

**VISCERAL@ISBI 2015**

**VISCERAL Anatomy3 Organ Segmentation Challenge**

co-located with IEEE International Symposium on Biomedical Imaging 2015

New York, NY, USA, April 16, 2015

**Proceedings**



Orcun Goksel  
Oscar Alfonso Jiménez del Toro  
Antonio Foncubierta Rodríguez  
Henning Müller  
(Eds.)

© 2015 for the individual papers by the papers' authors.  
Copying permitted only for private and academic purposes.  
Re-publication of material from this volume requires permission by the copyright owners.

*Corresponding editor's address:*  
Prof. Dr. Orcun Goksel  
ETH Zürich – Computer Vision Lab  
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 in images with wide field-of-view (e.g. showing entire abdomen, trunk, or the whole body) has been limited. VISCERAL Anatomy benchmark series aim to address this need by providing a common image and test dataset and corresponding segmentation challenges for a wide range of anatomical structures and image modalities. This proceedings summarize the techniques submitted for Anatomy<sup>3</sup> benchmark, the results of which were also presented at the ISBI VISCERAL Challenge session on April 16<sup>th</sup> 2014, as part of the IEEE International Symposium on Biomedical Imaging (ISBI) in New York, NY, USA.

The challenge participants used an online evaluation system, where they submitted their algorithms in a virtual machine environment. The organisers then run the virtual machines on the test images and populated the segmentation results in a participant viewable results board. Then, the participants could at their discretion upload their results to a public leaderboard. The results from the methods presented here were published in the online leaderboard two weeks before the challenge session.

The short papers in this proceedings were submitted by the participants to describe their specific methodologies used to generate their results. At the session, participants had a chance to present their methods as oral presentations.

We thank the authors for their submissions and the program committee for their hard work.

Proceeding editors  
On behalf of VISCERAL Consortium

**Session Chair**

Orçun Göksel, ETH Zürich, Switzerland

**Proceedings Editors**

Orçun Göksel, ETH Zürich, Switzerland

Oscar Jiménez, University of Applied Sciences Western Switzerland, Switzerland

Antonio Foncubierta, ETH Zürich, Switzerland

Henning Müller, University of Applied Sciences Western Switzerland, Switzerland

**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

Bjoern Menze, Munich University of Technology, Germany

Marc-André Weber, University of Heidelberg, Germany

Tomàs Salas Fernandez, Catalan Agency for Health Information, Assessment and Quality, Spain

## Contents

### PART I: ORGANIZATION AND EVALUATION

#### Overview of the VISCERAL Challenge at ISBI 2015

*Orcun Goksel, Antonio Foncubierta-Rodríguez, Oscar Alfonso Jiménez del Toro, Henning Müller, Georg Langs, Marc-André Weber, Bjoern Menze, Ivan Eggel, Katharina Gruenberg, Marianne Winterstein, Markus Holzer, Markus Krenn, Georgios Kontokotsios, Sokratis Metallidis, Roger Schaer, Abdel Aziz Taha, András Jakab, Tomàs Salas Fernandez, Allan Hanbury*

6

### PART II: PARTICIPANT SUBMISSIONS

#### Good Features for Reliable Registration in Multi-Atlas Segmentation

*Fredrik Kahl, Jennifer Alvé, Olof Enqvist, Frida Fejne, Johannes Ulé, Johan Fredriksson, Matilda Landgren, Viktor Larsson*

12

#### Fully Automatic Multi-Organ Segmentation Based on Multi-Boost Learning and Statistical Shape Model Search

*Baochun He, Cheng Huang, Fucang Jia*

18

#### Hierarchic Anatomical Structure Segmentation Guided by Spatial Correlations (AnatSeg-Gspac): VISCERAL Anatomy3

*Oscar Alfonso Jiménez del Toro, Yashin Dicente Cid, Adrien Depeursinge, Henning Müller*

22

#### Multi-modal Multi-Atlas Segmentation using Discrete Optimisation and Self-Similarities

*Mattias P. Heinrich, Oskar Maier, Heinz Handels*

27

#### Efficient and fully automatic segmentation of the lungs in CT volumes

*Yashin Dicente Cid, Oscar Alfonso Jiménez del Toro, Adrien Depeursinge, Henning Müller*

31

# Overview of the VISCERAL Challenge at ISBI 2015

Orcun Goksel<sup>1</sup>, Antonio Foncubierta-Rodríguez<sup>1</sup>, Oscar Alfonso Jiménez del Toro<sup>2</sup>, Henning Müller<sup>2</sup>, Georg Langs<sup>3</sup>, Marc-André Weber<sup>4</sup>, Bjoern Menze<sup>5</sup>, Ivan Eggel<sup>2</sup>, Katharina Gruenberg<sup>4</sup>, Marianne Winterstein<sup>4</sup>, Markus Holzer<sup>3</sup>, Markus Krenn<sup>3</sup>, Georgios Kontokotsios<sup>6</sup>, Sokratis Metallidis<sup>2</sup>, Roger Schaer<sup>2</sup>, Abdel Aziz Taha<sup>6</sup>, András Jakab<sup>3</sup>, Tomàs Salas Fernandez<sup>7</sup>, Allan Hanbury<sup>6</sup>

ETH Zürich, Switzerland<sup>1</sup>; HES-SO Valais, Sierre, Switzerland<sup>2</sup>;  
MUW, Vienna, Austria<sup>3</sup>; University of Heidelberg, Germany<sup>4</sup>; TUM, Munich, Germany<sup>5</sup>;  
TUWien, Vienna, Austria<sup>6</sup>; AQuAS, Barcelona, Spain<sup>7</sup>

## Abstract

This is an overview paper describing the data and evaluation scheme of the VISCERAL Segmentation Challenge at ISBI 2015. The challenge was organized on a cloud-based virtual-machine environment, where each participant could develop and submit their algorithms. The dataset contains up to 20 anatomical structures annotated in a training and a test set consisting of CT and MR images with and without contrast enhancement. The test-set is not accessible to participants, and the organizers run the virtual-machines with submitted segmentation methods on the test data. The results of the evaluation are then presented to the participant, who can opt to make it public on the challenge leaderboard displaying 20 segmentation quality metrics per-organ and per-modality. Dice coefficient and mean-surface distance are presented herein as representative quality metrics. As a continuous evaluation platform, our segmentation challenge leaderboard will be open beyond the duration of the VISCERAL project.

## 1 Introduction

In this challenge, a set of annotated medical imaging data was provided to the participants, along with a powerful complimentary cloud-computing instance (8-core CPU with 16GB RAM) where participant algorithms can be developed and evaluated. The available data contains segmentations of several different anatomical structures in different image modalities, e.g. CT and MRI. Annotated

---

*Copyright © by the paper's authors. Copying permitted only for private and academic purposes.*

In: O. Goksel (ed.): Proceedings of the VISCERAL Anatomy Grand Challenge at the 2015 IEEE International Symposium on Biomedical Imaging (ISBI), New York, NY, Apr 16<sup>th</sup>, 2015 published at <http://ceur-ws.org>

structures in the training and testing data corpus included the segmentations of left/right kidney, spleen, liver, left/right lung, urinary bladder, rectus abdominis muscle, 1<sup>st</sup> lumbar vertebra, pancreas, left/right psoas major muscle, gallbladder, sternum, aorta, trachea, left/right adrenal gland.

As training, 20 volumes each were provided for four different image modalities and field-of-views, with and without contrast enhancement, which add up to 80 volumes in total. In each volume, up to 20 structures were segmented. The missing annotations are due to poor visibility of the structures in certain image modalities or due to such structures being outside the field-of-view. Accordingly, in all 80 volumes, a total of 1295 structures are segmented. A breakdown of annotations per anatomy can be seen in Figure 1.

