switzerland[2]
Medical University of Vienna, Austria[3]
University of Heidelberg, Germany[4]
Vienna University of Technology, Austria[5]
Catalan Agency for Health Information, Assessment and Quality, Spain[6]

Abstract

The VISual Concept Extraction challenge in RAdioLogy (VISCERAL) project has been developed as a cloud–based infrastructure for the evaluation of medical image data in large data sets. As part of this project, the ISBI 2014 (International Symposium for Biomedical Imaging) challenge was organized using the VISCERAL data set and shared cloud–framework. Two tasks were selected to exploit and compare multiple state–of–the–art solutions designed for big data medical image analysis. Segmentation and landmark localization results from the submitted algorithms were compared to manually annotated ground truth in the VISCERAL data set. This paper presents an overview of the challenge setup and data set used as well as the evaluation metrics from the various results submitted to the challenge. The participants presented their algorithms during an organized session at ISBI 2014. There were lively discussions in which the importance of comparing approaches on tasks sharing a common data set was highlighted.

In: O. Goksel (ed.): Proceedings of the VISCERAL Organ Segmentation and Landmark Detection Benchmark at the 2014 IEEE International Symposium on Biomedical Imaging (ISBI), Beijing, China, May 1st, 2014
published at http://ceur-ws.org

1 Introduction

Computational approaches that can be scaled to large amounts of medical data are needed to tackle the ever–growing data resources obtained daily from the hospitals [Doi05]. Handling this enormous amount of medical data during clinical routine by health professionals has complexity and scaling limitations. It is also very time–consuming, and hence requires unsupervised and automatic methods to perform the necessary data analysis and processing for data interpretation. There are already many algorithms and techniques for big data analysis, however, most research groups do not have access to large-scale annotated medical data to develop such approaches for medical images. Distributing these big data sets (on the order of terabytes) requires efficient and scalable storing and computing capabilities. Evaluation campaigns and benchmarks can objectively compare multiple state–of–the art algorithms to determine the optimal solution for a certain clinical task [HMLM14, GSdHKCDF$^+$13].

The Visual Concept Extraction Challenge in Radiology (VISCERAL) project was developed as a cloud–based infrastructure for the evaluation of medical image analysis techniques on large data sets [LMMH13]. The shared cloud environment in which the VISCERAL project takes place allows access and processing of these data without having to duplicate the data or move it to participants' side. Since the data are stored centrally, and not distributed outside the cloud environment, the legal and ethical requirements of such data sets can also be satisfied, so also confidential data sets can be benchmarked in this way as only a small training data set can be accessed by participants [EILI$^+$10]. The cloud infrastructure is provided and funded by the VISCERAL project. The participants are provided with computationally powerful virtual machines that can be accessed remotely in the shared cloud infrastructure while working on the training data and tuning their algorithms. Participant access is withdrawn during the evaluation phase and only the organizers access the machines. The algorithms are brought to the data to perform automated processing and data mining. The evaluation of the performance of these methods can therefore be done with real clinical imaging data and the outcomes can be reused to improve the methods.

The whole body 3D medical imaging data including manual labels that is provided by VISCERAL includes a small subset with ground truth annotated by experienced radiologists. Through evaluation campaigns, challenges, benchmarks and competitions, tasks of general interest can be selected to compare the algorithms on a large scale. This manually annotated gold corpus can be used to identify high quality methods that can also be combined to create a much larger "reasonably annotated" data set, satisfactory but perhaps not as reliable as manual annotation. Using fusion techniques this silver corpus will be created with the agreement between the segmentations of the algorithms on a large–scale data set. This maximizes the gain of manual annotation and also identifies strong differences between participating systems on the annotated organs.

2 ISBI Challenge Framework

The registration procedure for the ISBI challenge was that of the VISCERAL series benchmark that includes several campaigns. The participants filled information details and uploaded a signed participation agreement form, which corresponds to ethics requests for usage of the data. Since the VISCERAL data set is stored on the Azure Cloud, each participant then received access to an Azure virtual cloud–computing instance. There were 5 operating systems available to choose from including Windows 2012, Windows 2008, Ubuntu Server 14.04 LTS, openSUSE 13.1 and CentOS 6.5. All cloud– computing instances have an 8–core CPU with 16 GB RAM to provide the same computing capabilities to different solutions proposed. The participant gets administrator rights on their virtual machine (VM) and can access remotely to deploy their algorithms and add any supporting library/applications to their VM. The VISCERAL training data set can then be accessed

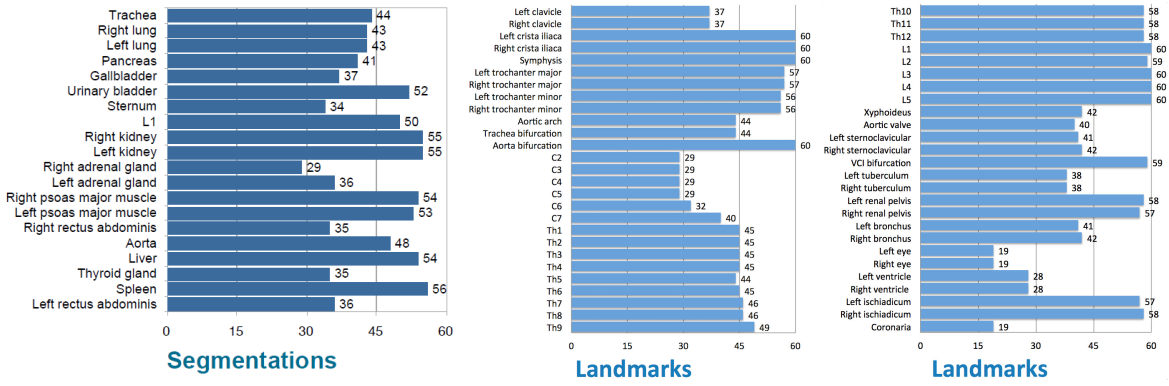

Figure 1: The ISBI training set.

and downloaded securely within the VMs through secured URL links.

2.1 Data Set

The medical images contained in the VISCERAL data set have been acquired during daily clinical routine work. Data sets of children (<18 years) were not included based on the recommendations of the ethical committee. In the provided data sets multiple organs are visible and depicted in a resolution sufficient to reliably detect an organ and delineate its borders. This is to enforce that a large number of organs and structures can be segmented in one data set. The data set consists of computed tomography (CT) scans and magnetic resonance (MR) imaging with and without contrast enhancement to evaluate the participants algorithms on several modalities, contrasts and MR sequence directions, making sure that algorithms are not optimized for one specific machine or protocol.

The available training set from VISCERAL Anatomy[2] benchmark was used by the participants of the ISBI VISCERAL challenge. The contents of this dataset are elaborated below.

2.1.1 CT Scans

There are 15 unenhanced whole–body CT volumes acquired from patients with bone marrow neoplasms, such as multiple myeloma, to detect osteolysis. The field–of–view spans from and including the head to the knee (see Fig. 2, A). The in–plane resolution ranges between 0.977/0.977 to 1.405/1.405 mm, and the in–between plane resolution is 3 mm or higher.

15 contrast–enhanced CT scans of the trunk that have been acquired in patients with malignant lymphoma are also included. They have a large field–of–view from the corpus mandibulae to the lower part of the pelvis (see Fig. 2, B). They have an in–plane resolution of between 0.604/0.604 and 0.793/0.793 mm, and an in–between plane resolution of at least 3 mm or higher.

2.1.2 MR Scans

15 whole–body MR scans in two sequences (30 in total) are also part of the training set. They were acquired in patients with multiple myeloma to detect focal and/or diffuse bone marrow infiltration. Both a coronal T1–weighted and fat-suppressed T2–weighted or STIR (short tau inversion recovery) sequence of the whole body are available for each of the 15 patients. The field–of–view starts and includes the head and ends at the feet (see Fig. 2, C). The in–plane resolution is 1.250/1.250 mm, and the in–between plane resolution is 5 mm.

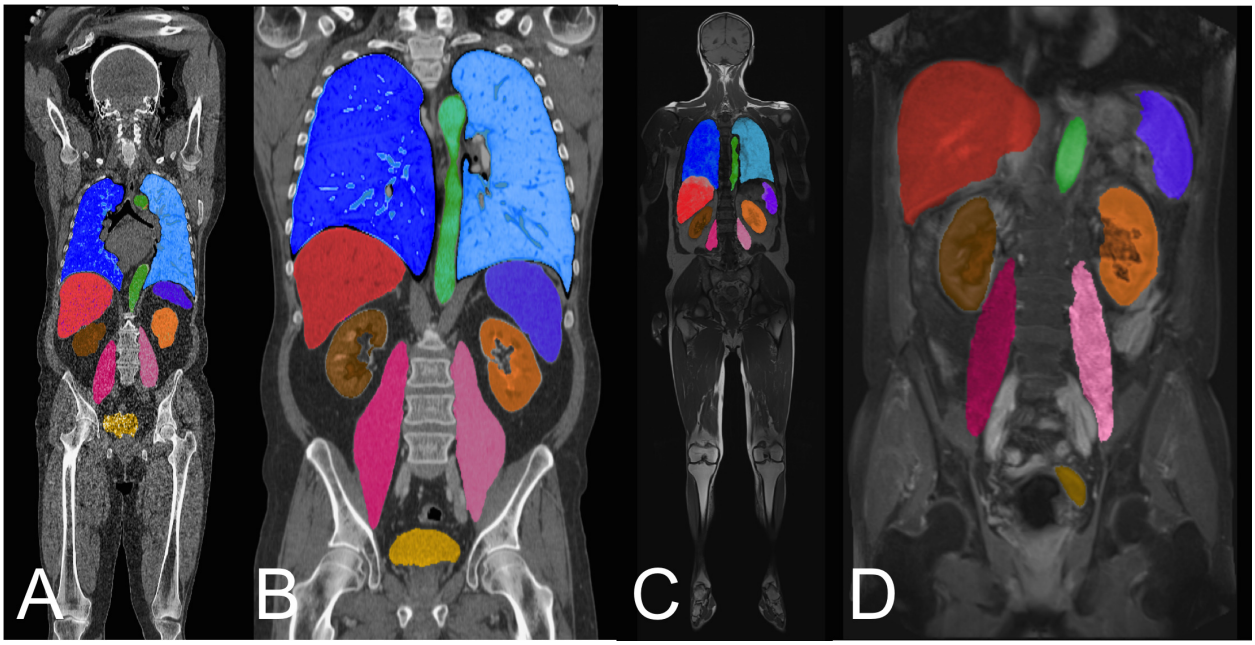

Figure 2: Sample data set volumes. A) Whole–body unenhanced CT; B) contrast–enhanced CT of the trunk; C) whole–body unenhanced MR; D) contrast–enhanced MR of the abdomen

To improve the segmentation of smaller organs (such as the adrenal glands), 15 T1 contrast–enhanced fat saturated MR scans of the abdomen are also included. They were acquired in oncological patients with likely metastases within the abdomen. The field–of–view starts at the top of the diaphragm and extends to the lower part of the pelvis (see Fig. 2, D). They have an in plane resolution of between 0.840/0.804 to 1.302/1.302 mm, and an in–between plane resolution of 3 mm.

2.1.3 Annotated Structures and Landmarks

There are in total 60 manually annotated volumes in this ISBI challenge training set. The available data contains segmentation and landmarks of several different anatomical structures in different imaging modalities, e.g. CT and MRI.

The two categories of annotations and results are:

- Region segmentations: These regions correspond to anatomical structures (e.g. right lung), or sub–parts in volume data. The 20 anatomical structures that make up the training set are: trachea, left/right lungs, sternum, vertebra L1, left/right kidneys, left/right adrenal glands, left/right psoas major muscles, left/right rectus abdominis, thyroid gland, liver, spleen, gallbladder, pancreas, urinary bladder and aorta. Not all structures are visible or within the field–of–view in the images, therefore leading to varying numbers of annotations per structure (see Fig. 1 for a detailed break–down).

- Landmarks: Anatomical landmarks are the locations of selected anatomical structures that should be identifiable in the different image sequences available in the data set. There can be up to 53 anatomical landmarks (see Fig. 1) located in the data set volumes: left/right clavicles, left/right crista iliaca, symphysis, left/right trochanter major, left/right trochanter minor, aortic arch, trachea bifurcation, aorta bifurcation, vertebrae C2-C7, Th1-Th12, L1-L5, xyphoideus, aortic valve, left/right sternoclavicular, VCI bifurcation, left/right tuberculums,

left/right renal pelvises, left/right bronchus, left/right eyes, left/right ventricles, left/right ischiadicum and coronaria.

In total the 60 training set volumes containing 890 manually segmented anatomical structures and 2420 manually located anatomical landmarks make up the training set. Some of the anatomical structures in the volumes were not segmented if the annotators considered there was insufficient tissue contrast to perform the segmentation or to locate the landmark. Other structures are missing or not included in the training set because of anatomical variations (e.g. missing kidney) or radiologic pathological signs (e.g. aortic aneurysm). Landmarks are easy and quick to annotate whereas precise organ segmentation is time–consuming even when using automatic tools.

2.1.4 Test Set

The test set contains 20 manually annotated volumes. Each modality (whole–body CT, thorax and abdomen contrast–enhanced CT, whole–body MR and abdomen contrast enhanced MR) is represented by 5 volumes. The anatomical structures and landmarks contained in the selected volumes were used to evaluate the participants' algorithms.