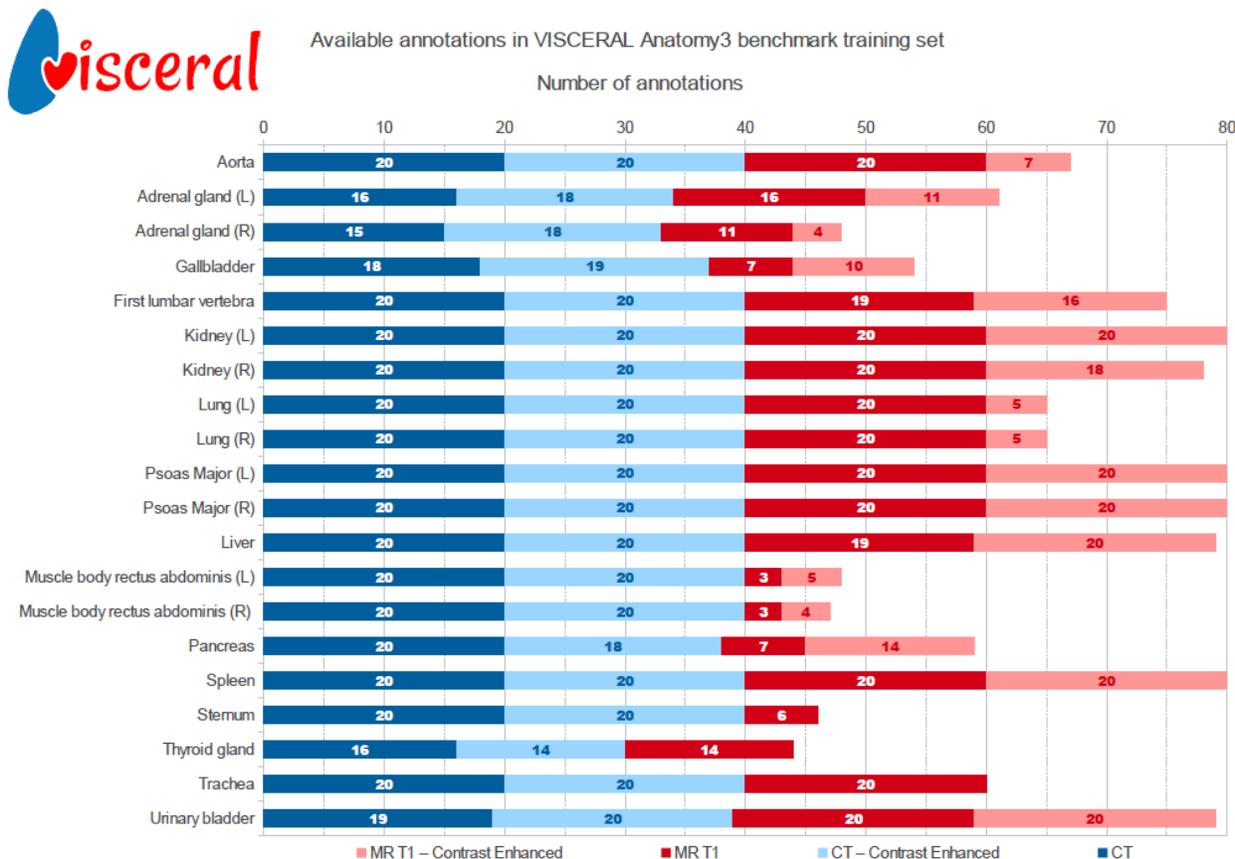


Figure 1: Number of annotations in the Anatomy3 training set classified by modality and organ.

Participants did not need to segment all the structures involved in such data, but rather they could attempt any single anatomical structure or a combination thereof. For instance, an algorithm that could segment only some organs in some of the modalities was evaluated only in those categories for which it outputted any results. Accordingly, our evaluation results were presented in a per-anatomy, per-modality fashion depending on the attempted segmentation task/s by each participating algorithm. This is, indeed, in line with the VISCERAL vision of creating a single, large, and multi-purpose medical image dataset, on which different research groups can test their specific applications and solutions.

Participants first registered for a benchmark account at the VISCERAL registration website. Among the options during the registration, they could request their choice of operating system (Linux, Windows, etc) for the virtual machine (VM), in order to get access to the VM and the data. Having signed the data usage agreement and uploaded it to the participant dashboard, they

could then access the VM for algorithm development and also use the training data accessible therein. Participants could additionally download the training dataset via FTP for offline training.

Participants accordingly developed and installed their algorithms in the VM, while adapting and testing them on the training data. They then prepared their executable on the VM according to the input/output specifications announced by us earlier in the Anatomy3 Guidelines for Participation, and submitted their VMs (through "Submit VM" button in the online participant dashboard) for evaluation on the test data. We subsequently ran their VM (and hence their algorithm) on the test data, and computed the relevant metrics. This evaluation process could be performed several times during the training phase, nevertheless, we limited submissions to once per week, in order to prevent the participants "training on the test data". The participants received feedback from their evaluations in a private leaderboard and had the option to make their results publicly available on the online public leaderboard, which included the results considered in our benchmark results.

## 2 Evaluation

For the Anatomy3 benchmark, a different evaluation approach was implemented compared to the previous Anatomy benchmarks [LMMH13, JdTGM<sup>+</sup>14]. For this benchmark, participants had the opportunity to submit their algorithms several times, giving them the opportunity to improve their algorithms prior to the final evaluation analysis during ISBI 2015. They could also choose to make any of their results from the test-set public at any time. To allow a continuous workflow with this evaluation approach, the steps during the evaluation phase were automated to a large extent.

The continuous evaluation approach included the following steps:

1. The participant registers for the challenge; fills, signs, and uploads the participant agreement.
2. The organizers provide the participant with a virtual machine (VM) from the VISCERAL cloud infrastructure.
3. The participant implements a segmentation algorithm in the VM according to the benchmark specifications.
4. The VM is submitted by the participant using the participant dashboard.
5. The organizers isolate the VM to prevent the participant from accessing it during the evaluation phase.
6. The participant executable is run in a batch-script to test if its output files correspond with those expected by the evaluation routines.
7. If the previous step is successful, the evaluation proceeds for all the volumes in the test set.
8. Each generated output segmentation file is uploaded by the batch script to the cloud storage reserved for that participant.
9. Once all the images in the test-set are processed by the participant executable, the output segmentations are cleared from the VM, which is in turn returned to the participant.
10. The output segmentations uploaded in the cloud storage are then evaluated against the ground-truth (manual annotations) and the results are presented in the participant dashboard.
11. The participant can then analyze and interpret the results of their submission, and choose to make them public or not on the public leaderboard.
12. The participant is allowed to submit again for testing only after a minimum of one week from their latest submission.

Participant	Affiliation	Modality	Submission Timestamp	Left Kidney	Right Kidney	Spleen	Liver	Left Lung	Right Lung	Urinary Bladder	Muscle Body Of Left Rectus Abdominis	Muscle Body Of Right Rectus Abdominis	Lumbar Vertebra 1	Thyroid	Pancreas	Left Psoas Major Muscle	Right Psoas Major Muscle	Gallbladder	Aorta	Sternum	Trachea	Left Adrenal Gland	Right Adrenal Gland
Oscar Jimenez	HES-SO	CTce	28 Mar 2015 07:42:00	0.91	0.889	0.73	0.887	0.959	0.963	0.679	0.474	0.453	0.523	0.41	0.423	0.794	0.799	0.484	0.721	0.762	0.855	0.331	0.342
Oscar Jimenez	HES-SO	CT	28 Mar 2015 07:42:00	0.784	0.79	0.703	0.866	0.972	0.975	0.698	0.551	0.519	0.718	0.549	0.408	0.806	0.787	0.276	0.761	0.753	0.92	0.373	0.355
Dr. Mattias Heinrich	University of Luebeck	MRT1cefs	25 Mar 2015 15:10:00	0.831	0.778	0.708	0.788			0.356						0.78	0.758						
Chunliang Wang	Linköping University	CT	06 Mar 2015 13:00:00	0.896	0.796	0.91	0.936	0.961	0.97	0.713						0.828	0.817						
Chunliang Wang	Linköping University	CTce	06 Mar 2015 13:00:00	0.945	0.959	0.909	0.949	0.972	0.971	0.856						0.83	0.845						
Fredrik Kahl	Signals and Systems, Chalmers University of Technology	CT	23 Mar 2015 08:00:00			0.87	0.921			0.763				0.424	0.383			0.19				0.282	0.22
Fredrik Kahl	Signals and Systems, Chalmers University of Technology	CT	29 Mar 2015 13:00:00	0.934	0.915			0.972	0.975		0.746	0.679	0.775			0.861	0.847		0.847	0.83	0.931		
Yashin Dicente Cid	HES-SO	CT	30 Mar 2015 06:30:00					0.972	0.974														
Yashin Dicente Cid	HES-SO	CTce	30 Mar 2015 06:30:00					0.974	0.973														

Figure 2: A snapshot of VISCERAL Anatomy3 public leaderboard at the time of ISBI 2015 challenge.

### 3 Benchmark results

Detailed results from 20 metrics can be seen in the online leaderboard<sup>1</sup>, a snapshot of which at the time of ISBI 2015 Anatomy3 challenge is shown in Fig. 2. Participant evaluation results are summarized in tables 1 and 2, respectively for Dice coefficient and mean surface distance, as commonly-used segmentation evaluation metrics. The former is an overlap metric, describing how well an algorithm estimates target anatomical region. The latter is a surface distance metric, summarizing the overall surface estimation error by a given algorithm. The participant row in the tables contains the citation for the publication contribution within this Anatomy3 proceedings Part II.

In the Dice results table, the highest ranking methods per-modality per-organ are marked in bold. Any other method within 0.01 (1%) Dice of this are also considered a winner (or a tie) due to the insignificance of the difference. Dice values below a threshold are considered unsuccessful segmentations, and thus are not declared as a winner – even though the reader should note that depending on particular clinical application such results can potentially still be useful. This threshold was selected as 0.6 Dice, coinciding with a gap in the reported participant results.

The results corresponding to the same bold values in the Dice table are also marked in the mean surface distance table, in order to facilitate comparison of the segmentation surface errors for the best methods in terms of the Dice metric. For successfully segmented organs (defined by the empirical 0.6 Dice cutoff), both metrics agree on the results for all structures and modalities, except for the first lumbar vertebra in CT. The reader should note that the mean surface distances are presented in voxels, therefore the values between modalities (e.g. MR-ce and CT) are not directly comparable in the latter table.

According to these tables, there are different algorithms performing well for different anatomy. In contrast-enhanced MR modality, we had only a single participant, Heinrich *et al.*, potentially

<sup>1</sup> The leaderboard is accessible at <http://visceral.eu:8080/register/Leaderboard.xhtml>

MODALITY	MR ce	CT contrast-enhanced (ce)			CT			
PARTICIPANT	Heinrich <i>et al.</i>	Jiménez <i>et al.</i>	He <i>et al.</i>	Cid <i>et al.</i>	Kahl <i>et al.</i>	Jiménez <i>et al.</i>	He <i>et al.</i>	Cid <i>et al.</i>
Left kidney	<b>0.862</b>	<b>0.91</b>	<b>0.91</b>	-	<b>0.934</b>	0.784	-	-
Right kidney	<b>0.855</b>	0.889	<b>0.922</b>	-	<b>0.915</b>	0.79	-	-
Spleen	<b>0.724</b>	0.73	<b>0.896</b>	-	<b>0.87</b>	0.703	<b>0.874</b>	-
Liver	<b>0.837</b>	0.887	<b>0.933</b>	-	<b>0.921</b>	0.866	<b>0.923</b>	-
Left lung	-	0.959	<b>0.966</b>	<b>0.974</b>	<b>0.972</b>	<b>0.972</b>	0.952	<b>0.972</b>
Right lung	-	<b>0.963</b>	<b>0.966</b>	<b>0.973</b>	<b>0.975</b>	<b>0.975</b>	0.957	<b>0.974</b>
Bladder	0.494	<b>0.679</b>	-	-	<b>0.763</b>	0.698	-	-
Pancreas	-	0.423	-	-	0.383	0.408	-	-
Gallbladder	-	0.484	-	-	0.19	0.276	-	-
Thyroid	-	0.41	-	-	0.424	0.549	-	-
Aorta	-	<b>0.721</b>	-	-	<b>0.847</b>	0.761	-	-
Trachea	-	<b>0.855</b>	-	-	<b>0.931</b>	0.92	-	-
Sternum	-	<b>0.762</b>	-	-	<b>0.83</b>	0.753	-	-
1 <sup>st</sup> lumbar vertebra	-	0.523	-	-	<b>0.775</b>	0.718	-	-
Left adrenal gland	-	0.331	-	-	0.282	0.373	-	-
Right adrenal gland	-	0.342	-	-	0.22	0.355	-	-
Left psoas major	<b>0.801</b>	<b>0.794</b>	-	-	<b>0.861</b>	0.806	-	-
Right psoas major	<b>0.772</b>	<b>0.799</b>	-	-	<b>0.847</b>	0.787	-	-
Left rectus abdominis	-	0.474	-	-	<b>0.746</b>	0.551	-	-
Right rectus abdominis	-	0.453	-	-	<b>0.679</b>	0.519	-	-

Table 1: Segmentation results in terms of DICE coefficient classified by modality and organ.

MODALITY	MR ce	CT contrast-enhanced (ce)			CT			
PARTICIPANT	Heinrich <i>et al.</i>	Jiménez <i>et al.</i>	He <i>et al.</i>	Cid <i>et al.</i>	Kahl <i>et al.</i>	Jiménez <i>et al.</i>	He <i>et al.</i>	Cid <i>et al.</i>
Left kidney	<b>0.251</b>	<b>0.172</b>	<b>0.171</b>	-	<b>0.147</b>	1.209	-	-
Right kidney	<b>0.3</b>	0.243	<b>0.131</b>	-	<b>0.229</b>	1.307	-	-
Spleen	<b>1.138</b>	2.005	<b>0.385</b>	-	<b>0.534</b>	1.974	<b>0.36</b>	-
Liver	<b>0.935</b>	0.514	<b>0.203</b>	-	<b>0.299</b>	0.78	<b>0.239</b>	-
Left lung	-	0.071	<b>0.069</b>	<b>0.05</b>	<b>0.045</b>	<b>0.043</b>	0.101	<b>0.05</b>
Right lung	-	<b>0.065</b>	<b>0.078</b>	<b>0.052</b>	<b>0.043</b>	<b>0.038</b>	0.094	<b>0.046</b>
Bladder	2.632	<b>1.879</b>	-	-	<b>1.057</b>	1.457	-	-
Pancreas	-	3.804	-	-	4.478	5.521	-	-
Gallbladder	-	3.603	-	-	9.617	5.938	-	-
Thyroid	-	3.337	-	-	2.163	1.466	-	-
Aorta	-	<b>0.899</b>	-	-	<b>0.542</b>	0.938	-	-
Trachea	-	<b>0.223</b>	-	-	<b>0.083</b>	0.103	-	-
Sternum	-	<b>1.094</b>	-	-	<b>0.798</b>	1.193	-	-
1 <sup>st</sup> lumbar vertebra	-	4.504	-	-	<b>2.424</b>	1.953	-	-
Left adrenal gland	-	3.115	-	-	3.298	2.672	-	-
Right adrenal gland	-	2.66	-	-	7.046	3.445	-	-
Left psoas major	<b>0.493</b>	<b>0.742</b>	-	-	<b>0.443</b>	0.595	-	-
Right psoas major	<b>0.569</b>	<b>0.757</b>	-	-	<b>0.55</b>	0.775	-	-
Left rectus abdominis	-	6.068	-	-	<b>1.614</b>	0.355	-	-
Right rectus abdominis	-	6.6	-	-	<b>1.922</b>	4.032	-	-

Table 2: Segmentation results in terms of mean surface distance in pixels (which may have different physical meanings based on the resolution of a particular modality).

due to the difficulty of automatic segmentations in this modality. This group thus became the unchallenged winner of MRce for the structures they participated in. Note that the surface error results are reported in voxels, where MRce has a significantly lower resolution than the other modalities. In CTce, He *et al.* performed the best for the 6 structures they participated in, with some ties with Jimenez *et al.* The latter group segmented all the given structures in CTce, some of them with satisfactory accuracy, while for the others with potentially unusable results. We had the most participants for the CT modality, in which the lungs –a relatively easier segmentation problem– were segmented successfully by most participants; potentially close to the accuracy of inter-subject annotations. For most other structures for which successful segmentations were achieved in CT, Kahl *et al.* were the winner of the challenge. Nevertheless, for structures where lower fidelity segmentations (below the 0.6 Dice cutoff) were attained, Jimenez *et al.* are seen to provide better segmentations estimations; likely due to their segmentation approach being atlas-based. It is also observed that, despite the relatively good contrast of CT, several structures (prominently the pancreas, gallbladder, thyroid, and adrenal glands) are still quite challenging to segment from CT — potentially due to the lower sensitivity of CT to those structures also complicated by the difficult-to-generalize shapes of these anatomies.

## 4 Conclusions

The VISCERAL Anatomy3 Challenge had a total of 23 virtual machines allocated for participants at a time, although not all participants ultimately submitted results for the challenge. Most participants relied on atlas-based segmentation methods, although there were also techniques that use anatomy-based reasoning and locational relations. By using an online leaderboard evaluation method, more participants are expected to submit results for our Anatomy3 challenge in the future.

## 5 Acknowledgments

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2014) under grant agreement n° 318068 VISCERAL.

## References

- [JdTGM<sup>+</sup>14] Oscar Alfonso Jiménez del Toro, Orcun Goksel, Bjoern Menze, Henning Müller, Georg Langs, Marc-André Weber, Ivan Eggel, Katharina Gruenberg, Markus Holzer, Georgios Kotsios-Kontokotsios, Markus Krenn, Roger Schaer, Abdel Aziz Taha, Marianne Winterstein, and Allan Hanbury. VISCERAL – VISual Concept Extraction challenge in RAdioLogY: ISBI 2014 challenge organization. In Orcun Goksel, editor, *Proceedings of the VISCERAL Challenge at ISBI*, number 1194 in CEUR Workshop Proceedings, pages 6–15, Beijing, China, May 2014.
- [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.

# Good Features for Reliable Registration in Multi-Atlas Segmentation

Fredrik Kahl<sup>1,2</sup>    Jennifer Alvé<sup>1</sup>    Olof Enqvist<sup>1</sup>    Frida Fejne<sup>1</sup>  
Johannes Ulén<sup>2</sup>    Johan Fredriksson<sup>2</sup>    Matilda Landgren<sup>2</sup>    Viktor Larsson<sup>2</sup>

<sup>1</sup>Department of Signals and Systems  
Chalmers University of Technology, Sweden

<sup>2</sup>Centre for Mathematical Sciences  
Lund University, Sweden

## Abstract

This work presents a method for multi-organ segmentation in whole-body CT images based on a multi-atlas approach. A robust and efficient feature-based registration technique is developed which uses sparse organ specific features that are learnt based on their ability to register different organ types accurately. The best fitted feature points are used in RANSAC to estimate an affine transformation, followed by a thin plate spline refinement. This yields an accurate and reliable nonrigid transformation for each organ, which is independent of initialization and hence does not suffer from the local minima problem. Further, this is accomplished at a fraction of the time required by intensity-based methods. The technique is embedded into a standard multi-atlas framework using label transfer and fusion, followed by a random forest classifier which produces the data term for the final graph cut segmentation. For a majority of the classes our approach outperforms the competitors at the VISCERAL Anatomy Grand Challenge on segmentation at ISBI 2015.