2.2 ISBI VISCERAL Challenge Submission

The participants can select the structures and modalities in which they choose to participate. The outputs are therefore evaluated per structure and per modality. The evaluation of the ISBI challenge has been organized differently than the general VISCERAL evaluation framework to allow for the evaluation results to complete in the given relatively short time–frame. For this challenge, the test set volumes were made available in the cloud some weeks ahead of the challenge. The participants themselves computed the annotations (segmentations and/or landmark locations) in their VMs and stored them on their VM storage. The files could then be submitted within their VM through an uploading script provided to the participants. The script stored their output files in a corresponding cloud container created for the challenge individual for each participant. A list containing the available ground truth segmentations of the test set filtered duplicates or output files with incorrect file names. It also ensured all files were coherent with the participant ID list from the organizers.

2.3 Evaluation Software

To evaluate the output segmentations and landmark locations against the ground truth, the VISCERAL evaluation tool was used. This software was also included in the VM assigned to each participant. This evaluation tool has different evaluation metrics implemented such as (1) distance–based metrics, (2) spatial overlap metrics and (3) probabilistic and information theoretic metrics. The most suitable subset of the metrics was used in the analysis of the results and all metrics were made available to the participants. For the output segmentations of the ISBI challenge the following evaluation metrics were selected:

- DICE coefficient [ZWB+04]
- Adjusted Rand Index [VPYM11]
- Interclass Correlation [GJC01]
- Average distance [KCAB09]

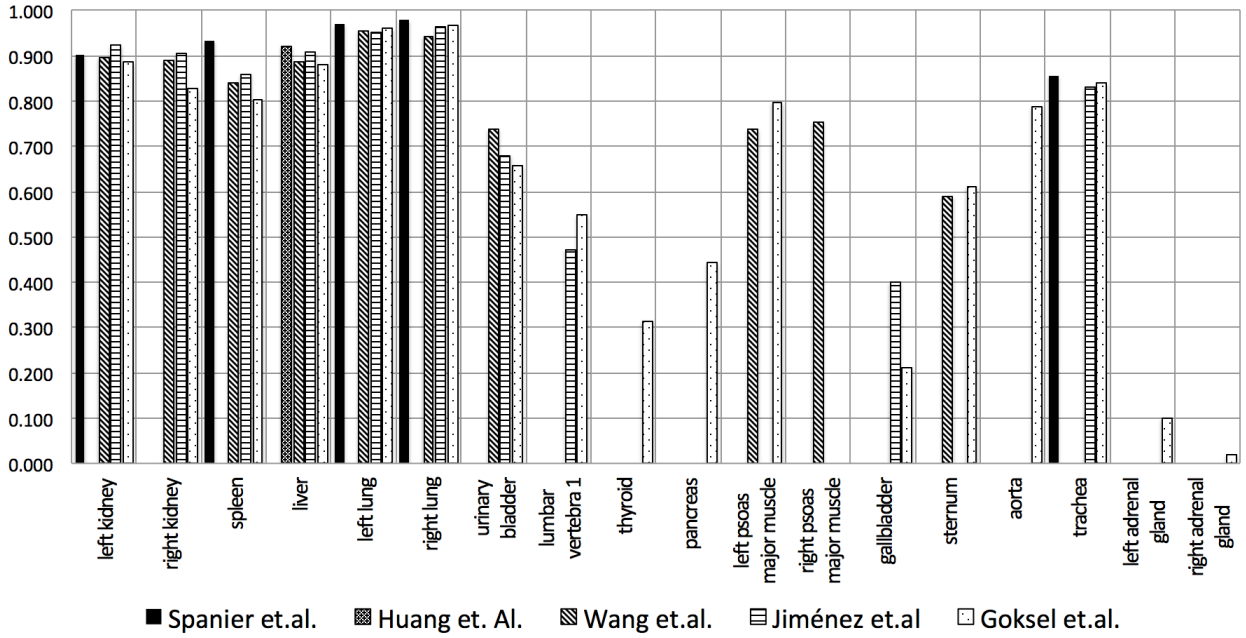

Figure 3: Anatomical structure segmentation task: DICE coefficient results. Contrast–enhanced CT scans of the Thorax and Abdomen.

Only one label is considered per image. The voxel value can be either zero (background) or one for the voxels containing the segmentation. A threshold is set at 0.5 to create binary images in case the output label has a fuzzy membership or a probability map.

For the landmark localization evaluation the same VISCERAL tool measures the landmark–specific average error (Euclidean distance) error between all the results and the manually located landmarks. The percentage of detected landmarks per volume (i.e. landmarks detected / landmarks in the volume) is also computed.

2.4 Participation

The ISBI training and test set volumes were made available through the Azure cloud framework for all the registered participants of the VISCERAL Anatomy[2] benchmark. In total 18 groups got access to the challenge training set and the 60 training volumes of the data set. The research groups that submitted working virtual machines had a chance to present their methods and results at the "VISCERAL Organ Segmentation and Landmark Detection Challenge" at the 2014 IEEE International Symposium on Biomedical Imaging (ISBI).

A single–blind review process was applied to the initial abstract submissions. The accepted abstracts were then invited to submit a short paper presenting their methods and results in the challenge. There were 5 high–quality submissions accepted and included in these proceedings.

Spanier et al. [SJ14] submitted segmentations for five organs in CT contrast–enhanced volumes. Their multi–step algorithm combines thresholding and region growing techniques to segment each organ individually. It starts with the location of a region of interest and identification of the largest axial cross–section slices of the selected structure. It then improves the initial segmentation with morphological operators and a final step performs 3D region growing.

Huang et al. [HLJ14] proposed a coarse–to–fine liver segmentation using prior models for the shape, profile appearance and contextual information of the liver. An AdaBoost voxel–based classifier creates a liver probability map that is refined in the last step with free–form deformation with

Table 1: Anatomical structure segmentation task: DICE coefficient results table. Contrast–enhanced and unenhanced CT scans submissions.

Participant	Modality	Region	left kidney	right kidney	spleen	liver	left lung	right lung	urinary bladder	lumbar Vertebra 1	thyroid	pancreas	left psoas major muscle	right psoas major muscle	gallbladder	sternum	aorta	trachea	left adrenal gland	right adrenal gland	Average
			29663	**29662**	**86**	**58**	**1326**	**1302**	**237**	**29193**	**7578**	**170**	**32249**	**32248**	**187**	**2473**	**480**	**1247**	**30325**	**30324**	
	Measure		DICE coefficient																		
Spanier et.al.	Ctce	ThAb	0.902		0.934		0.970	0.979										0.856			**0.928**
Huang et.al.	Ctce	ThAb				0.922															**0.922**
	CT	wb				0.911															**0.911**
Wang et.al.	CT	wb	0.729	0.777	0.887	0.904	0.971	0.972	0.806				0.722	0.764		0.712					**0.824**
	Ctce	ThAb	0.896	0.890	0.842	0.887	0.956	0.942	0.738				0.737	0.752		0.590					**0.823**
Jiménez et.al.	CT	wb	0.678	0.649	0.677	0.823	0.969	0.967	0.616	0.440					0.271			0.855			**0.694**
	Ctce	ThAb	0.923	0.905	0.859	0.908	0.952	0.963	0.680	0.472					0.400			0.830			**0.789**
Goksel et.al.	Ctce	ThAb	0.885	0.827	0.803	0.882	0.960	0.966	0.657	0.548	0.315	0.442	0.797		0.212	0.612	0.787	0.839	0.099	0.019	**0.627**
	CT	wb	0.756	0.679	0.684	0.798	0.955	0.965	0.636	0.333	0.439	0.466	0.773	0.780	0.078	0.630	0.724	0.837	0.282	0.133	**0.608**

a gradient appearance model.

Wang et al. [WS14] segmented 10 anatomical structures in CT contrast–enhanced and unenhanced scans. Their multi–organ segmentation pipeline performs in a top–down approach by a model–based level–set segmentation of the ventral cavity. After dividing the cavity in thoracic and abdominopelvic cavity, the major structures are segmented and their location information is passed to the lower–level structures.

Jiménez del Toro et al. [JdTM14] segmented structures in CT and contrast–enhanced CT scans with a hierarchical multi–atlas approach. Based on the spatial anatomical correlations between the organs, the bigger and high–contrasted organs are first segmented. These then define the initial volume transformations for the smaller structures with less defined boundaries.

Goksel et al. [GGS14] submitted segmentations for both CT and MR anatomical structure segmentation. They also submitted results for the landmark localization task. For the segmentations they use a multi–atlas based technique that implements Markov Random Fields to guide the registrations. A multi–atlas template–based approach fuses the different estimations to detect the landmarks.

3 Results

There were approximately 500 structure segmentations and 211 landmark locations submitted to the VISCERAL ISBI challenge. Four participants submitted results for the segmentation tasks in multiple organs using whole–body CT or contrast–enhanced scans with results presented in Table 1 and Fig. 3. There was one participant that contributed segmentations on both the whole–body MR scans and the contrast–enhanced MR abdomen volumes with results presented in Table 3. Only one participant submitted landmark localization results, with Table 4 showing their evaluation results.

4 Conclusions and Future Work

The VISCERAL project has the evaluation of algorithms on large data sets as its main objective. The proposed VISCERAL infrastructure allows evaluations with private or restricted data, such as electronic health records, without the participants access to the test data by using a fully cloud–based approach. This infrastructure also avoid moving data, which is potentially hard for very large data sets. The algorithms are brought to the data and not the data to the algorithms.

Table 2: Anatomical structure segmentation task: Average Distance results table. Contrast–enhanced and unenhanced CT scans submissions.

Participant	Modality	Region	left kidney	right kidney	spleen	liver	left lung	right lung	urinary bladder	lumbar Vertebra 1	thyroid	pancreas	left psoas major muscle	right psoas major muscle	gallbladder	sternum	aorta	trachea	left adrenal gland	right adrenal gland	Average
			29663	**29662**	**86**	**58**	**1326**	**1302**	**237**	**29193**	**7578**	**170**	**32249**	**32248**	**187**	**2473**	**480**	**1247**	**30325**	**30324**	
	Measure		Average distance																		
Spanier et.al.	Ctce	ThAb	0.23		0.14		0.08	0.04										0.30			**0.16**
Huang et.al.	Ctce	ThAb				0.29															**0.29**
	CT	wb				0.31															**0.31**
Wang et.al.	CT	wb	3.63	1.21	0.45	0.46	0.07	0.06	0.78				1.77	1.15		1.24					**1.08**
	Ctce	ThAb	0.27	0.28	0.87	0.65	0.15	0.20	1.59				1.75	1.58		4.14					**1.15**
Jiménez et.al.	CT	wb	3.06	3.31	2.64	1.76	0.05	0.06	2.08	4.18					25.02			0.30			**4.24**
	Ctce	ThAb	0.15	0.37	0.60	0.37	0.11	0.07	1.69	4.22					4.34			0.58			**1.25**
Goksel et.al.	Ctce	ThAb	0.39	1.48	1.08	0.73	0.38	0.92	1.93	2.46	5.54	6.83	0.80		13.50	1.13	0.78	0.71	9.63	15.07	**3.73**
	MRT1	wb	21.23	16.73	0.83	0.90	159.38	0.28	1.44	1.21	1.01		0.94		8.18	5.44	0.45	0.53	12.09	3.64	**14.64**
	CT	wb	1.42	3.66	2.15	1.60	0.88	0.10	2.32	6.99	3.57	3.36	1.33	1.33	14.55	1.33	1.02	0.87	6.01	6.62	**3.28**
	MRT1cefs	Ab	0.34	3.76	0.82	0.83			5.91	9.14		5.91	1.63		11.54						**4.43**

Table 3: Evaluation metrics for the MR scan submissions.

Participant	Modality	Region	left kidney	right kidney	spleen	liver	left lung	right lung	urinary bladder	lumbar Vertebra 1	thyroid	pancreas	left psoas major muscle	right psoas major muscle	gallbladder	sternum	aorta	trachea	left adrenal gland	right adrenal gland	Average
			29663	**29662**	**86**	**58**	**1326**	**1302**	**237**	**29193**	**7578**	**170**	**32249**	**32248**	**187**	**2473**	**480**	**1247**	**30325**	**30324**	
	Measure		DICE coefficient																		
Goksel et.al.	MRT1	wb	0.548	0.589	0.646	0.817	0.486	0.909	0.577	0.623	0.488		0.765		0.044	0.359	0.783	0.747	0.144	0.268	**0.550**
	MRT1cefs	Ab	0.888	0.732	0.785	0.861			0.334	0.084		0.356	0.654		0.000						**0.587**
	Measure		Average distance																		
Goksel et.al.	MRT1	wb	21.23	16.73	0.83	0.90	159.4	0.28	1.44	1.21	1.01		0.94		8.18	5.44	0.45	0.53	12.09	3.64	**14.64**
	MRT1cefs	Ab	0.34	3.76	0.82	0.83			5.91	9.14		5.91	1.63		11.54						**4.43**

Both gold corpus and silver corpus will be available as a resource to the community. The ISBI test set volumes and annotations are now available and are part of the VISCERAL Anatomy[2] benchmark training set.

So far, both past VISCERAL anatomy benchmarks have addressed organ segmentation and landmark localization tasks. There are two more benchmarks under development in the VISCERAL project, a retrieval benchmark and a detection benchmark. The retrieval benchmark will be the retrieval of similar cases based on both visual information and radiology reports. The detection benchmark will focus in the detection of lesions in MR and CT images.

In the future, the automation of the evaluation process is intended to reduce the need for intervention from the organizers to a minimum and to provide faster evaluation feedback to the participants. The participants will then be able to submit their algorithms through the cloud virtual machines and obtain the calculated metrics directly from the system. Such a system could then store the results from all the algorithms submitted and perform an objective comparison with state-of-the art algorithms. Through the involvement of the research community, the VISCERAL framework could produce novel tools for the clinical work flow that has substantial impact on diagnosis quality

Table 4: Landmarks results.