## 1 Introduction

Segmentation is a key problem in medical image analysis, and may be used for numerous applications in medical research and clinical care. In this paper, a pipeline for the segmentation of whole-body CT images into 20 different organs is presented. The approach is based on multi-atlas segmentation, see [KSK<sup>+</sup>10, HKA<sup>+</sup>10, WSD<sup>+</sup>13] and the references therein, an approach which is known to produce state-of-the-art results for several segmentation tasks. The method requires pair-wise registrations from a set of atlas images to the unknown target image.

---

*Copyright © by the paper's authors. Copying permitted only for private and academic purposes.*

In: O. Goksel (ed.): Proceedings of the VISCERAL Anatomy Grand Challenge  
at the 2015 IEEE International Symposium on Biomedical Imaging (ISBI), New York, NY, Apr 16<sup>th</sup>, 2015  
published at <http://ceur-ws.org>

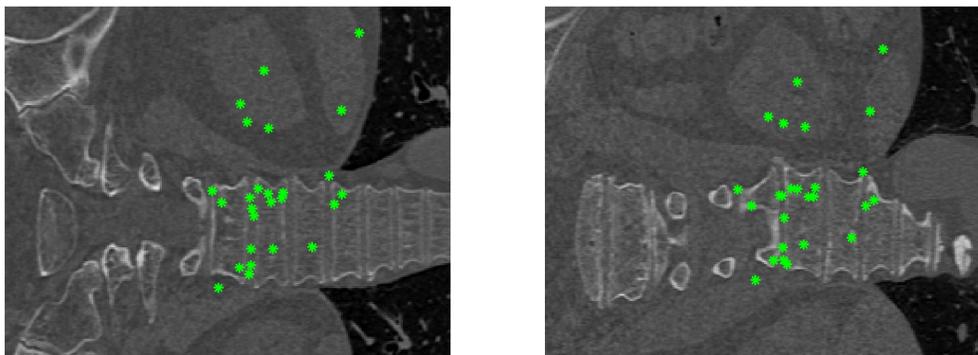


Figure 1: Two CT slices of a target (*left*) and atlas (*right*) with corresponding features after RANSAC for Lumbar Vertebra 1.

In principle, there are two different approaches to image registration, feature-based and intensity-based registration, see the surveys [KBG<sup>+</sup>11, SDP13]. Intensity-based methods are capable of producing accurate registrations but are sensitive to initialization and often slow. Feature-based methods are usually faster, but may risk failing due to many outlier correspondences between the images. Our approach is an adapted feature-based method that utilizes the speed of general feature-based methods while trying to eliminate the risk of establishing incorrect point-to-point correspondences between the images by identifying reliable feature points. We show that reliable organ localization can be computed using (i) robust optimization techniques and (ii) learned feature correspondences.

## 2 Proposed Solution

Our system segments each organ independently of each other using a multi-atlas approach. The pipeline has three steps:

1. Feature-based registration with RANSAC.
2. Label fusion with a random forest classifier.
3. Graph cut segmentation with a Potts model.

These steps will now be described in more detail.

**1. Feature-based registration with RANSAC.** In order to register an atlas image to the target, a feature-based approach is used. Sparse features are extracted according to Svärm et al. [SEKO15], which uses a method similar to SIFT for feature detection and SURF for feature description. Typically around 8,000–10,000 features are extracted from a  $512 \times 512 \times 800$  CT image, which takes less than 30s. Correspondences are obtained by matching a *subset* of the features in the atlas image to the features in the target. The matching is done with a symmetric neighbor approach, where each descriptor is matched to its nearest neighbor. The organ-specific subset of atlas features is determined as a pre-processing step in the following way. For each atlas image and organ, *golden transformations* are established to the other atlas images using the ground truth segmentations. Then, based on these transformations, one can check which features in the atlas image of interest that are most accurate, and rank the features accordingly. We have found empirically that using the top 300 best features for each organ provides robust and reliable registration. Standard RANSAC

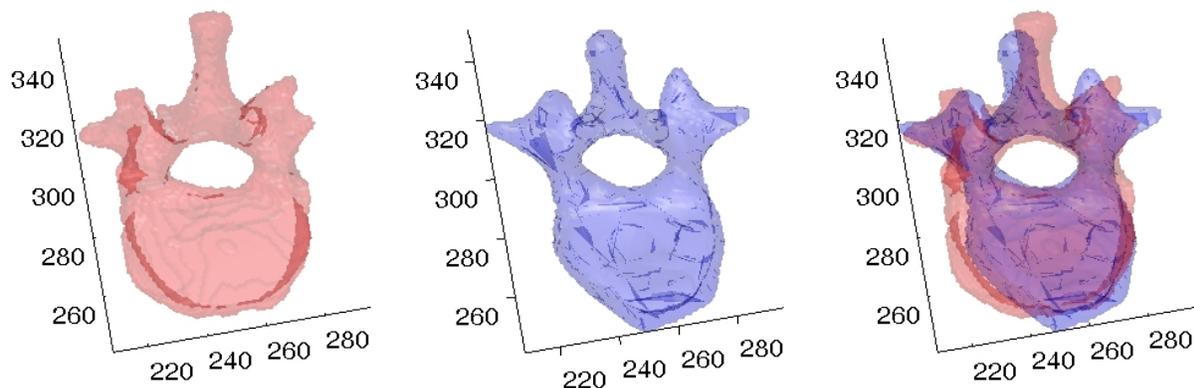


Figure 2: Example registration of Lumbar Vertebra 1. *Left*: the atlas ground truth mask. *Middle*: the TPS warped target mask. *Right*: The masks overlaid in the same coordinate system.

with the truncated  $\ell_2$  as cost function is used in order to remove outliers. The optimization is run 500,000 iterations and the truncation threshold was set to 30 mm. See Figure 1 for an example.

Finally, a coordinate transformation from the atlas to the target image is computed by applying thin plate splines (TPS) to the remaining correspondences, and thereafter used in order to transfer the labels of the atlas to the target image. The thin plate spline method proposed in [CR00] was used. One registration takes less than 10s in total. See Figure 2 for an example.

For several of the organs, the registrations are refined with a standard intensity-based method. More specifically, we used `NiftyReg` [ORS<sup>+</sup>01] which takes around 100–200s per registration.

**2. Label fusion with a random forest classifier.** The pairwise feature-based registrations give us a rough idea where the organ is located. In order to fuse the transferred labels, one can compute an average voxel map, denoted  $P$ . Hence, for voxel  $i$ , the map  $P(i)$  gives a number between 0 and 1 which can (intuitively) be interpreted as the probability of voxel  $i$  belonging to the organ. For example, if half of the atlas images say that voxel  $i$  should be organ, then  $P(i) = 0.5$ .

The map  $P$  largely ignores the local appearance around the target organ and in order to improve the accuracy of the estimate, the map  $P$  along with a few other features is used to train a random forest classifier. We use `Sherwood` [CSK11] to train and evaluate large random forest instances efficiently. Given a target volume  $I$  and the map  $P$ , we begin by smoothing both using a Gaussian kernel with standard deviation  $\sigma = 1$ , resulting in two new volumes denoted  $I_s$  and  $P_s$ . Using cross validation we also determine a threshold level,  $\tau$ , for  $P$  and construct a distance map  $D$ , where each voxel in  $D$  equals the (signed) distance to the boundary surface of the binary volume  $P > \tau$ , that is, the map  $P$  thresholded at  $\tau$ . For each volume  $I$  we thus obtain 5 features per voxel  $i$ :  $I(i)$ ,  $I_s(i)$ ,  $P(i)$ ,  $P_s(i)$  and  $P_t(i)$ . The output of the classifier produces yet another map, denoted  $P_r$ , which is a refined estimate of the location of the organ. As previously,  $P_r$  can be interpreted as the probability for each voxel belonging to the organ of interest. See Figure 3 for an example.

**3. Graph cut segmentation with a Potts model.** The random forest classifier generates a new estimate  $P_r(i)$  for each voxel  $i$ , but each decision in the classifier is taken independently of the output of the neighboring voxels of  $i$ . This has a tendency of producing noisy boundary estimates and therefore we will regularize the solution by using a standard Potts model. The final solution can then be computed with graph-cuts.

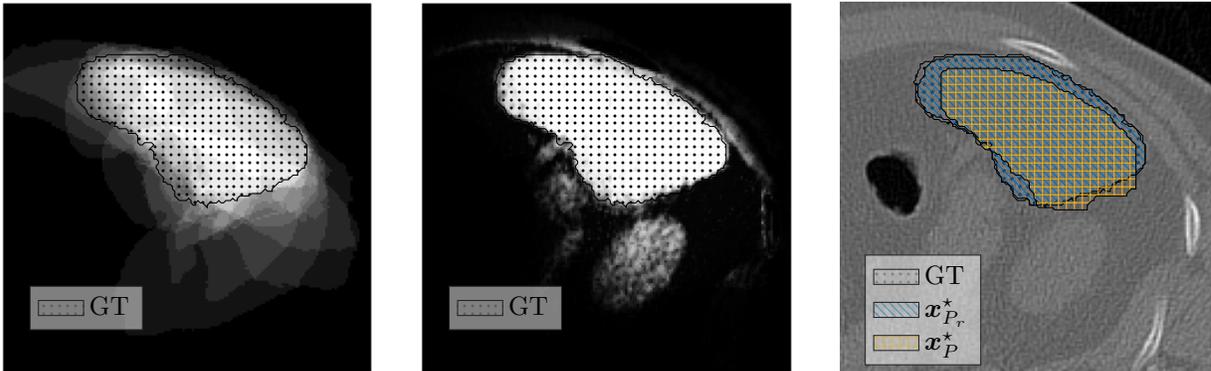


Figure 3: Example of the resulting probability estimates and segmentation of the spleen for one CT slice; in each image the ground truth (GT) is indicated. *Left*: the initial probability,  $P$ . *Middle*: the probability given by random forest,  $P_r$ . *Right*: the resulting segmentation  $\mathbf{x}_P^*$  using  $P$  and  $\mathbf{x}_{P_r}^*$  using  $P_r$  overlaid on the original image.

The Potts model penalizes neighboring voxels if they take different labels. Let  $x_i$  be a Boolean indicator variable for voxel  $i$ , i.e.,  $x_i \in \{0, 1\}$ . Then, for two neighboring voxels  $x_i$  and  $x_j$ , the cost should be zero if  $x_i = x_j$  and  $\lambda$  otherwise, where  $\lambda$  is a positive scalar. This cost can compactly be written as  $\lambda x_i(1 - x_j)$ . Further, the data cost for voxel  $i$  is set to take value  $1/2 - P_r(i)$  if  $x_i = 1$  and zero otherwise. This favors voxels with probabilities in the interval  $[0.5, 1]$  to be foreground and voxels with  $[0, 0.5]$  to be background.

In summary the final segmentation,  $\mathbf{x}^*$ , is given by the solution to the optimization problem:

$$\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x} \in \{0,1\}^n} \sum_{i=1}^n x_i \left( \frac{1}{2} - P_r(i) \right) + \lambda \sum_{i=1}^n \sum_{j \in \mathcal{N}(i)} \mu_{ij} x_i (1 - x_j), \quad (1)$$

where  $\lambda$  is a regularization weight and  $\mu_{ij}$  compensates for anisotropic resolution. For all organs we use a 6-connected neighborhood  $\mathcal{N}$ . In order to save memory and speed-up calculations we only process a volume around the zero level of the distance map  $D$  with a 20 voxels margin. The function in (1) is submodular and is minimized efficiently using the graph-cut implementation of [JSH12].

### 3 Experimental Results

All the tuning parameters in our system have been set by leave-one out cross validation on the first 15 of the 20 whole-body CT images available in the VISCERAL challenge. The 5 remaining images have been used to validate the performance of the random forest classifier and the graph-cut segmentation. In the training phase, the first 15 images have served as the atlas set, while in the final version all 20 images are utilized in the atlas.

Our system has been evaluated on a test set of 10 whole-body CT images by the organizers of the VISCERAL Anatomy Grand Challenge at ISBI 2015. Note that this test set is only available to the organizers. The final results are given in Table 1 together with the best competitors to date:

- *CMIV* - “Center for Medical Image Science and Visualization, Linköping University”,
- *HES-SO* - “University of Applied Sciences Western Switzerland” and
- *SIAT* - “Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences”.

<i>Organ</i>	<i>Our</i>	<i>CMIV</i>	<i>HES-SO</i>	<i>SIAT</i>
Left Kidney	<b>0.934</b>	0.896	0.784	-
Right Kidney	<b>0.915</b>	0.796	0.790	-
Spleen	0.870	<b>0.910</b>	0.703	0.874
Liver	0.921	<b>0.936</b>	0.866	0.923
Left Lung	<b>0.972</b>	0.961	<b>0.972</b>	0.952
Right Lung	<b>0.975</b>	0.970	<b>0.975</b>	0.957
Unirary Bladder	<b>0.763</b>	0.713	0.698	-
Muscle Body of Left Rectus Abdominis	<b>0.746</b>	-	0.551	-
Muscle Body of Right Rectus Abdominis	<b>0.679</b>	-	0.519	-
Lumbar Vertebra 1	<b>0.775</b>	-	0.718	-
Thyroid	0.424	-	<b>0.549</b>	-
Pancreas	0.383	-	<b>0.408</b>	-
Left Psoas Major Muscle	<b>0.861</b>	0.828	0.806	-
Right Psoas Major Muscle	<b>0.847</b>	0.817	0.787	-
Gallbladder	0.190	-	<b>0.276</b>	-
Sternum	<b>0.847</b>	-	0.761	-
Aorta	<b>0.830</b>	-	0.753	-
Trachea	<b>0.931</b>	-	0.92	-
Left Adrenal Gland	0.282	-	<b>0.373</b>	-
Right Adrenal Gland	0.220	-	<b>0.355</b>	-
Average	<b>0.718</b>	-	0.678	-

Table 1: Final results measured in DICE metric for whole-body CT images. Our approach gives the best results for 13 out of the 20 organs. Here '-' means that no segmentation was provided.

## 4 Conclusions

We have demonstrated that by using a feature-based approach to multi-atlas segmentation, it is possible to reliably locate and segment organs in whole-body CT images with state-of-the-art results. Still, there is room for improvement. For example, there is no guarantee that the system produces a valid organ shape. We are currently working on ways to directly incorporate such shape priors in the framework. Further, the speed of the system can be improved, for example, by circumventing the need to perform 20 pairwise registrations for every new target image [ANEK15].

## References

- [ANEK15] J. Alvé, A. Norlén, O. Enqvist, and F. Kahl. Überatlas: Robust speed-up of feature-based registration and multi-atlas segmentation. *Scandinavian Conference on Image Analysis*, 2015. To appear.
- [CR00] H. Chui and A. Rangarajan. A new algorithm for non-rigid point matching. *IEEE Conference on Computer Vision and Pattern Recognition*, 2:44–51, 2000.
- [CSK11] A. Criminisi, J. Shotton, and E. Konukoglu. Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. *Foundations and Trends® in Computer Graphics and Vision*, 7:81–227, 2011.
- [HKA<sup>+</sup>10] RA. Heckemann, S. Keihaninejad, P. Aljabar, D. Rueckert, JV. Hajnal, and A. Hammers. Improving intersubject image registration using tissue-class information benefits

- robustness and accuracy of multi-atlas based anatomical segmentation. *NeuroImage*, 51(1):221–227, 2010.
- [JSH12] O. Jamriška, D. Sýkora, and A. Hornung. Cache-efficient graph cuts on structured grids. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3673–3680, 2012.
- [KBG<sup>+</sup>11] F. Khalifa, GM. Beache, G. Gimel’farb, JS. Suri, and AS. El-Baz. State-of-the-art medical image registration methodologies: A survey. In *Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies*, pages 235 – 280. Springer, 2011.
- [KSK<sup>+</sup>10] HA. Kirisli, M. Schaap, S. Klein, LA. Neefjes, AC. Weustink, T. van Walsum, and WJ. Niessen. Fully automatic cardiac segmentation from 3d cta data: a multi-atlas based approach. *Proceedings of SPIE*, 2010.
- [ORS<sup>+</sup>01] S. Ourselin, A. Roche, G. Subsol, X. Pennec, and N. Ayache. Reconstructing a 3D structure from serial histological sections. *Image and vision computing*, 19(1):25–31, 2001.
- [SDP13] A. Sotiras, C. Davatzikos, and N. Paragios. Deformable medical image registration: A survey. *IEEE Transactions on Medical Imaging*, 32(7):1153–1190, 2013.
- [SEKO15] L. Svärm, O. Enqvist, F. Kahl, and M. Oskarsson. Improving robustness for inter-subject medical image registration using a feature-based approach. *International Symposium on Biomedical Imaging*, 2015.
- [WSD<sup>+</sup>13] H. Wang, JW. Suh, SR. Das, J. Pluta, C. Craige, and PA. Yushkevich. Multi-atlas segmentation with joint label fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(3):611–623, 2013.

# Fully Automatic Multi-organ Segmentation based on Multi-boost Learning and Statistical Shape Model Search

Baochun He, Cheng Huang, 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 multi-organ segmentation based on multi-boost learning and statistical shape model search was proposed. First, simple but robust Multi-Boost Classifier was trained to hierarchically locate and pre-segment multiple organs. To ensure the generalization ability of the classifier relative location information between organs, organ and whole body is exploited. Left lung and right lung are first localized and pre-segmented, then liver and spleen are detected upon its location in whole body and its relative location to lungs, kidney is finally detected upon the features of relative location to liver and left lung. Second, shape and appearance models are constructed for model fitting. The final refinement delineation is performed by best point searching guided by appearance profile classifier and is constrained with multi-boost classified probabilities, intensity and gradient features. The method was tested on 30 unseen CT and 30 unseen enhanced CT (CTce) datasets from ISBI 2015 VISCERAL challenge. The results demonstrated that the multi-boost learning can be used to locate multi-organ robustly and segment lung and kidney accurately. The liver and spleen segmentation based on statistical shape searching has shown good performance too.