Participant	Modality	Region		aorta_bifurcation	aortic_arch	clavicle_left	clavicle_right	crista_iliaca_left	crista_iliaca_right	symphysis	trachea_bifurcation	trochanter_major_left	trochanter_major_right	trochanter_minor_left	trochanter_minor_right	Average
			Measure	Euclidean distance												
Goksel et al.	CT	wb	average	19.05	17.68	9.27	5.69	7.70	6.12	8.01	3.99	34.37	36.18	5.16	4.06	**13.11**
	Ctce	ThAb	average	36.22	16.18	16.26	32.35	13.93	10.38	15.59	3.35	37.84	38.31	11.22	12.64	**20.36**
	MRT1	wb	average	252.49	43.67	13.05	23.31	23.29	19.21	122.45	61.20	29.57	44.40	18.51	62.40	**59.46**
	MRT1cefs	Ab	average	61.28				88.92	57.65	50.86		30.49	59.81	28.54	34.84	**51.55**

and treatment success. Having all tools and algorithms in the same cloud environment can also help us to combine tools and approaches with very little additional effort, which expectedly yields better results.

5 Acknowledgments

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 318068 VISCERAL. We would also like to thank Microsoft research for their financial and information support in using the Azure cloud for the benchmark.

References

[Doi05] K Doi. Current status and future potential of computer–aided diagnosis in medical imaging. *British Journal of Radiology*, 78:3–19, 2005.

[EILI+10] Bernice Elger, Jimison Iavindrasana, Luigi Lo Iacono, Henning Müller, Nicolas Roduit, Paul Summers, and Jessica Wright. Strategies for health data exchange for secondary, cross–institutional clinical research. *Computer Methods and Programs in Biomedicine*, 99(3):230–251, September 2010.

[GGS14] Orcun Goksel, Tobias Gass, and Gabor Szekely. Segmentation and landmark localization based on multiple atlases. In Orcun Goksel, editor, *Proceedings of the VISCERAL Challenge at ISBI*, CEUR Workshop Proceedings, pages 37–43, Beijing, China, May 2014.

[GJC01] Guido Gerig, Matthieu Jomier, and Miranda Chakos. A new validation tool for assessing and improving 3D object segmentation. In Wiro J. Niessen and Max A. Viergever, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2001*, volume 2208 of *Lecture Notes in Computer Science*, pages 516–523. Springer Berlin Heidelberg, 2001.

[GSdHKCDF+13] Alba García Seco de Herrera, Jayashree Kalpathy-Cramer, Dina Demner Fushman, Sameer Antani, and Henning Müller. Overview of the ImageCLEF 2013 medical tasks. In *Working Notes of CLEF 2013 (Cross Language Evaluation Forum)*, September 2013.

[HLJ14] Cheng Huang, Xuhui Li, and Fucang Jia. Automatic liver segmentation using multiple prior knowledge models and free–form deformation. In Orcun Goksel, editor, *Proceedings of the VISCERAL Challenge at ISBI*, CEUR Workshop Proceedings, pages 22–24, Beijing, China, May 2014.

[HMLM14] Allan Hanbury, Henning Müller, Georg Langs, and Bjoern H. Menze. Cloud–based evaluation framework for big data. In Alex Galis and Anastasius Gavras, editors, *Future Internet Assembly (FIA) book 2013*, Springer LNCS, pages 104–114. Springer Berlin Heidelberg, 2014.

[JdTM14] Oscar Alfonso Jiménez del Toro and Henning Müller. Hierarchical multi–structure segmentation guided by anatomical correlations. In Orcun Goksel, editor, *Proceedings of the VISCERAL Challenge at ISBI*, CEUR Workshop Proceedings, pages 32–36, Beijing, China, May 2014.

[KCAB09] Hassan Khotanlou, Olivier Colliot, Jamal Atif, and Isabelle Bloch. 3d brain tumor segmentation in MRI using fuzzy classification, symmetry analysis and spatially constrained deformable models. *Fuzzy Sets and Systems*, 160(10):1457–1473, 2009. Special Issue: Fuzzy Sets in Interdisciplinary Perception and Intelligence.

[LMMH13] Georg Langs, Henning Müller, Bjoern H. Menze, and Allan Hanbury. Visceral: Towards large data in medical imaging – challenges and directions. *Lecture Notes in Computer Science*, 7723:92–98, 2013.

[SJ14] Assaf B. Spanier and Leo Joskowicz. Rule–based ventral cavity multi–organ automatic segmentation in ct scans. In Orcun Goksel, editor, *Proceedings of the VISCERAL Challenge at ISBI*, CEUR Workshop Proceedings, pages 16–21, Beijing, China, May 2014.

[VPYM11] Nagesh Vadaparthi, Suresh Varma Penumatsa, Srinivas Yarramalle, and P. S. R. Murthy. Segmentation of Brain MR Images based on Finite Skew Gaussian Mixture Model with Fuzzy C–Means Clustering and EM Algorithm. *International Journal of Computer Applications*, 28:18–26, 2011.

[WS14] Chunliang Wang and Örjan Smedby. Automatic multi–organ segmentation using fast model based level set method and hierarchical shape priors. In Orcun Goksel, editor, *Proceedings of the VISCERAL Challenge at ISBI*, CEUR Workshop Proceedings, pages 25–31, Beijing, China, May 2014.

[ZWB+04] Kelly H. Zou, Simon K. Warfield, Aditya Bharatha, Clare M.C. Tempany, Michael R. Kaus, Steven J. Haker, William M. Wells III, Ferenc A. Jolesz, and Ron Kikinis. Statistical validation of image segmentation quality based on a spatial overlap index[1]: scientific reports. *Academic Radiology*, 11(2):178 – 189, 2004.

Rule-Based Ventral Cavity Multi-Organ Automatic Segmentation in CT Scans

Assaf B. Spanier
School of Eng. and Computer Science
The Hebrew Univ. of Jerusalem, Israel
assaf.spanier@mail.huji.ac.il

Leo Joskowicz
School of Eng. and Computer Science
The Hebrew Univ. of Jerusalem, Israel
josko@cs.huji.ac.il

Abstract

We describe a new method for the automatic segmentation of multiple organs of the ventral cavity in CT scans. The method is based on a set of rules that determine the order in which the organs are isolated and segmented, from the simplest one to the most difficult one. First, the body is isolated from the background. Second, the trachea and the left and right lungs are segmented based on their air content. Third, the spleen and the kidneys – the organs with high blood content – are segmented. Finally, the kidney is segmented based on the surrounding organs segmentation. Each organ is individually segmented with a four-step procedure that consists of: 1) definition of an inclusive region of interest; 2) identification of the largest axial cross-section slice; 3) removal of background structures by morphological operations, and; 4) 3D region growing segmentation. Our method is unique in that it uses the same generic segmentation approach for all organs and in that it relies on the segmentation difficulty of organs to guide the segmentation process. Experimental results on 15 CT scans of the VISCERAL Anatomy2 Challenge training datasets yield a Dice volume overlap similarity score of 79.1 for the trachea, 97.4 and 97.6 for the left and right lungs, 89.2 for the spleen, and 92.8 for the left kidney. For the 5 CT scans test datasets, the Dice scores are 97.9, 97.0, 85.6, 93.4 and 90.2, respectively. Our method achieved an overall DICE score of 92.8 and was ranked first among the five methods that participated in the challenge.

In: O. Goksel (ed.): Proceedings of the VISCERAL Organ Segmentation and Landmark Detection Benchmark at the 2014 IEEE International Symposium on Biomedical Imaging (ISBI), Beijing, China, May 1^{st}, 2014
published at http://ceur-ws.org

1 Introduction

The increasing amount of medical imaging data acquired in clinical practice constitutes a vast database of untapped diagnostically relevant information. Today, only a small fraction of this information is used during clinical routine or research due to the complexity, richness, high dimensionality, and data size [1].

Content-based image retrieval (CBIR) techniques have been proposed to access this information and to identify similar cases to assist radiologists in the clinical decision support process [2]. The segmentation of individual ventral cavity organs in CT scans is expected to improve the diagnostic accuracy and performance of CBIR systems. While the manual delineation of these organs is considered the gold standard, this is a tedious and very time-consuming process which is impractical for all but a few dozen datasets for research. Consequently, a plethora of methods for automatic segmentation of ventral body cavity organs in CT scans have been proposed. Liver segmentation methods are thoroughly summarized and reviewed by Mharib et al. [4]. Lungs segmentation from CT scans has been addressed by Sluimer et al. [5]. Kidney segmentation methods are described in Freiman et al [6]. While very different from each other, all these methods target a single organ and do not use information about other organs' segmentations. Thus, multi-organ segmentation requires a specific method for each organ, which yields variable quality results and quickly becomes unmanagable as the number of organs to be segmented grows. It is thus desirable to develop a single, generic approach that can be customized for each organ and that uses the information about other organs' segmentations.

The rule-based approach to medical image segmentation calls for using each organ anatomical context and prior knowledge about its location and its extension for enhancing, improving, and automating the segmentation process. In this pipeline-oriented approach, the organs of interest are successively extracted from the CT scan. Previous research has focused mainly on liver segmentation [3]. In this paper, we extend and generalize the rule-based approach to the automatic segmentation of multiple ventral cavity organs in CT scans.

2 Method

The basic premise of the rule-based paradigm is to sequentially extract different organs based on prior information on the organs of interest and their characteristics in the CT scan. Simple and context-free organs are segmented first, followed by more complex and context-based identfication and delineation. Our proposed approach extends the established rule-based approach by providing a unified, generic four-step approach that is customized for each organ and incorporates information about other organs prior segmentations.

2.1 Generic organ segmentation framework

In our generic framework, the segmentation of each organ is performed in four successive steps (Fig. 1):

1. Definition of the organ's Binary Inclusive Region Of Interest (BI-ROI) based on the target organ intensity values.
2. Identification of the organ's Largest Axial Cross Section Slice (LACSS). This is the CT scan slice where the organ has the largest axial area.
3. Removal of remaining background structures from the LACSS by morphological operations.
4. Organ segmentation by 3D region growing starting from the LACSS inside the BI-ROI.

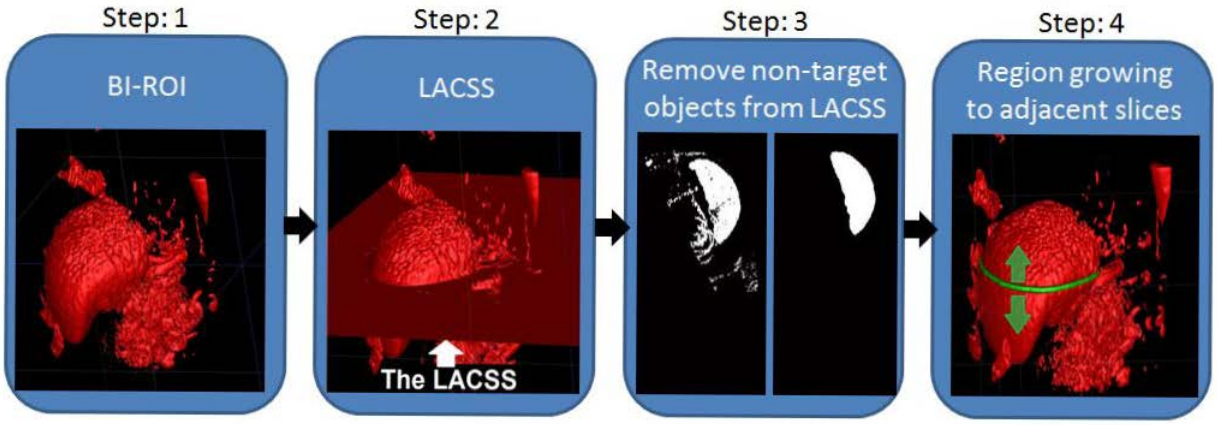

Figure 1: The four steps of the generic organ segmentation framework exemplified on the spleen.

We start with a preprocessing step that isolates the patient body from the background (air and scan gantry) based on location and intensity values. The generic four-step framework is then applied to the ventral body cavity organs in the following order. First, the breathing system organs are segmented: the trachea and the left and right lungs. Next, the organs with high blood content are segmented: the spleen, the liver, and the left and right kidneys. This organ segmentation order prevents ambiguous assignment of the same image region to multiple organs, as previously segmented image regions are excluded from the segmentation process. Due to space limitations, we illustrate below each step for the breathing system only.

2.2 Breathing system segmentation

The breathing system consists of the trachea and the left and right lung.

Step 1: Definition of the BI-ROI: Binary Inclusive Region of Interest

We perform a simple thresholding with the Houndsfiel Unit (HU) of air and fat (< -500HU). This results in a binary map consisting of air, fat, and other background structures. Then, the trachea and the lungs are separated from the from the undesired surrounding fat by finding the the largest connected component. The resulting structure includes the breathing system defines the trachea and lungs BI-ROI (Fig. 2a). This BI-ROI is further refined for the trachea and the left and right lungs.

Step 2: Identification of the LACSS: Largest Axial Cross Section Slice

The Largest Axial Cross Section Slice (LACSS) of the trachea and the lungs are identified by finding the CT slices in the BI-ROI with the narrowest and widest perimeters, respectively (Figs. 2b and 2c). Note that the lungs slice contains two connected components, for the left and right lungs.

Step 3: Removal of background structures

No further background removal is required for the trachea and lungs, since the lungs LACSS contains exactly two connected components corresponding to the left and right lungs and the trachea LACSS contains exactly one connected component (Fig. 2).