---

*Copyright © by the paper's authors. Copying permitted only for private and academic purposes.*

In: O. Goksel (ed.): Proceedings of the VISCERAL Anatomy Grand Challenge at the 2015 IEEE International Symposium on Biomedical Imaging (ISBI), New York, NY, Apr 16<sup>th</sup>, 2015 published at <http://ceur-ws.org>

## 1 Introduction

Abdominal organ segmentation is an essential step in the multi-organ visualization, clinical diagnosis and therapy. Up to now, some methods [Okada12, Wang14] have been proposed, and all of them showed that information about the spatial relationship among organs is very beneficial to automatic 3D multi-organ localization. Previous studies also indicated that segmentation in a hierarchical way is more robust [Wang14, Selver14]. In our previous work [Li14], we used Adaboost and statistic shape model (SSM) prior knowledge to segment liver successfully. Now we extend this framework in multi-organ segmentation as shown in Figure 1. The differences are in two-fold. Firstly, Multi-Boost [Ben12] is employed to classify two organs one time in a top-down order. The last organ segmentation result will be used to classify the next level organs. Secondly, to acquire a customized specific shape result, free searching is directed by K Nearest Neighbor (KNN) and is constrained with voxel-based information such as probability, intensity and gradient features.

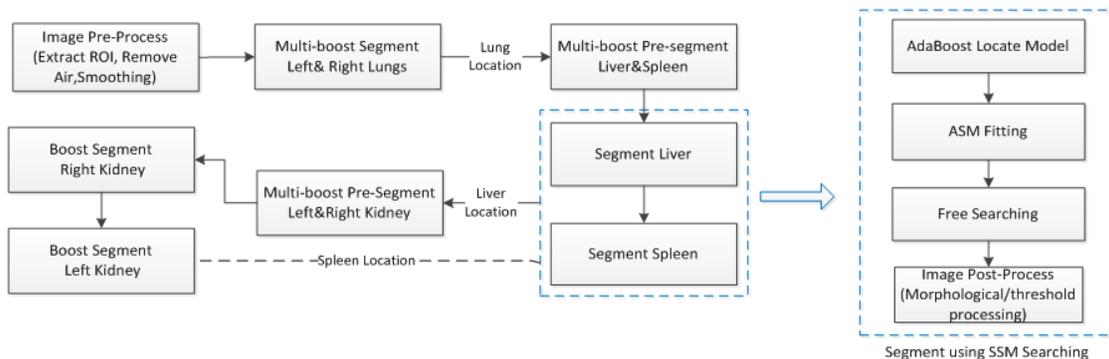


Figure 1: The framework of multi-organ segmentation

## 2 Method

### 2.1 Model Construction

SSM model was constructed from 20 CT and 20 CTce training binary segmentations. At first, reasonable region of interest (ROI) of the training binary images is extracted and generalized Procrustes aligned. Then one smooth and normal reference mesh is obtained using marching cubes method. Finally a set of corresponding shapes are created by elastic registration of the reference shape to the aligned binary images. The SSM is constructed by Statismo toolkit [Luthi12] and represented by Simplex mesh. The local appearance model of each organ is established by a KNN classifier trained on both intensity and gradient profiles information inside, outside and at the true organ boundary as suggested in [Heimann07].

### 2.2 Multi-organ Localization

Image features such as intensity, location and contextual information are used to train a multi-boost classifier. To ensure the generalization ability of the classifier relative location information between organs, organ and whole body were exploited. Template matching is employed to extract the organ ROI as shown in Figure 2(a). Localization and segmentation is performed in a top-down order - first left and right lung, then liver and spleen, at last left and right kidney, as seen in Figure 2(b). Thresholding was applied to the probability image of the boosting classified ROI image to get the pre-segmentation mask. Due to good boosting classification precision for lung and kidney, the pre-segmentation mask is used as the final segmentation.

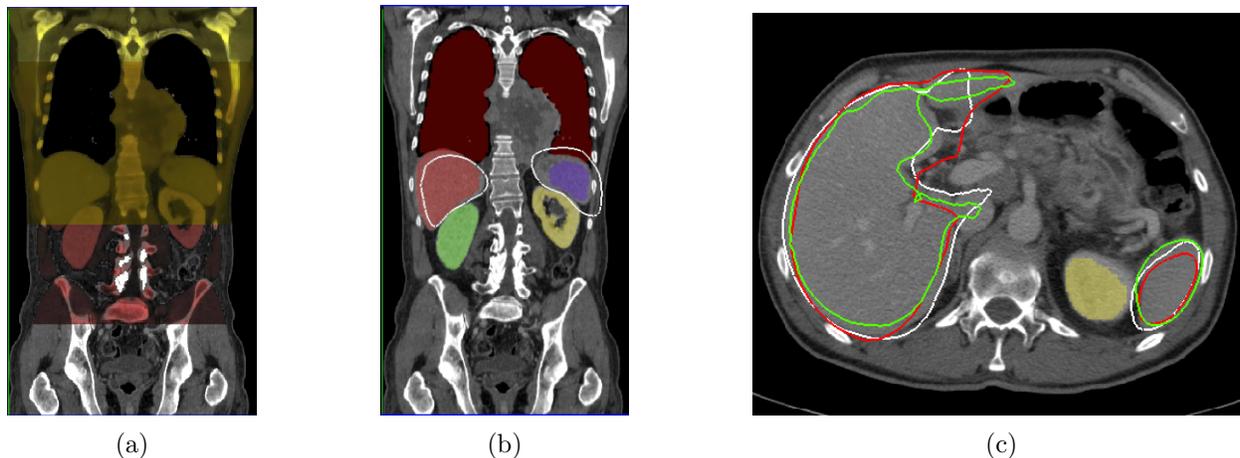


Figure 2: Three steps in the multi-organ segmentation framework: (a) Image preprocessing; (b) Model localization and segmentation of lung and kidney; (c) Shape fitting for liver and spleen, with pre-segmentation distance map (red), continued by boundary profile search (white), finally free-searching directed by the boundary profile classifier (green).

### 2.3 Active Shape Model Search

Similarity and shape transform parameters are initialized first by registration of SSM shape to the distance map of the pre-segmentation image. Appearance model is utilized for accurate parameters searching [Cootes95]. Previous trained KNN-classifier shifts each landmark to its optimal displacement position, similarity and shape parameters are then calculated through matrix operations. This process is performed iteratively until the parameters converge.

### 2.4 Appearance Profile Classifier directed Boundary Searching

In this step, the goal is to find the optimal confidence position for each mesh vertex. Due to high accuracy of the KNN, it is still used as boundary profile classification method. However, in step 2.3, the best positions calculated by KNN may overflow or fail to reach the true boundary as illustrated in Figure 2(c). The target position around the one searched by KNN is named as KNN position for convenience. The points around the KNN position are selected as candidate points. Each candidate point is assigned by previous Adaboost probability obtained in step 2.2, where both the intensity and the gradient are scaled to  $[-1,1]$ . The point with maximum voting value will be the optimal confidence position. To preserve the smoothness of the shape, the point can only move to the computed best position in a constrained step. This process stops after iteration of user-specified numbers.

## 3 Results

Twenty non-contrast CT and twenty contrast enhanced CT (CTce) training volumes were used for each multi-boost classifier and KNN boundary classifier training. SSM was built on all thirty datasets. There are 2562 landmarks for the mean liver shape model and 1520 ones for the mean spleen shape model. The experiment was run on 30 unseen CT and CTce datasets and evaluated by Dice coefficient and average Hausdorff distance (AvgD). The evaluation results are shown in Table 1.

Table 1: Multi-Organ Segmentation Results

Organ	Non-Contrast CT		Contrast-Enhanced CT	
	Dice Coefficient	AvgD (mm)	Dice Coefficient	AvgD (mm)
Left Lung	0.952	0.101	0.966	0.069
Right Lung	0.957	0.094	0.966	0.078
Liver	0.923	0.239	0.933	0.203
Spleen	0.874	0.360	0.896	0.385
Left Kidney	–	–	0.910	0.171
Right Kidney	–	–	0.922	0.131

## 4 Conclusions

In this paper, a robust and automatic multi-organ segmentation method was proposed. The method exploits and combines different prior knowledge, such as interrelations of organs, intensity, boundary profiles and shape variation information, for robust model localization, model fitting and free searching. The method has been validated on ISBI 2015 VISCERAL challenge and showed good performance. Future work will extend the framework to more abdominal organ segmentation.

## 5 Acknowledgments

This work was supported by the grants as follows: NSFC-Guangdong Union Foundation (Grant No. U1401254); Guangdong Science and Technology Project (Grant No. 2012A080203013 and 2012A030400013).

## References

- [Okada12] T. Okada, M. G. Linguraru, M. Hori. et al. Multi-organ segmentation in abdominal CT images. *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, DOI: 10.1109/EMBC.2012.6346840, pp. 3986 - 3989, 2012.
- [Selver14] M. A. Selver Segmentation of abdominal organs from CT using a multi-level, hierarchical neural network strategy. *Computer Methods and Programs in Biomedicine*, 113(3):830-852, 2014.
- [Wang14] C. Wang, O. Smedby. Automatic multi-organ segmentation in nonenhanced CT datasets using hierarchical shape priors. *The 22nd International Conference on Pattern Recognition (ICPR)*, pp. 3327-3332, 2014.
- [Li14] X. Li, C. Huang, F. Jia, Z. Li, C. Fang, Y. Fan. Automatic liver segmentation using statistical prior models and free-form deformation. *In: Menze, B., et al. (eds.) MCV 2014. LNCS*, vol. 8848, pp. 181-188. Springer, Heidelberg (2014).
- [Ben12] D. Benbouzid, R. Busa-Fekete, N. Casagrande, F. -D. Collin, B. Kgl. MULTIBOOST: a multi-purpose boosting package. *The Journal of Machine Learning Research*, 13, pp.549-553, 3/1/2012
- [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.