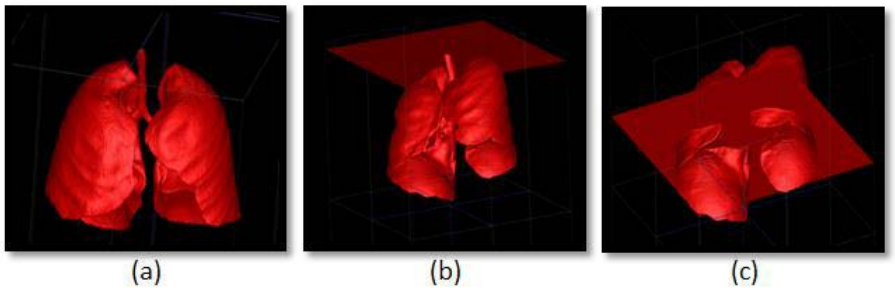

Figure 2: Illustration of the results of the first two steps of the generic organ segmentation framework on the breathing system: a) Binary Inclusive Region of Interest; b) Largest Axial Cross Section Slice plane for the trachea, and c) Largest Axial Cross Section Slice plane for the lungs.

Step 4: Segmentation by 3D region growing

The trachea and the left and right lungs are segmented by 3D region growing. The process starts at the LACSS and proceeds to adjacent CT slices withing the volume defined by the BI-ROI.

First, the distance map between the LACSS contour (Fig 3a) and the adjacent slice (Fig 3b) is computed, with the LACSS contour distance set to 0 (Fig 3c). Next, the adjacent slice and the distance map are intersected to identify regions of high change (Fig 3d). In the resulting intersection, we define a series of windows along the contour (Fig 3e) and compute the intensity histogram in each window. Finally, windows whose histograms have a positive kurtosis are considered as segmentation leakage, e.g. windows R_1 and R_2 in Fig 3e. These windows contain undesired structures whose voxels are removed from the organ segmentation. Windows whose histograms have zero or negative kurtosis, e.g. window R_3 in Fig 3e, are considered smooth and are thus segmented as part of the organ of interest. This process is repeated throughout the slices of the image until the number of segmented pixels in the slice is below a predefined threshold.

The idea behind this step is that ventral cavity organs are relatively smooth, so two adjacent slices of the same organ cannot exceed some level of variability. From a geometric point of view, we constrain the expansion of the region growing boundary curve to have variable speed according to the context. Our method takes into account the local geometry of the curvature: when the magnitude of the curvature is above a predefined threshold, we stop its propagation and allow it to continue only in low curvature regions. The rationale is once again that ventral body cavity organs should preserve some level of smoothness constraint. The final result is the 3D segmentation of the organ (Fig 4).

3 Experimental Results

We evaluated our method on two sets of scans of the VISCERAL Anatomy2 Challenge. The training and test datasets consist of 15 and 5 CT clinical scans, respectively, acquired in 2004-08. Datasets of patients younger than 18 years were not included following the recommendation of the local ethical committee (S-465/2012, approval date Feb. 21th 2013). The CT scans in-plane resolution is 0.604-0.793/0.604-0.793mm; the in-between plane resolution is >=3mm. A VISCERAL team radiologist manually produced ground-truth segmentations for each scan.

Table 1 summarizes the results for each type of dataset and organ: training and test datasets, left lung, right lung, trachea, spleen and left kidney. Note that the DICE similarity coefficients are high or very high, with a relatively small standard deviation. Our method achieved an overall DICE score of 92.8 and was ranked first among the five methods that participated in the challenge. Fig. 4 shows four representative examples of the multi-organ segmentation results.

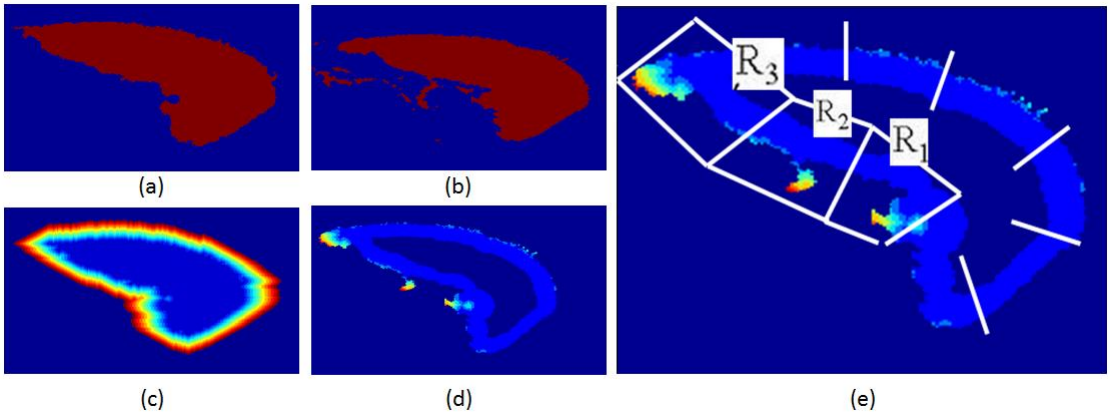

Figure 3: Illustration of step 4, 3D region growing, on the right lung: a) initial LACSS; b) adjacent LACSS; c) distance map; d) intersection image of the adjacent LACSS and distance map, and e) windows along the contour.

Training dataset	Left Lung	Right Lung	Trachea	Spleen	Left Kidney
	97.4	97.6	79.1	89.2	92.8
Test dataset	Left Lung	Right Lung	Trachea	Spleen	Left Kidney
	97.9	97.0	85.6	93.4	90.2

Table 1: Results: Mean Dice similarity coeffficient and standard deviation for the training and test datasets on each organ.

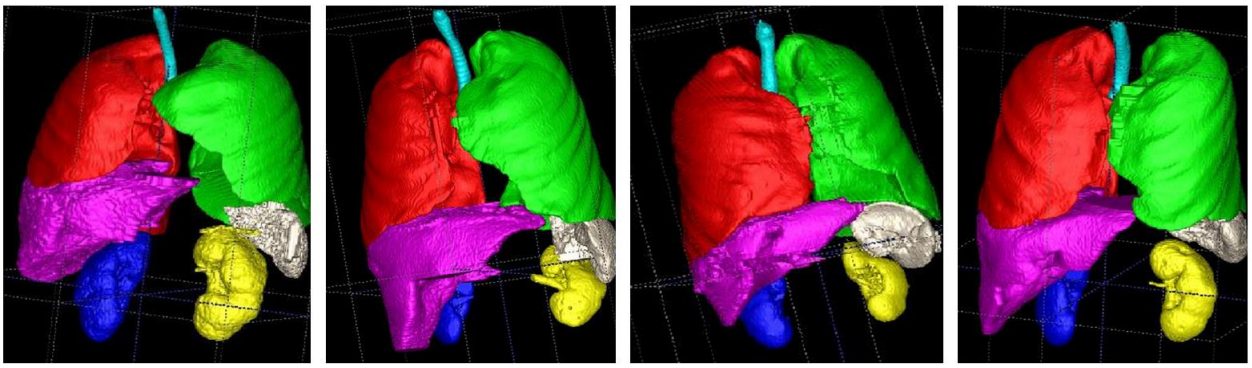

Figure 4: Multi-organ segmentation results of four representative datasets of the VISCERAL Anatomy2 Challenge.

4 Conclusions

We have developed a generic framework for the segmentation of ventral body cavity organs in CT scans. Our approach consists of four-step pipeline method that takes into account prior information about the locations of the organs and their appearance in CT scans. We have shown that the method is applicable to a variety of ventral body cavity organs including the trachea, the left and right lungs, the spleen, and the left kidney.

Current and future research is incorporating other structures, including the right kidney and the liver. We are also extending the 3D region growing step to include different smoothing criteria in different regions of the organ, to eliminate and avoid leakage to neighboring organs.

References

[1] Langs G, Hanbury A, Menze B. VISCERAL: towards large data in medical imaging challenges and directions. *Proc. MICCAI LNCS: Medical Content-Based Retrieval for Clinical Decision Support*, pp 92-98, 2013.

[2] Rubin DL, Akgl CB, Napel S, Beaulieu CF, Greenspan H, Acar B. Content-based image retrieval in radiology: current status and future directions. *Journal of Digital Imaging*, pp 24-32, 2011.

[3] Schmidt G. et al. Cognition network technology for a fully automated 3D segmentation of the liver. *Proc. MICCAI Workshop on 3D Segmentation in the Clinic: A Grand Challenge*, pp 125-133, 2007.

[4] Mharib AM, Rahman A, Mashohor S, Binti R. Survey of liver CT image segmentation methods. *Artificial Intelligence Review*, pp 37-42, 2012.

[5] Sluimer I, Schilham A, Prokop M, van Ginneken B. Computer analysis of computed tomography scans of the lung: a survey. *IEEE Transactions on Medical Imaging* Vol 25(4), pp. 385, 2006.

[6] Freiman M, Kronman A, Esses SJ, Joskowicz L, Sosna J. Non-parametric iterative model constraint graph min-cut for automatic kidney segmentation. *Proc. MICCAI, LNCS* 3217, pp. 73-80, 2010.

Automatic Liver Segmentation using Multiple Prior Knowledge Models and Free-Form Deformation

Cheng Huang, Xuhui Li, Fucang Jia
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
1068 Xueyuan Avenue, Xili University Town, Shenzhen, 518055, China
Email: fc.jia@siat.ac.cn

Abstract

In this paper, an automatic and robust coarse-to-fine liver image segmentation method is proposed. Multiple prior knowledge models are built to implement liver localization and segmentation: voxel-based AdaBoost classifier is trained to localize liver position robustly, shape and appearance models are constructed to fit liver shape and appearance models to original CT images. Free-form deformation is incorporated into segmentation process to improve the model's ability of refining liver boundary. The method was tested on IBSI 2014 VISCERAL challenge datasets and the result demonstrates that the proposed method is robust and efficient.

1 Introduction

Accurate and robust liver segmentation in CT images is an indispensable part in liver quantitative diagnosis and surgery planning, while variation in liver shape, appearance and fuzzy boundary remain challenging. Recently, prior knowledge models learned from big data play an important role in successful clinical image segmentation. In this study, integrating of discriminative and generative models in a hybrid scheme was presented to assist liver localization and segmentation: machine learning based voxel classifier, active shape model (ASM) [Cootes95] including statistical shape model (SSM) prior and local appearance model. Finally, the final fitted model was free-form deformed to true liver boundary under appearance model guidance. The coarse-to-fine liver image segmentation framework including liver localization, model reconstruction, model fitting and free-form deformation is illustrated in Figure 1.

In: O. Goksel (ed.): Proceedings of the VISCERAL Organ Segmentation and Landmark Detection Benchmark at the 2014 IEEE International Symposium on Biomedical Imaging (ISBI), Beijing, China, May 1st, 2014
published at http://ceur-ws.org

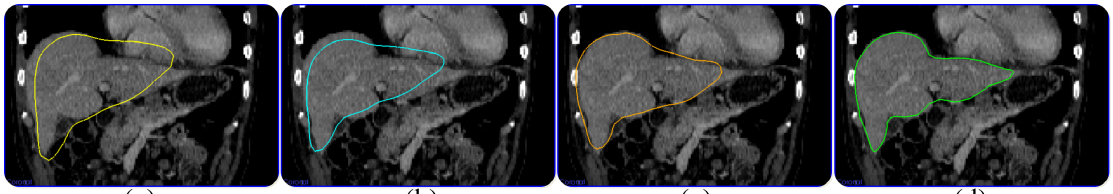

Figure 1: The four steps of liver segmentation framework: (a) liver model location; (b) registration with liver distance map; (c) shape fitting under appearance guidance; (d) free-form deformation.

2 Method

2.1 Liver localization

An atlas image based rigid registration with correlation coefficient histogram metric was used to detect the region of interest (ROI) of liver. A set of image features such as region mean intensity, variance, location, histogram and contextual features were extracted to train an AdaBoost classifier, by which a liver probability map was generated, and the position of the liver was robustly estimated.

2.2 Model reconstruction

The SSM of liver was constructed from training CT images and corresponding binary segmentations. Firstly, pose training described in [Huang13] was applied to resample all the images. For shape correspondence establishment, one reference mesh was obtained by marching cubes method, all other training segmentations were elastic registered to the reference mesh, landmarks were sampled equally on each training mesh. The SSM was constructed by Statismo toolkit [Luthi12] and represented by simplex mesh.

The local appearance model of liver was established by a K Nearest Neighbor (KNN)-classifier trained on both intensity and gradient profiles information inside, outside and at the true liver boundary as suggested in [Heimann07]. For each landmark, profiles perpendicular to the surface are sampled from all training volumes and stored as boundary samples. Additional non-boundary samples were acquired by shifting the profiles towards the inside and outside of the liver.

2.3 Shape and appearance profile fitting

For the image to be segmented, a liver probability map was derived by AdaBoost classifier, and the binary mask can be obtained at threshold 0.5. The distance map image was applied to register to the point sets of the mean shape model, and the mesh vertexes of deformed mean shape were fitted to liver boundary location with major shape variation constraints.

The appearance model is utilized to drive the model toward the precise liver boundary. Local appearance features for all landmarks are extracted at different positions perpendicular to the model surface. Previous trained KNN-classifier shifts landmarks to the optimal displacement position with maximum boundary probability.

2.4 Free-form deformation

Once appearance profile fitting has converged, the deformed shape model were then free-form deformed to the more accurate position. Free deformation was implemented based on deformable simplex mesh [Montagnat97] segmentation. The internal force strives to keep the deformable mesh close to the best fitting SSM, and the external forces tries to move all vertices to the locations where intensity or gradient appearance model predicts the highest boundary probability. Previous

KNN-classifier was integrated as external force to deform to conquer local specific variation of liver shape.

3 Result

Seven CT and seven CTce IBSI VISCERAL challenge 2014 datasets were employed to train AdaBoost classifier. Additional fifty manually segmented datasets were used to train the prior shape and appearance models. There are 1252 landmarks in the liver shape model, each landmark is sampled with 11 points in the landmark normal direction in the profile model. The experiment was tested on 8 CT and 8 CTce datasets. The four evaluation metric scores are as follows: average dice coefficient were 0.924 and 0.925, interclass correlation were 0.924 and 0.925, adjusted rand index were 0.923 and 0.920 and average distance were 0.222mm and 0.261mm for CT and CTce modality respectively.

4 Conclusion

In this paper, a robust and automatic liver segmentation method is proposed. The method exploits different prior knowledge to represent contextual, profile appearance and shape variation of liver, relies on different registration to construct liver model, liver localization, model fitting and refined deformation. The method has been validated on ISBI VISCERAL challenge and showed good performance. In future, we will adapt the method to other visceral organs segmentation.

5 Acknowledgments

The work was supported by the National High-tech R&D Program of China (863 Program) (No.2012AA022305).

References

[Cootes95] T. Cootes, C. J. Taylo, D. H. Cooper, J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[Huang13] C. Huang, F. Jia, C. Fang, Y. Fan, Q. Hu. Automatic liver detection and segmentation from 3D CT images: a hybrid method using statistical pose model and probabilistic atlas. *International Journal of Computer Assisted Radiology and Surgery*, 8(S1):237-238, 2013.

[Heimann07] T. Heimann, H. P. Meinzer, I. Wolf. A statistical deformable model for the segmentation of liver CT volumes. *MICCAI Workshop: 3D Segmentation in the clinic: A grand challenge*, 161-166, 2007.

[Luthi12] M. Luthi, R. Blanc, T. ALbrecht, T. Gass, O. Goksel, P. Buchler, M. Kistler, H. Bousleiman, M. Reyes, P. Cattin, T. Vetter. Statismo - A framework for PCA based statistical models. *The Insight Journal*, 1:1-18, 2012.

[Montagnat97] J. Montagnat, H. Delingette. Volumetric medical images segmentation using shape constrained deformable models. *CVRMed-MRCAS'97, Springer Berlin Heidelberg*, 13-22, 1997.

Automatic multi-organ segmentation using fast model based level set method and hierarchical shape priors

Chunliang Wang
Center for Medical Imaging Science and Visualization(CMIV),Linköping University
Linköping, Sweden
chunliang.wang@liu.se

Örjan Smedby
Department of Medical and Health Sciences (IMH), Linköping University
Linköping, Sweden
orjan.smedby@liu.se

Abstract

An automatic multi-organ segmentation pipeline is presented. The segmentation starts with stripping the body of skin and subcutaneous fat using threshold-based level-set methods. After registering the image to be processed against a standard subject picked from the training datasets, a series of model-based level set segmentation operations is carried out guided by hierarchical shape priors. The hierarchical shape priors are organized according to the anatomical hierarchy of the human body, starting with ventral cavity, and then divided into thoracic cavity and abdominopelvic cavity. The third level contains the individual organs such as liver, spleen and kidneys. The segmentation is performed in a top-down fashion, where major structures are segmented first, and their location information is then passed down to the lower level to initialize the segmentation, while boundary information from higher-level structures also constrains the segmentation of the lower-level structures. In our preliminary experiments, the proposed method yielded a Dice coefficient around 90% for most major thoracic and abdominal organs in both contrast-enhanced CT and non-enhanced datasets, while the average running time for segmenting ten organs was about 10 minutes.

1 Introduction

Automatic segmentation of anatomical structures has great value for both clinical and epidemiological studies. Some common examples include using a brain segmentation tool for quantitative mea-

In: O. Goksel (ed.): Proceedings of the VISCERAL Organ Segmentation and Landmark Detection Benchmark at the 2014 IEEE International Symposium on Biomedical Imaging (ISBI), Beijing, China, May 1st, 2014
published at http://ceur-ws.org

surements of brain structure changes to study Alzheimer's disease [FSB+02], using an automated lung segmentation method to define the region of interest for computer-aided diagnosis (CAD) methods for more efficient screening and earlier detection of tumors, and using liver segmentation for surgery planning to achieve more precise and better cancer treatment [HVGS+09]. Besides these single organ applications, the multi-organ segmentation methods have broader applications, such as radiotherapy planning and semantic image segmentation and content retrieving [SKM+10]. Many automated organ segmentation methods have been proposed in the literature, such as the active shape model (ASM) [CTCG95], atlas-based methods [ISR+09] and machine-learning-based methods [ZBG+08]. The robustness of these single-organ approaches is usually unsatisfactory. This is related to the fact that the boundary between two organs may be inadequately defined due to limited resolution and intensity similarity. Even with the help of shape priors, most algorithms still have difficulties in discriminating between another organ and anatomical variation of the same organ. Recently, a number of multi-organ segmentation approaches have been proposed, thanks to the improving performance of modern computers and the increasing recognition of the advantages of considering multi-organ simultaneous in the image models. Okada et al. proposed a hierarchical organization of organ ASMs [OYH+08], where the inter-organ position changing is decoupled from the individual organs morphological variations. Promising results were obtained in their upper-abdominal organ segmentation in contrast-enhanced CT scans. Wolz et al. proposed a hierarchical atlas registration and weighting scheme, which sequentially picks the close-looking atlases, best-matching organ atlases and best-fitting segmentation patches in a three-level coarse-to-fine registration pipeline [WCM+12]. A few machine learning based methods were also reported [MSW+11, KSZ+11]. In [WS14], we proposed an automatic multi-organ segmentation method using hierarchical-shape-prior guided level sets. The hierarchical shape priors are organized according to the anatomical hierarchy of the human body, so that the major structures with less population variation are at the top, and smaller structures with higher irregularities are linked at a lower level. The segmentation is performed in a top-down fashion, where major structures are segmented first, and their location information is then passed down to the lower level to initialize the segmentation, while boundary information from higher-level structures also constrains the segmentation of the lower-level structures. The proposed method delivered relatively accurate results in non-enhanced CT datasets [WS14]. In this paper, we extend the framework to process both non-enhanced and contrast-enhance CT datasets, by introducing an iterative organ intensity estimation step.

2 Methods

Figure 1 summarizes the processing pipeline of the proposed segmentation framework, which can be roughly divided into three phases: preprocessing, hierarchical shape model guided multi-organ segmentation and iterative organ intensity estimation. Detailed descriptions of these phases are given in the following sections.

2.1 Preprocessing

A skin and subcutaneous fat stripping step is first carried out to remove the large variation of the subcutaneous fat distribution among the population. This is done with a two-step threshold-based level set segmentation combined with mathematical morphology operations. First, the surface of the human body is segmented with a threshold of 300 HU and an initial seed region set to cover the whole volume. The resulting mask is then processed with an erosion operator to remove the skin. Finally a second round threshold-based level set segmentation is carried out with the threshold set to 0 HU. After subcutaneous fat stripping, the musculoskeletal figure of a patient tends to vary less from patient to patient. A straightforward rigid registration is carried out between the

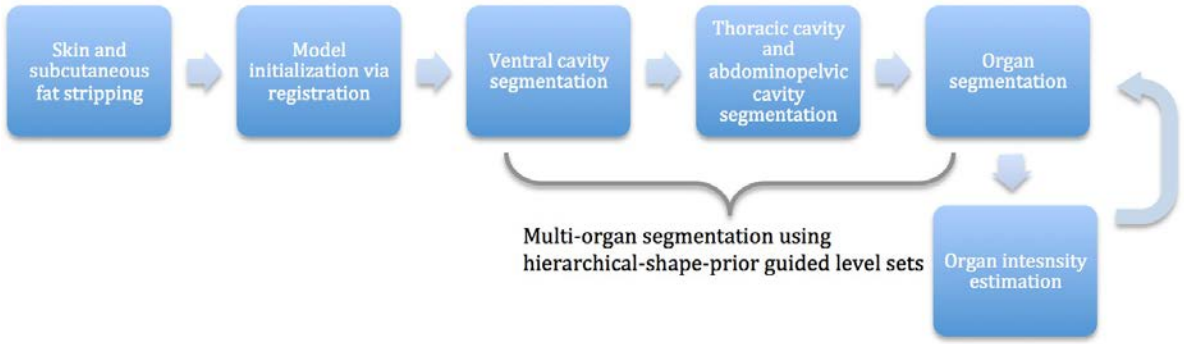

Figure 1: The processing pipeline of the proposed multi-organ segmentation framework.

unseen patient and a selected standard subject. This standard subject (common-looking subject) was manually selected by visually comparing the appearance among the sample group. The air-filled lung areas in both datasets are set to the fat tissue intensity to reduce their influence on the registration, so the skeletons are better aligned. The transformation matrix from the registration step is used to initialize the position of the hierarchical shape model. For non-enhanced CT datasets, a cropping step is introduced to limit the remaining processing to the torso. The largest torso cross-section area is estimated by finding the largest connected region (2D) within the musculoskeletal figure among all axial slices. The starting and ending slice of the torso is then defined as the first slice, on either side of the largest torso cross-section slice, in which the width of the largest connected region (2D) is below half the width of the largest torso cross-section area.

2.2 Hierarchical shape model guided multi-organ segmentation

The hierarchical shape model used in this study is shown in Figure 2. To generate statistical shape priors for individual structures, all segmentation masks of the corresponding organ are registered to the common-looking subject. To link a statistical shape prior to its parent structures space, the statistical mean shape is registered against a trust zone created by thresholding the probability atlas of that anatomical structure in the upper-level structures space. More detailed description of building the hierarchical shape priors can be found in [WS14].

The segmentation is performed in a top-down fashion, i.e. ventral cavity is first segmented, and then divided into thoracic cavity and abdominopelvic cavity. The third level contains the individual organs such liver, spleen and kidneys. The location information of a higher level structure is passed down to the lower level to initialize the segmentation. Within the same level, structures are segmented sequentially from left to right as the order listed in Figure 2. Segmented regions are set to different empirically defined likelihood values to guide the following segmentation.

2.3 Iterative organ intensity estimation

In the proposed hierarchical-shape-prior guided level set framework, the external speed function is an intensity mapping function, which is similar to the threshold function in the threshold-based level set method proposed by Lefohn et al. [LCW03]. In [WS14], the upper and lower thresholds are empirically defined beforehand for different structures. Since the intensity of some organs in contrast-enhanced CT scans can vary depending on the circulation rate and acquisition timing, we introduced an iterative approach to estimate the intensity range of heart, liver, kidney and spleen. An organs upper and lower threshold are estimated to be $M + 1.5\sigma$ and $M - 1.5\sigma$, where M and σ is the mean and standard deviation of the voxel intensity within the current segmented area. All voxels with intensity lower than 30HU are excluded from the calculation of M and σ.

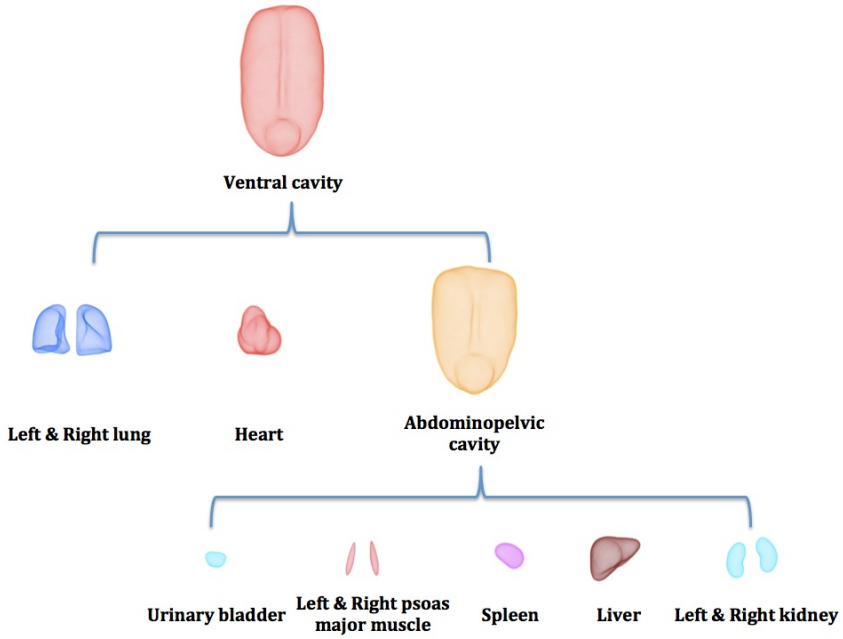

Figure 2: The hierarchical shape model used in this study

The intensity estimation is repeated every 15 iterations of the model fitting process. The iterative intensity estimation stops when the changing rates of M and σ are both lower than a threshold (5 HU). The fixed thresholds reported in [WS14] are used as the initial setting for these organs in the beginning of organ segmentation.

2.4 Model-guided level set using coherent propagation

In this study, the model-based level set method proposed by Leventon et al. [LGF02](Leventon, Grimson, and Faugeras 2002) is adapted for individual structure segmentation at different levels. Making this method efficient and accurate is essential for the usability and robustness of the whole framework. In our earlier papers [WFS11, WFS14]), we proposed a fast level set method using coherent propagation, which achieved 10100 times speed-up in various segmentation tasks when compared with the sparse field level set algorithm. In [WS14], we extended the coherent propagation method to model-based level sets, which can not only speed up the level set propagation, but also reduce the frequency of shape-prior registration by taking advantage of the convergence detection of the coherent propagation. In this new framework, the model fitting operation is only repeated if the contour has moved a certain distance from the previously estimated model.