# Hierarchic Anatomical Structure Segmentation Guided by Spatial Correlations (AnatSeg-Gspac): VISCERAL Anatomy3

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

Yashin Dicente Cid  
yashin.dicente@hevs.ch

Adrien Depeursinge  
adrien.depeursinge@hevs.ch

Henning Müller  
henningmueller@hevs.ch

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

## Abstract

Medical image analysis techniques require an initial localization and segmentation of the corresponding anatomical structures. As part of the VISCERAL Anatomy segmentation benchmarks, a hierarchical multi-atlas multi-structure segmentation approach guided by anatomical correlations is proposed (AnatSeg-Gspac). The method defines a global alignment of the images and refines locally the anatomical regions of interest for the smaller structures. In this paper, the method is evaluated in the VISCERAL Anatomy3 benchmark in twenty anatomical structures in both contrast-enhanced and non-enhanced computed tomography (CT) scans. AnatSeg-Gspac obtained the lowest average Hausdorff distance in 19 out of the 40 possible structure scores in the test set CT scans.

## 1 Introduction

Medical image analysis and computer-aided diagnosis initially require an accurate location and segmentation of the anatomical structures present. The time expensive task of manually annotating the current large amounts of medical image data daily produced restricts the implementation of further analysis by computer algorithms [Doi05]. Different approaches have been proposed to automatically detect multiple or single anatomical structures within the patient images [LSL<sup>+</sup>10, CRK<sup>+</sup>13]. The VISual Concept Extraction challenge in RAdioLogY project (VISCERAL<sup>1</sup>) organizes public

---

*Copyright © by the paper's authors. Copying permitted only for private and academic purposes.*

In: O. Goksel (ed.): Proceedings of the VISCERAL Anatomy Grand Challenge at the 2015 IEEE International Symposium on Biomedical Imaging (ISBI), New York, NY, Apr 16<sup>th</sup>, 2015 published at <http://ceur-ws.org>

<sup>1</sup><http://www.visceral.eu/>, as of 1 April 2015

benchmarks to test multiple segmentation approaches on the same available medical dataset for an objective evaluation of the algorithms [JdTGM<sup>+</sup>14]. The VISCERAL data set has been manually annotated by radiologists and includes real medical images obtained from clinical routine in hospitals. The benchmarks are set up in a cloud environment platform designed to host large amounts of medical data with equal computing instances for the participating research groups [LMMH13].

A hierarchic Anatomical structure Segmentation Guided by spatial correlations (AnatSeg-Gspac)[JdTM13, JdTGM<sup>+</sup>14, JdTM14b] has been previously proposed and tested in the first two VISCERAL Anatomy benchmarks. This approach requires no interaction from the user and generates a robust segmentation for multiple anatomical structures with short re-training phase for new scan parameters or additional structures [JdTM14a]. The evaluation and results of AnatSeg-Gspac in the VISCERAL Anatomy3 benchmark are presented in the following sections.

## 2 Materials and Methods

### 2.1 Dataset

For the VISCERAL Anatomy3 benchmark 20 CT contrast-enhanced of the trunk (CTce) and 20 CT whole body unenhanced (CTwb) with their manual annotations (up to 20 anatomical structures), were provided to the participants for training. For the implementation of AnatSeg-Gspac in this benchmark a subset of volumes (7) with all or the majority of manual annotations were selected per modality as atlases. Further information on the VISCERAL data set can be found in [JdTGM<sup>+</sup>14].

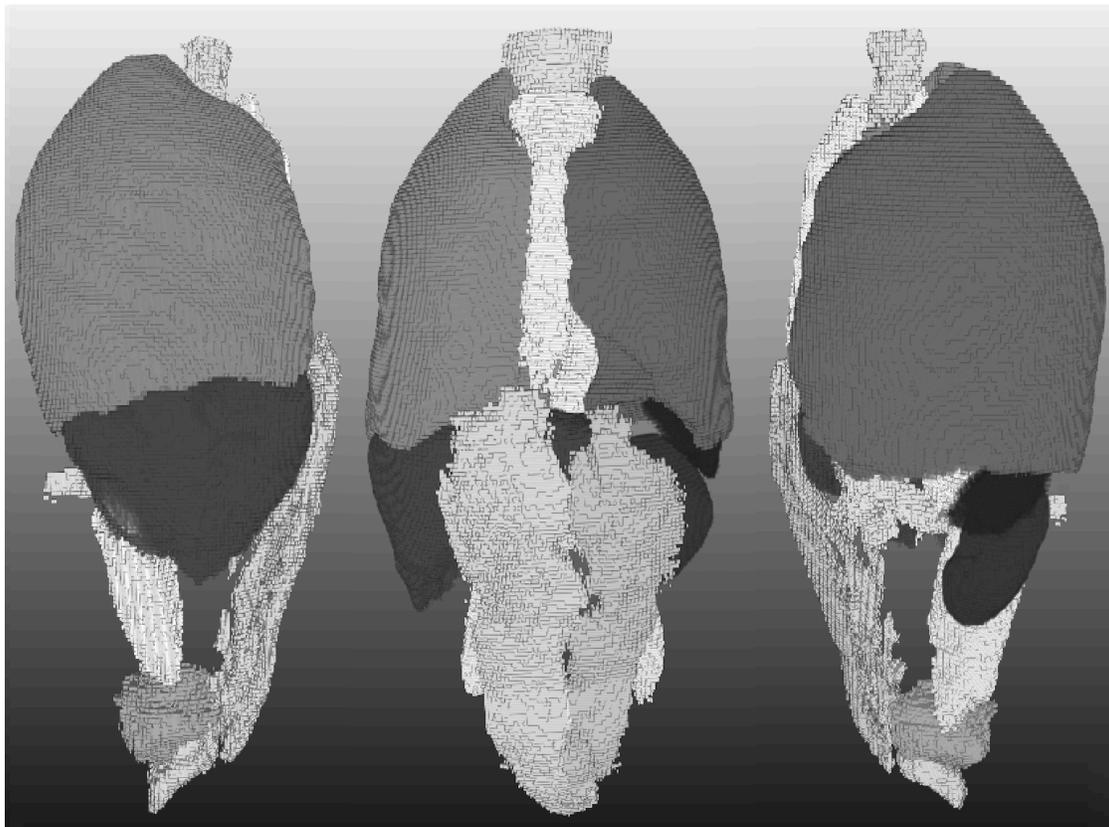


Figure 1: 3D rendering sample output using AnatSeg-Gspac.

## 2.2 AnatSeg-Gspac

The proposed method performs a hierarchic multi-atlas multi-structure segmentation defining anatomical regions of interest in their spatial domain. The bigger and high contrast anatomical structures are used as reference for smaller structures with low contrast, which are consequently harder to segment. The registration pipeline has been optimized to reduce the amount of registrations needed for the smaller structures obtaining also a robust localization. In Figure 1, a sample segmentation output for one unused training volume including all the anatomical structures evaluated in the VISCERAL benchmarks is shown. Further information on the AnatSeg-Gspac method can be found in the previously referenced papers.

## 3 Evaluation

For the Anatomy3 benchmark the test set included 10 CTce volumes and 10 CTwb scans. Twenty different evaluation metrics are provided to the participants about their algorithm performance for each anatomical structure. The evaluation phase is performed in the Azure cloud by the organizers with no intervention from the participants.

Table 1: Average Hausdorff distance results in trunk CTce test set of the VISCERAL Anatomy3 benchmark (Anatomy3 Leaderboard, <http://www.visceral.eu/Leaderboard/>, as of 1 April 2015). Competitive scores were obtained for kidneys and lungs compared to other organ-specific methods.

Average Hausdorff Distance																					
CTce	left kidney	right kidney	spleen	liver	left lung	right lung	urinary bladder	muscle body of left rectus abdominis	muscle body of right rectus abdominis	lumbar Vertebra 1	thyroid	pancreas	left psoas major muscle	right psoas major muscle	gallbladder	sternum	aorta	trachea	left adrenal gland	right adrenal gland	
	29663	29662	86	58	1326	1302	237	40358	40357	29193	7578	170	32249	32248	187	2473	480	1247	30325	30324	
Dicente et al.					0.050	0.052															
Jia et al.	0.171	0.131	0.385	0.203	0.069	0.078															
JiménezdelToro et al.	0.172	0.243	2.005	0.514	0.071	0.065	1.879	<b>6.068</b>	<b>6.600</b>	<b>4.504</b>	<b>3.337</b>	<b>3.804</b>	<b>0.742</b>	0.757	<b>3.603</b>	<b>0.899</b>	<b>1.094</b>	<b>0.223</b>	<b>3.115</b>	<b>2.660</b>	
Wang et al.	0.137	0.072	0.573	0.174	0.076	0.070	0.485						0.934	0.734							

Table 2: Average Hausdorff distance results in CT whole body (CTwb) test set of the VISCERAL Anatomy3 benchmark (Anatomy3 Leaderboard, <http://www.visceral.eu/Leaderboard/>, as of 1 April 2015). The method AnatSeg-Gspac generates robust segmentations for the big structures like the lungs (best benchmark scores highlighted). Moreover, it also shows overall better results particularly for small structures like thyroid, gallbladder and both adrenal glands.