3 Results

.The proposed method was trained on 7 training CT datasets and tested on 5 non-enhanced CT datasets and 5 contrast-enhanced CT datasets. These CT images are down-sampled to 333 mm resolution, whereas the segmentation results are up-sampled to the original resolution for evaluation. All these datasets were obtained from the Visceral Benchmark 1 site (visceral.eu) [HML$^+$12]. Overall, the proposed method yielded a Dice coefficient around 90% for most major organs. Detailed results are listed in Table 1. The average processing time for segmenting all ten major organs is about 10 minutes (excluding the resampling steps) on an 8-core Mac Pro (2.26GHz). Figure 3 shows an example of the segmentation results at different stages from one testing dataset.

Table 1: SEGMENTATION RESULTS

Organ Name	Non-enhanced CT		Contrast-enhanced CT	
	Dice coefficient (%)	Average Hausdorff distance	Dice coefficient (%)	Average Hausdorff distance
Liver	0.904	0.46	0.887	0.65
Spleen	0.887	0.45	0.842	0.87
Left lung	0.971	0.07	0.956	0.15
Right lung	0.972	0.06	0.942	0.20
Left kidney	0.729	3.63	0.896	0.27
Right kidney	0.777	1.21	0.890	0.28
Bladder	0.806	0.78	0.738	1.59

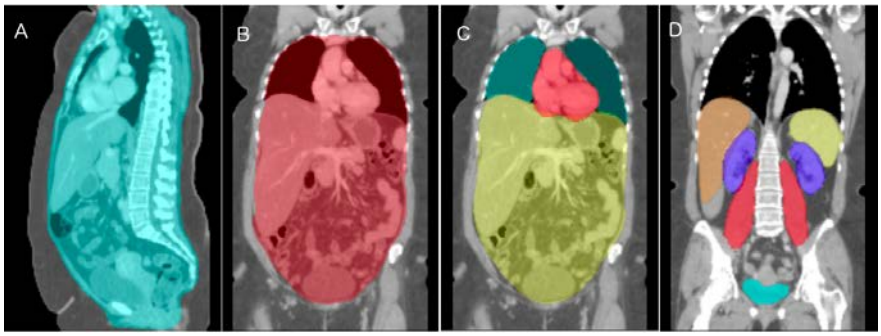

Figure 3: Segmentation results at different stages. A, segmentation result after skin and subcutaneous fat stripping; B, segmentation result of the ventral cavity; C, segmentation result of the second level structures; D, segmentation results of the third level structures.

4 Discussion and Conclusion

The proposed segmentation method has a number of limitations. First, the statistical shape priors for different structures were trained on 7 subjects, which can over-constrain the segmented area (cf. liver segmentation in Figure 3D). Second, as the top-down strategy suffers from the accumulated error being passed down along the hierarchy tree, a bottom-up feedback path should be added to allow the lower structure to recover the higher level errors. Future work also includes improving segmentation accuracy by using more edge-based image terms and extending the framework to handle MRI images. In conclusion, a multi-organ segmentation framework using hierarchical shape priors is presented. This method gradually improves the estimation of the organ location by first segmenting out large and regular-shaped structures. The appearance of organs is iteratively estimated based on statistical analysis of preliminary segmentation results. Preliminary results on non-enhanced and contrast-enhance CT datasets are encouraging.

References

[CTCG95] Timothy F. Cootes, Christopher J. Taylor, David H. Cooper, and Jim Graham. Active shape models-their training and application. *Computer vision and image understanding*, 61(1):3859, 1995.

[FSB+02] Bruce Fischl, David H. Salat, Evelina Busa, Marilyn Albert, Megan Dieterich, Christian Haselgrove, Andre van der Kouwe, Ron Killiany, David Kennedy, Shuna Klave-

ness, Albert Montillo, Nikos Makris, Bruce Rosen, and Anders M. Dale. Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3):341–355, January 2002.

[HML+12] Allan Hanbury, Henning Mller, Georg Langs, Marc Andr Weber, Bjoern H. Menze, and Tomas Salas Fernandez. Bringing the algorithms to the data: Cloudbased benchmarking for medical image analysis. In *Information Access Evaluation. Multilinguality, Multimodality, and Visual Analytics*, page 2429. Springer, 2012.

[HVGS+09] T. Heimann, B. Van Ginneken, M.A. Styner, Y. Arzhaeva, V. Aurich, C. Bauer, A. Beck, C. Becker, R. Beichel, G. Bekes, F. Bello, G. Binnig, H. Bischof, A. Bornik, P. Cashman, Ying Chi, A. Cordova, B.M. Dawant, M. Fidrich, J.D. Furst, D. Furukawa, L. Grenacher, J. Hornegger, D. Kainmuller, R.I. Kitney, H. Kobatake, H. Lamecker, T. Lange, Jeongjin Lee, B. Lennon, Rui Li, Senhu Li, H.-P. Meinzer, G. Nemeth, D.S. Raicu, A.-M. Rau, E.M. van Rikxoort, M. Rousson, L. Rusko, K.A. Saddi, G. Schmidt, D. Seghers, A. Shimizu, P. Slagmolen, E. Sorantin, G. Soza, R. Susomboon, J.M. Waite, A. Wimmer, and I. Wolf. Comparison and evaluation of methods for liver segmentation from CT datasets. *IEEE Transactions on Medical Imaging*, 28(8):1251–1265, 2009.

[ISR+09] I. Isgum, M. Staring, A. Rutten, M. Prokop, M.A. Viergever, and B. van Ginneken. Multi-atlas-based segmentation with local decision fusion-application to cardiac and aortic segmentation in CT scans. *IEEE Transactions on Medical Imaging*, 28(7):1000–1010, July 2009.

[KSZ+11] Timo Kohlberger, Michal Sofka, Jingdan Zhang, Neil Birkbeck, Jens Wetzl, Jens Kaftan, Jrme Declerck, and S. Kevin Zhou. Automatic multi-organ segmentation using learning-based segmentation and level set optimization. In *Medical Image Computing and Computer-Assisted InterventionMICCAI 2011*, page 338345. Springer, 2011.

[LCW03] A. Lefohn, J. Cates, and R. Whitaker. Interactive, gpu-based level sets for 3D segmentation. In *Proceedings of the MICCAI conference*, page 564572, 2003.

[LGF02] M. E Leventon, W. E.L Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, volume 1, page 316323, 2002.

[MSW+11] Albert Montillo, Jamie Shotton, John Winn, Juan Eugenio Iglesias, Dimitri Metaxas, and Antonio Criminisi. Entangled decision forests and their application for semantic segmentation of CT images. In Gbor Szkely and Horst K. Hahn, editors, *Information Processing in Medical Imaging*, number 6801 in Lecture Notes in Computer Science, pages 184–196. Springer Berlin Heidelberg, January 2011.

[OYH+08] Toshiyuki Okada, Keita Yokota, Masatoshi Hori, Masahiko Nakamoto, Hironobu Nakamura, and Yoshinobu Sato. Construction of hierarchical multi-organ statistical atlases and their application to multi-organ segmentation from CT images. In *Medical Image Computing and Computer-Assisted InterventionMICCAI 2008*, page 502509. Springer, 2008.

[SKM+10] Sascha Seifert, Michael Kelm, Manuel Moeller, Saikat Mukherjee, Alexander Cavallaro, Martin Huber, and Dorin Comaniciu. Semantic annotation of medical images. In *SPIE Medical Imaging*, page 762808762808, 2010.

[WCM+12] Robin Wolz, Chengwen Chu, Kazunari Misawa, Kensaku Mori, and Daniel Rueckert. Multi-organ abdominal CT segmentation using hierarchically weighted subject-specific atlases. In *Medical Image Computing and Computer-Assisted InterventionMICCAI 2012*, page 1017. Springer, 2012.

[WFS11] Chunliang Wang, Hans Frimmel, and Örjan Smedby. Level-set based vessel segmentation accelerated with periodic monotonic speed function. In *Proceedings of the SPIE Medical Imaging Conference*, pages 79621M–79621M–7, 2011.

[WFS14] Chunliang Wang, Hans Frimmel, and Örjan Smedby. Fast level-set based image segmentation using coherent propagation. *Medical Physics*, 41(7):073501, 2014.

[WS14] Chunliang Wang and Örjan Smedby. Automatic multi-organ segmentation in non-enhanced CT datasets using hierarchical shape priors. In *ICPR2014 Acepted*, 2014.

[ZBG+08] Yefeng Zheng, Adrian Barbu, Bogdan Georgescu, Michael Scheuering, and Dorin Comaniciu. Four-chamber heart modeling and automatic segmentation for 3-d cardiac CT volumes using marginal space learning and steerable features. *Medical Imaging, IEEE Transactions on*, 27(11):16681681, 2008.

Hierarchical Multi–structure Segmentation Guided by Anatomical Correlations

Oscar Alfonso Jiménez del Toro
oscar.jimenez@hevs.ch

Henning Müller
henningmueller@hevs.ch

University of Applied Sciences Western Switzerland
University and University Hospitals of Geneva, Switzerland

Abstract

Many medical image analysis techniques require an initial localization and segmentation of anatomical structures. As part of the VISCERAL benchmarks on Anatomy segmentation, a hierarchical multi–atlas multi–structure segmentation approach guided by anatomical correlations is proposed. The method begins with a global alignment of the volumes and refines the alignenment of the structures locally. The alignment of the bigger structures is used as reference for the smaller and harder to segment structures. The method is evaluated in the ISBI VISCERAL testset on ten anatomical structures in both contrast–enhanced and non–enhanced computed tomography scans. The proposed method obtained the highest DICE overlap score in the entire competition for some structures such as kidneys and gallbladder. Similar segmentation accuracies compared to the highest results of the other methods proposed in the challenge are obtained for most of the other structures segmented with the method.

1 Introduction

Anatomical structure segmentation in medical imaging is a fundamental step for further image analysis and computer–aided diagnosis [Doi05]. With the ongoing increase in medical image data, it is necessary to develop fast and automatic algorithms that can process a large quantity of images with high accuracy and sufficient speed for clinical daily use. Although many different methods have already been proposed [LSL+10, CRK+13], it is uncommon to test multiple approaches on the same available dataset. The Visual Concept Extraction Challenge in Radiology (VISCERAL[1]) benchmarks have been organized with the objective to evaluate the available state-of-the-art segmenting

In: O. Goksel (ed.): Proceedings of the VISCERAL Organ Segmentation and Landmark Detection Benchmark at the 2014 IEEE International Symposium on Biomedical Imaging (ISBI), Beijing, China, May 1st, 2014
published at http://ceur-ws.org

[1] http://www.visceral.eu/, as of 27 April 2014

approaches on a large public dataset. Twenty anatomical structures in four imaging modalities, enhanced and non–enhanced magnetic resonance (MR) and computed tomography volumes, are included in both the training and testing sets provided to the participants. The benchmarks are handled in a novel cloud environment that allows to distribute large quantities of volumes and implement algorithms of the research groups under the same conditions (regarding computing power etc.) inside the cloud [LMMH13].

Multi–atlas based segmentation is an approach that requires little or no interaction from the user. It has been evaluated showing high accuracy and consistent reproducibility in different anatomical structures [LSL+10, RBMMJ04]. In this method, an atlas includes a patient volume and a label volume, created by manual annotation, that identifies the location of one or more structures in the patient volume. The target is the query volume where the location of the structures is unknown. Using image registration, the spatial relationship between the target and atlas volume is estimated. The label volumes are transformed taking the coordinate transformation obtained from the registration. Afterwards the labels are fused resulting in a single label volume that provides an estimated location of the label in the target volume. When multiple atlases are used, the local errors of the registration will be removed by a per–voxel classification.

The proposed method was tested on computed tomography scans with ten different anatomical structures. The method can be extended and applied to the other modalities and any of the anatomical structures in the VISCERAL dataset.

2 Method

All volumes are resampled to obtain isotropic 1mm voxels. Afterwards they are down–sampled to half their size in all three dimensions to speed up the registrations and resampled to their original size for the label fusion.

2.1 Image registration

The atlas patient volume, considered as moving volume $V_A(x)$, is registered to the fixed query volume $V_Q(x)$ using the image registration implementation of Elastix software[2] [KSM+10]. The registration is evaluated in every iterative optimization by a cost function C of the parameterized coordinate transformation T_μ from the moving atlas volume V_A to the query volume V_Q. The adaptive stochastic gradient descent optimizer proposed in [KPSV09] is applied. A coordinate transformation is obtained by minimizing the value of C with respect to the transformation:

$$\hat{\mu} \quad = \quad \underset{\mu}{arg\ min}\ C(T_\mu; V_Q, V_A), \tag{1}$$

the subscript μ indicates that the transformation was parameterized with a vector μ that contains the transformation parameters. Normalized Cross–Correlation (NCC) is selected as the similarity metric for cost function C.

2.2 Hierarchical anatomical structure alignment

The anatomy can differ considerably from patient to patient, particularly the spatial relations between the different structures in the same patient volume [JdTM13]. Since multiple structures are segmentation targets in the VISCERAL benchmark, a hierarchical selection of the registrations improves the segmentations of all the structures. A global affine registration is followed by individual affine registrations using local binary masks to enforce the spatial correlation of each

[2]Elastix: http://elastix.isi.uu.nl, 2014.[Online; accesed 27–April–2014].

anatomical structure separately. These masks are obtained from the morphological dilation of the output labels of the different atlases registered in the previous step. The registrations of the bigger structures are used as a starting point for the closely related smaller structures, which are harder to segment. Most of the registrations of the initial bigger structures (liver, lungs, urinary bladder) will be reused in the method which makes it faster than segmenting each structure individually from the start. The method is repeated for the non-rigid registrations of all the target structures. Also the creation of regions-of-interest with the local masks speeds up the image registrations and improves the output estimations.

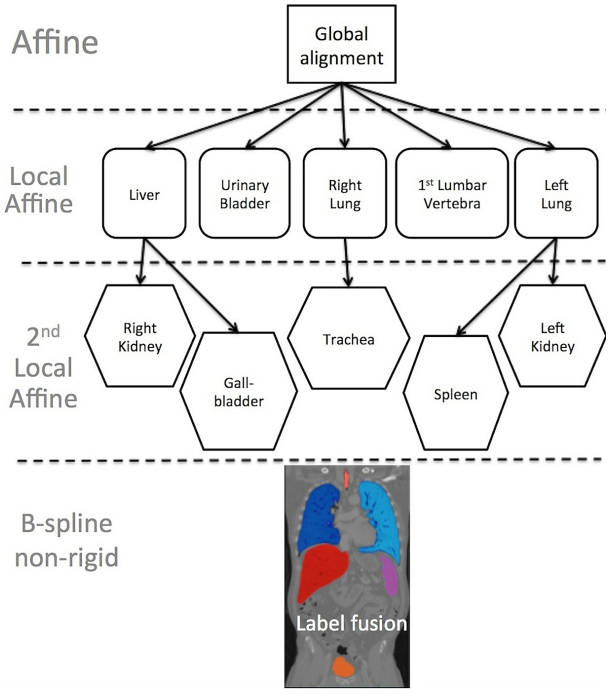

Figure 1: Method Pipeline.

2.3 Non-rigid registration

After each anatomical structure has its own independent ROI mask, the volumes are registered again but using a non–rigid B–spline transformation model. This non–rigid registration allows local deformations obtaining a higher spatial similarity between the volumes. The B–spline registration was also performed in a multi–resolution approach with an adaptive stochastic gradient descent optimizer. This final registration step has a higher computational cost than the affine registration. The transformed labels are updated using the coordinate transformation parameters from the B–spline registration. The new transformed label volumes for each structure constitute the individual votes that will be used for the label fusion step.

2.4 Label fusion

A different label volume is obtained for every atlas registered to the target volume. In order to combine the information obtained from the multiple atlases registered, the output labels are fused in a single label for the target volume. Defining a majority voting threshold is a commonly used label fusion method. An optimal threshold is found for each of the different structures on a per–voxel basis with this approach. Majority voting has also the advantage of providing more than one output segmentation varying the threshold parameter with no additional computations required.

3 Experimental Setup

Ten CT volumes were used to evaluate the performance of the algorithm for the International Symposium on Biomedical Imaging (ISBI) 2014 VISCERAL challenge. Five of them are contrast–enhanced (ceCT) with a field–of–view from below the skull base to the pelvis. The other five are non–enhanced whole body CT scans (wbCT). For the ten CT volumes, ten structures were included in the proposed segmentation method: liver, 2 kidneys, 2 lungs, urinary bladder, spleen, trachea, first lumbar vertebra and gallbladder.

An initial global affine registration is followed by individual affine registrations of the independent structures using local masks as described in the method. The liver, both lungs, $1st$ lumbar vertebra and urinary bladder were segmented with individual affine and non–rigid registrations. The gallbladder and right kidney have the affine alignment of the volume after the liver registrations as a starting point. The left lung affine alignment is used for the spleen and the left kidney. The right lung affine alignment is refined for the trachea segmentation. All structures are refined with non–rigid b–spline registration for the final estimation.

According to the results of the VISCERAL Benchmark 1, an individual majority vote threshold was selected in each structure for the label fusion.

4 Results

The method obtained a total average DICE of 0.789 for ten structures in ceCT and 0.694 for the same ten structures in wbCT (Table 1). All the overlap scores were higher in ceCT and in close relation to the results from the other participants in the challenge for the same anatomical structures. The method obtained the best DICE score of the ISBI Visceral challenge for the left kidney, right kidney and the gallbladder in ceCT. For wbCT the method had the best DICE in the 1st lumbar vertebra, gallbladder and trachea.

Table 1: Average Segmentation Accuracy

Structure	Reference structure	DICE CTwb	DICE ceCT
Liver	none	0.823	0.908
Right lung	none	0.967	0.963
Left lung	none	0.969	0.952
Urinary bladder	none	0.616	0.68
1st Lumbar vertebra	none	0.44	0.472
Right kidney	liver	0.649	0.905
Gallbladder	liver	0.271	0.4
trachea	right lung	0.855	0.83
Spleen	left lung	0.677	0.859
Left kidney	left lung	0.678	0.923

5 Conclusions

The proposed method showed robustness in the segmentation of multiple structures from two different modalities of the challenge using a relatively small dataset. The overlap accuracies are consistent for most of the evaluated anatomical structures and obtained some of the best structure overlap of the challenge when compared to the other proposed methods in the same testset.

Due to the flexibility of the method for adding more structures, 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 for the VISCERAL benchmark 2 Anatomy with a much bigger testset.

6 Acknowledgments

This work was supported by the EU/FP7 through VISCERAL (318068).

References

[CRK+13] Antonio Criminisi, Duncan Robertson, Ender Konukoglu, Jamie Shotton, Sayan Pathak, Steve White, and Khan Siddiqui. Regression forests for efficient anatomy detection and localization in computed tomography scans. *Medical Image Analysis*, 17(8):1293–1303, 2013.

[Doi05] K Doi. Current status and future potential of computer–aided diagnosis in medical imaging. *British Journal of Radiology*, 78:3–19, 2005.

[JdTM13] Oscar Alfonso Jiménez del Toro and Henning Müller. Multi–structure atlas–based segmentation using anatomical regions of interest. In *MICCAI workshop on Medical Computer Vision*, Lecture Notes in Computer Science. Springer, 2013.

[KPSV09] Stefan Klein, Josien P.W. Pluim, Marius Staring, and Max A. Viergever. Adaptive stochastic gradient descent optimisation for image registration. *International Journal of Computer Vision*, 81(3):227–239, 2009.

[KSM+10] Stefan Klein, Marius Staring, Keelin Murphy, Max A. Viergever, and Josien P.W. Pluim. Elastix: a toolbox for intensity–based medical image registration. *IEEE Transactions on medical imaging*, 29(1):196–205, 2010.

[LMMH13] Georg Langs, Henning Müller, Bjoern H. Menze, and Allan Hanbury. Visceral: Towards large data in medical imaging – challenges and directions. *Lecture Notes in Computer Science*, 7723:92–98, 2013.

[LSL+10] Marius George Linguraru, Jesse K. Sandberg, Zhixi Li, Furhawn Shah, and Ronald M. Summers. 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.

[RBMMJ04] Torsten Rohlfing, Robert Brandt, Randolf Menzel, and Calvin R. Maurer Jr. 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.

Segmentation and Landmark Localization Based on Multiple Atlases

Orcun Goksel Tobias Gass Gabor Szekely

Computer Vision Lab, ETH Zurich, Switzerland

Abstract

In this work, we present multi-atlas based techniques for both segmentation and landmark detection. We focus on modality and anatomy independent techniques to be applied to a wide range of input images, in contrast to methods customized to a specific anatomy or image modality. For segmentation, we use label propagation from several atlases to a target image via a Markov random field (MRF) based registration method, followed by label fusion by majority voting weighted by local cross-correlations. For landmark localization, we use a consensus based fusion of location estimates from several atlases identified by a template-matching approach. Results in IEEE ISBI 2014 VISCERAL challenge as well as VISCERAL Anatomy1 challenge are presented herein.

1 Introduction

Segmentation and landmark detection are two very common problems in medical image analysis, as they both pertain to several clinical applications. Although there exist methods customized for specific anatomy and modality, generic methods are valuable as they are applicable in a wide range of applications without much effort for customization. Regarding the two tasks above, in this work we use modality and anatomy independent techniques to treat the diverse dataset from the Anatomy challenge series of the VISCERAL (Visual Concept Extraction Challenge in Radiology) Consortium. The methods are detailed below, also presenting our results from the said challenges.

2 Segmentation

For segmentation, a multi-atlas based technique is used by registering several atlases individually to a target image using our implementation of the MRF-based deformable registration method in [GKT+08]. These registrations are then used to propagate the anatomical labels (ground-truth annotations) from each atlas image into the target coordinate frame. At voxel level, a majority

In: O. Goksel (ed.): Proceedings of the VISCERAL Organ Segmentation and Landmark Detection Benchmark at the 2014 IEEE International Symposium on Biomedical Imaging (ISBI), Beijing, China, May 1st, 2014
published at http://ceur-ws.org

voting is held to decide the winning label where each label votes based on the locally-normalized cross-correlation (LNCC) [CBD+03] of the registered atlas to the given target at that location.

2.1 Atlas-based segmentation using registration via MRF

Finding an optimal displacement vector field $\hat{T}$ can be defined as the minimization of a functional:

$$\hat{T} = \arg\min_{T} E\left(T, X, A_n\right) \tag{1}$$

where X is a target image and A_n is an atlas image and E is the registration energy. MRFs provide an efficient means for the minimization of such energy, with the main advantage being that it does not rely on the gradient of the criterion and therefore is less prone to poor local optima. In order to use MRFs for solving the minimization problem, this energy is decomposed into unary (ψ) and pairwise (Ψ) potentials over discrete labels as follows:

$$E\left(T, X, A_n\right) = \sum_{p \in \Omega} \left(\psi_p\left(l_p\right) + \sum_{q \in \mathcal{N}(p)} \lambda\, \Psi_{pq}\left(l_p, l_q\right) \right), \tag{2}$$

where Ω is the discretized image space. The continuous displacement space is sampled discretely, so that each registration label l_p and l_q in the set of all registration labels L_R maps to a unique displacement vector $\vec{d_p}$. The unary potentials are then a local similarity metric, measuring the fit between the deformed atlas image and the target image at a location i. The pairwise potentials correspond to prior assumptions over the displacement, which is often implemented as a smoothing over the neighborhood $\mathcal{N}$ as justified by a first-order Markov assumption. λ is the pairwise weight.

For robust and smooth solution of (2), an efficient method was proposed in [GKT+08] that seeks the displacements $\vec{d}$ of control points in a multi-resolution cubic B-spline framework. We use our implementation of this method with four levels of detail, where the coarsest grid resolution has three nodes along the shortest edge of an input image and a spacing as isotropic as possible given that a control-point is required on each corner of the image. Each following level of the resolution hierarchy has twice the resolution compared to the previous step. At each level of detail, we sample four displacements in each cardinal direction, yielding 25 displacement samples in total. Within each direction, samples are equidistant, with the largest displacement set to 0.4 times the control grid spacing. This was shown to guarantee diffeomorphic deformations [RAH+06]. For each level of detail, we re-run the MRF registration with the displacements being re-scaled by the golden ratio 0.618. The resulting displacements are then composed onto the previous deformation. This guarantees that the result is still diffeomorphic and sub-pixel accuracy can be achieved. For the unary potentials we use normalized cross correlation (NCC) of patches centered around each control point with their radius equivalent to control grid spacing. Euclidean distance between displacements of neighboring control grids is used to penalize non-smooth deformations. Tree-reweighted message passing (TRW-S) [Kol06] is employed to find a solution to each energy minimization instance.

A segmentation candidate of the target image X based on the atlas A_n is then obtained by applying the resulting displacement field $\hat{T}$ to the known segmentation S_{A_n} as follows:

$$S_{X,n} = S_{A_n}(\hat{T}). \tag{3}$$

2.2 Label fusion via weighted majority voting

Although MRF-based registration is a relatively robust method, it can only guarantee a locally optimal solution and is therefore susceptible to poor initialization. Furthermore, for a grossly different atlas, correspondences for registration may not be guaranteed. Accordingly, the segmentation from

a single atlas may not be satisfactory. It was shown in different research fields that combination of multiple weak information sources can surpass average accuracy. Multiple segmentations from different atlases were combined in [HHA+06].

Assume that N segmentation candidates of a target image X are computed from N atlases via (3). Let final target segmentation S_X be an image of the same size as X, where pixels take on values from the set $L_S=\{1,\ldots,N_S\}$ such that each discrete value corresponds to an organ or anatomical structure. An intuitive and straight-forward method to combine multiple segmentation estimates is then to choose the most frequent segmentation label (majority voting, MV) at each location p:

$$S_X^{\mathrm{MV}}(p) = \arg\max_{l_S \in L_S} \sum_n \delta\left(l_S, S_{X,n}(p)\right). \tag{4}$$

Such majority voting does not take into account the individual quality of each registration and therefore the resulting segmentation. We assume that post-registration image similarity between the deformed atlas and the target image is an indicator of segmentation reliabilty and can be used to locally assign weights w to each individual segmentation. The resulting weighted majority vote (wMV) can then be formalized as follows:

$$S_X^{\mathrm{wMV}}(p) = \arg\max_{l_S \in L_S} \sum_n w_n(p)\, \delta\left(l_S, S_{X,n}(p)\right). \tag{5}$$

To obtain the weights w, we use local normalized cross correlation (LNCC,[CBD+03]) between image X and deformed atlas $A_n(T_n)$. The advantages of LNCC are its smoothness and fast computation time due to convolution with Gaussian kernels:

$$\begin{aligned} \mathrm{LNCC}(X,Y,p) &= \frac{\langle X,Y\rangle(p)}{\sigma_X(p)\,\sigma_Y(p)} & \langle X,Y\rangle(p) &= \overline{X\cdot Y}(p) - \overline{X}(p)\cdot\overline{Y}(p) \\ \overline{X} &= \mathcal{G}_{\sigma_\mathcal{G}} * X & \sigma_X^2(p) &= \overline{X^2}(p) - \overline{X}^2(p), \end{aligned} \tag{6}$$

where $*$ is the convolution operator and $\mathcal{G}_{\sigma_\mathcal{G}}$ is a Gaussian kernel with standard deviation $\sigma_\mathcal{G}$. From the LNCC metric, we compute the weights:

$$w_n(p) = \left(\frac{1-\mathrm{LNCC}_\sigma(X, A_n(T_n), p)}{2}\right)^\gamma, \tag{7}$$

which normalizes LNCC to the range $[0,1]$. γ is used to scale the similarity such that contributions from individual segmentations are well spread [IK09].

3 Landmark Detection

For anatomical landmark detection, we use a template based approach from multiple atlases, the location estimates from which are fused based on their consensus. We localize each landmark ℓ separately from the others using the two stages below. To localize the unknown voxel coordinates p^ℓ of landmark ℓ in the target image X, we perform the following template matching procedure from each atlas A_n where n represents the atlas index.

3.1 Determining template and search regions

The template is set as a box-shaped image region A_n^ℓ in the current atlas. Similarly, a box-shaped search region X^ℓ is defined in the target image. Both such regions are chosen targeting a physically isotropic region of interest (ROI) in corresponding image, while limiting the maximum number of ROI voxels to ensure efficient computation. Specifics of ROI selection are given in Table 1.

Table 1: Specifics of template and search ROI, where $|\cdot|$ represents image dimensions (per axis).

ROI box (cropped image)	Centered at	Targeted half-width (d_{HW})	Max size ($n_{\max}$)
Template image A_n^ℓ	$p_{A_n}^\ell$	20 mm	41^3 voxels
Search region X^ℓ	$p_{A_n}^\ell \cdot \frac{\lvert X\rvert}{\lvert A_n\rvert}$	–	241^3 voxels

When setting the template size, the trade-off between it containing sufficient image features and final localization precision was considered. Template half-width was set to 20 mm empirically via cross-validation in multiple modalities using different template sizes (e.g., 10, 20, 30, ...). The search region is centered around a gross estimate of the landmark location, which is the normalized voxel coordinates of the landmark from the atlas; using the fact that both the atlas and the target have similar fields of view (ie. both abdomen, thorax, or whole body). Note that our large search region covers most or all of the image in many modalities (e.g. in MRce) or at least a quadrant thereof (e.g. in CT), such that the searched landmark can be guaranteed to exist therein.

3.2 Similarity metric

For template matching, the template is convolved over the search region by computing two independent similarity metrics, *sum of squared differences* (SSD) and *normalized cross-correlation* (NCC), at each template location i with respect to search image. Both values are then normalized linearly to $[0, 1]$ such that they are both 1 at the best match location. A combined similarity metric $\mathrm{SSD}^a \cdot \mathrm{COR}^b$ is then computed, where the parameters a=2 and b=3 were determined empirically via cross-validation with several powers. The maximum of this combined metric gives the best match location estimate p_n^ℓ for landmark ℓ considered atlas A_n.

3.3 Statistical fusion of estimate location from atlases

From cross-validation trials with different techniques such as the mean and weighted average of location estimates, the *median* operator was determined to be the best method for fusing location estimates. Accordingly, each axis coordinate of the the target landmark location is found as the median value of those axes from atlas estimates. The entire process can be summarized as:

- Repeat for each atlas A_n :
  - → Crop landmark template A_n^ℓ centered at given landmark location $p_{A_n}^\ell$ in atlas image A_n
  - → Crop a large search region X^ℓ centered around a grossly approximated location in X
  - → Compute $\mathrm{SSD}\big(A_n^\ell, X^\ell\big)$ and $\mathrm{COR}\big(A_n^\ell, X^\ell\big)$
  - → $p_n^\ell = \arg\max_i \big(\mathrm{SSD}[i]^2 \cdot \mathrm{COR}[i]^3\big)$
- $p^\ell = \mathrm{median}\left\{p_n^\ell \mid \forall_n\right\}$

4 Results and Discussion

Throughout the results, the following abbreviations are used for the image modalities: whole-body CT images (CT), thorax+abdomen contrast-agent CT images (CTce), abdominal T1-weighted contrast-agent MR images (MRce), and whole-body T1-weighted MR images (MR).

We treated each modality separately. We used the training images from Anatomy1 benchmark as atlases, ie. depending on the modality of the test image, six atlases for CTce and seven for CT, MR, MRce. For convenience, we combined all organ segmentations for each atlas into a single multi-label

segmentation image, which was then deformed using the atlas-to-target registration $\hat{T}$ described in Sec. 2.1. In Tab. 2, Dice overlap metric results regarding our segmentation approach reported by VISCERAL for the test images can be seen both for the VISCERAL Anatomy1 benchmark and the ISBI challenge.

In order to compare our technique to other participants' in the ISBI challenge, we computed an average rank per organ per participant. For given test image, we assigned a rank to each method, ie. $\{1,2,...,P\}$ where P is the number of participants submitted a (non-blank) output for that test image. In Fig. 1 the average of such ranks for all given test images is seen per anatomy. Our submission was *the only entrant that aimed to segment all images in all modalities* and it achieved competitive results for many organs as seen in the given figure. We did not plot ranks for the MR modalities, since we were the only participant to submit such results.

It is notable that the multi-atlas fusion has significantly lower segmentation accuracy for organs which has low volume, e.g. urinary, gall bladder, and the adrenal glands. One possible explanation is that such small structures are difficult to register using full-body images. Due to such misalignments, overlap between multiple deformed atlas segmentations for such label can be small, resulting in the weighted majority voting not selecting that label.

Table 2: Our segmentation overlap (Dice) results in VISCERAL Anatomy1 and ISBI challenges.

	Anatomy1				ISBI challenge			
Modality	Ctce	MR	CT	MRce	Ctce	MR	CT	MRce
Kidney (L)	0.903	0.730	0.805	0.782	0.885	0.548	0.756	0.888
Kidney (R)	0.877	0.733	0.754	0.787	0.827	0.589	0.679	0.732
Spleen	0.802	0.668	0.688	0.689	0.803	0.646	0.684	0.785
Liver	0.899	0.822	0.830	0.847	0.882	0.817	0.798	0.861
Lung (L)	0.961	0.533	0.952	0.650	0.960	0.486	0.955	
Lung (R)	0.968	0.900	0.960	0.664	0.966	0.909	0.965	
Urinary bladder	0.676	0.656	0.640	0.280	0.657	0.577	0.636	0.334
Lumbar Vertebra 1	0.604	0.396	0.350	0.060	0.548	0.623	0.333	0.084
Thyroid	0.252	0.367	0.469		0.315	0.488	0.439	
Pancreas	0.465		0.438	0.356	0.442		0.466	0.356
Psoas major (L)	0.811	0.801	0.772	0.644	0.797	0.765	0.773	0.654
Psoas major (R)			0.787				0.780	
Gallblader	0.334	0.023	0.102	0.035	0.212	0.044	0.078	0.000
Sternum	0.595	0.358	0.648		0.612	0.359	0.630	
Aorta	0.785	0.744	0.723	0.616	0.787	0.783	0.724	
Trachea	0.847	0.736	0.822		0.839	0.747	0.837	
Adrenal gland (L)	0.204	0.109	0.165	0.000	0.099	0.144	0.282	
Adrenal gland (R)	0.164	0.215	0.138	0.107	0.019	0.268	0.133	

5 Conclusions

In the Anatomy1 and ISBI challenges organized by VISCERAL project, our landmark localization achieved in whole-body CT images an impressive 11 and 13 voxel average error, respectively in these challenges. No comparison to alternatives was possible since ours was the only entrant in landmark localization in both challenges. For segmentation, our multi-atlas based method that

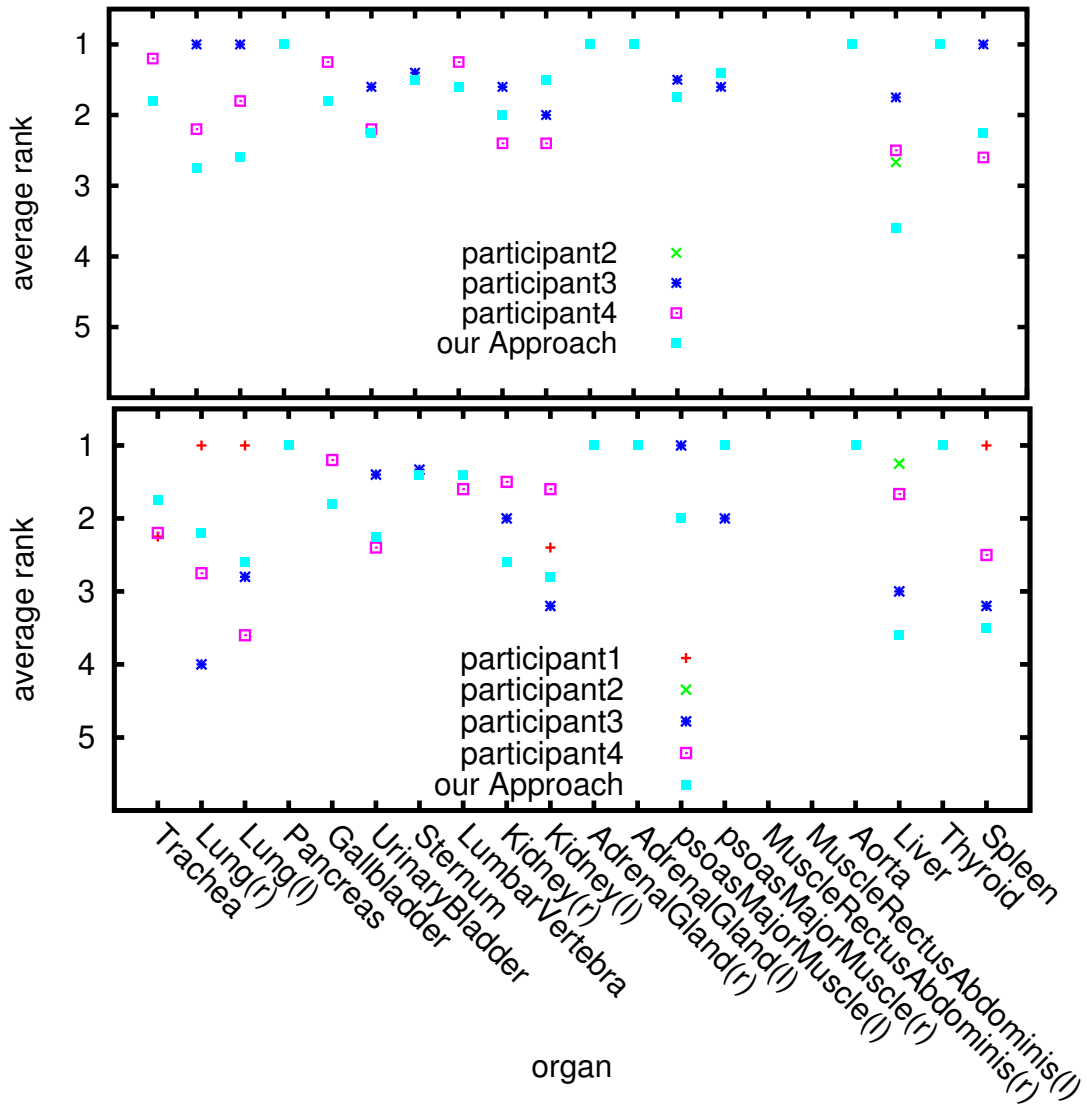

Figure 1: Average segmentation rank (by Dice coefficient) for each participant and organ for the CT (top) and CTce (bottom) modalities.

does not require any customisation to a specific modality or organ competed in all categories and ranked satisfactorily compared to the results of other participants.

References

[CBD+03] Pascal Cachier, Eric Bardinet, Didier Dormont, Xavier Pennec, and Nicholas Ayache. Iconic feature based nonrigid registration: the PASHA algorithm. *Computer Vision and Image Understanding*, 89(2-3):272–298, February 2003.

[GKT+08] Ben Glocker, Nikos Komodakis, Georgios Tziritas, Nassir Navab, and Nikos Paragios. Dense image registration through MRFs and efficient linear programming. *Medical Image Analysis*, 12(6):731–741, December 2008.

[HHA+06] Rolf a Heckemann, Joseph V Hajnal, Paul Aljabar, Daniel Rueckert, and Alexander Hammers. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage*, 33(1):115–26, October 2006.

[IK09] Juan Eugenio Iglesias and Nico Karssemeijer. Robust initial detection of landmarks in film-screen mammograms using multiple FFDM atlases. *IEEE Transactions on Medical Imaging*, 28(11):1815–24, November 2009.

[Kol06] Vladimir Kolmogorov. Convergent Tree-Reweighted Message Passing for Energy Minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28:1568–1583, 2006.

[RAH+06] Daniel Rueckert, Paul Aljabar, Rolf a Heckemann, Joseph V Hajnal, and Alexander Hammers. Diffeomorphic registration using B-splines. In *Medical Image Computing and Computer-Assisted Intervention*, pages 702–709, January 2006.