Average Hausdorff Distance																				
CTwb	left kidney	right kidney	spleen	liver	left lung	right lung	urinary bladder	muscle body of left rectus abdominis	muscle body of right rectus abdominis	lumbar Vertebra 1	thyroid	pancreas	left psoas major muscle	right psoas major muscle	gallbladder	sternum	aorta	trachea	left adrenal gland	right adrenal gland
	29663	29662	86	58	1326	1302	237	40358	40357	29193	7578	170	32249	32248	187	2473	480	1247	30325	30324
Dicente et al.					0.050	0.046														
Jia et al.			0.360	0.239	0.101	0.094														
JiménezdelToro et al.	1.209	1.307	1.974	0.780	<b>0.043</b>	<b>0.038</b>	1.457	3.550	4.032	<b>1.953</b>	<b>1.466</b>	5.521	0.595	0.775	<b>5.938</b>	0.938	1.193	0.103	<b>2.672</b>	<b>3.445</b>
Kahl et al.	0.147	0.229	0.534	0.299	0.045	0.043	1.057	1.614	1.922	2.424	2.163	4.478	0.443	0.550	9.617	0.542	0.798	0.083	3.298	7.046
Wang et al.	0.758	2.751	0.248	0.191	0.356	0.096	2.155						0.705	0.812						

The proposed method obtained the lowest average Hausdorff distance of the Anatomy3 benchmark in 12/20 structures in CTce (Table 1) and 7/20 structures in CTwb (Table 2). The DICE coefficient scores are also presented for all the methods submitted in the benchmark (Table 3 and Table 4).

Table 3: DICE coefficient results in the test set trunk CT contrast-enhanced (CTce) of the VISCERAL Anatomy3 benchmark (Anatomy3 Leaderboard, <http://www.visceral.eu/Leaderboard/>, as of 1 April 2015). The proposed AnatSeg-Gspac(Jiménez del Toro et al. in light grey) was the only submitted method that segmented all available anatomical structures in both CT modalities (enhanced and unenhanced).

DICE Coefficient																				
CTce	left kidney	right kidney	spleen	liver	left lung	right lung	urinary bladder	muscle body of left rectus abdominis	muscle body of right rectus abdominis	lumbar Vertebra 1	thyroid	pancreas	left psoas major muscle	right psoas major muscle	gallbladder	sternum	aorta	trachea	left adrenal gland	right adrenal gland
	29663	29662	86	58	1326	1302	237	40358	40357	29193	7578	170	32249	32248	187	2473	480	1247	30325	30324
Dicente et al.	0.974 0.973																			
Jia et al.	0.910	0.922	0.869	0.933	0.966	0.966														
JiménezdelToro et al.	0.910	0.889	0.730	0.887	0.959	0.963	0.679	<b>0.474</b>	<b>0.453</b>	<b>0.523</b>	<b>0.410</b>	<b>0.423</b>	0.794	0.799	<b>0.484</b>	<b>0.721</b>	<b>0.762</b>	<b>0.855</b>	<b>0.331</b>	<b>0.342</b>
Wang et al.	0.945	0.959	0.909	0.949	0.972	0.971	0.856	0.830 0.845												

Table 4: DICE coefficient results in the test set unenhanced CT of the whole body (CTwb) of the VISCERAL Anatomy3 benchmark (Anatomy3 Leaderboard, <http://www.visceral.eu/Leaderboard/>, as of 1 April 2015). Highlighted are the best DICE overlap scores obtained in the benchmark by AnatSeg-Gspac in 7 clinically relevant anatomical structures: left and right lungs, thyroid, pancreas, gallbladder, left and right adrenal gland.

DICE Coefficient																				
CTwb	left kidney	right kidney	spleen	liver	left lung	right lung	urinary bladder	muscle body of left rectus abdominis	muscle body of right rectus abdominis	lumbar Vertebra 1	thyroid	pancreas	left psoas major muscle	right psoas major muscle	gallbladder	sternum	aorta	trachea	left adrenal gland	right adrenal gland
	29663	29662	86	58	1326	1302	237	40358	40357	29193	7578	170	32249	32248	187	2473	480	1247	30325	30324
Dicente et al.	<b>0.972</b> 0.974																			
Jia et al.	0.874 0.923 0.952 0.957																			
JiménezdelToro et al.	0.784	0.790	0.703	0.866	<b>0.972</b>	<b>0.975</b>	0.698	0.551	0.519	0.718	<b>0.549</b>	<b>0.408</b>	0.806	0.787	<b>0.276</b>	0.761	0.753	0.920	<b>0.373</b>	<b>0.355</b>
Kahl et al.	0.934	0.915	0.870	0.921	<b>0.972</b>	<b>0.975</b>	0.763	0.746	0.679	0.775	0.424	0.383	0.861	0.847	0.190	0.847	0.830	0.931	0.282	0.220
Wang et al.	0.896	0.796	0.910	0.936	0.961	0.970	0.713	0.828 0.817												

## 4 Discussion and Conclusions

The proposed method showed robustness in the segmentation of multiple structures from two different imaging modalities using a small training set. Both the distance and overlap scores in this and the previous Anatomy benchmarks show AnatSeg-Gspac outperforms other algorithms in some of the smaller anatomical structures (e.g. both adrenal glands, gallbladder). It can also obtain the best overlap for bigger and high contrasted structures like the lungs.

A limitation of the method is the computation cost, mainly for the B-spline non-rigid registrations. Although the number of registrations and the size of registered regions are reduced using anatomical correlations, the execution time is around 13 hours for a complete CT volume. A faster

code implementation and better selection of the relevant atlases may reduce the number of needed registrations and thus the execution time of the method.

The method can be extended to the other imaging modalities and include more anatomical structures with short re-training phases. This is particularly important for its application with new or different scanners contained in large not annotated data sets. Further clinical image analyses, that may require the location of additional structures, might also benefit from this feature or include the output locations of the method as an initialization step.

## 5 Acknowledgments

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

## References

- [CRK<sup>+</sup>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.
- [JdTGM<sup>+</sup>14] Oscar Alfonso Jiménez del Toro, Orcun Goksel, Bjoern Menze, Henning Müller, Georg Langs, Marc-André Weber, Ivan Eggel, Katharina Gruenberg, Markus Holzer, Georgios Kotsios-Kontokotsios, Markus Krenn, Roger Schaer, Abdel Aziz Taha, Marianne Winterstein, and Allan Hanbury. VISCERAL – VISual Concept Extraction challenge in RADioLogY: ISBI 2014 challenge organization. In Orcun Goksel, editor, *Proceedings of the VISCERAL Challenge at ISBI*, number 1194 in CEUR Workshop Proceedings, pages 6–15, Beijing, China, May 2014.
- [JdT<sup>M</sup>13] 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.
- [JdT<sup>M</sup>14a] Oscar Jiménez del Toro and Henning Müller. Hierarchic multi-atlas based segmentation for anatomical structures: Evaluation in the visceral anatomy benchmarks. In *MICCAI workshop on Medical Computer Vision*, Lecture Notes in Computer Science. Springer, 2014.
- [JdT<sup>M</sup>14b] 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.
- [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<sup>+</sup>10] Marius George Linguraru, Jesse K. Sandberg, Zhixi Li, Fuhawn 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.

# Multi-modal Multi-Atlas Segmentation using Discrete Optimisation and Self-Similarities

Mattias P. Heinrich      Oskar Maier      Heinz Handels  
heinrich@imi.uni-luebeck.de

Institute of Medical Informatics, University of Lübeck, Germany

## Abstract

This work presents the application of a discrete medical image registration framework to multi-organ segmentation in different modalities. The algorithm works completely automatically and does not have to be tuned specifically for different datasets. A robust similarity measure, using the local self-similarity context (SSC), is employed and shown to outperform other commonly used metrics. Both affine and deformable registration are driven by a dense displacement sampling (deeds) strategy. The smoothness of displacements is enforced by inference on a Markov random field (MRF), using a tree approximation for computational efficiency. Consensus segmentations for unseen test images of the VISCERAL Anatomy 3 data are found by majority voting.

## 1 Introduction

Organ segmentations are an important processing step in medical image analysis, e.g. for image-guided interventions, radiotherapy, or improved radiological diagnostics. General solutions are preferable over organ specific models for large scale image processing. Machine learning approaches, in particular the popular random decision forests (RDF), have been recently used for multi-organ localisation [CSB09] and segmentation [GPKC12], yet for more challenging modalities (e.g. structural MRI) they have had limited success. This is partly due to the inhomogeneous intensity variations within and across MR scans. Registration-based multi-atlas segmentation can provide more robustness by using contrast-invariant similarity measures to guide the alignment of atlas to patient data. Here, we propose to employ a discrete registration model, which can capture large deformations to accurately segment volumes with large differences in patient anatomy and geometry. Combined with a robust multi-modal similarity metric (self-similarity context) it can be applied to registering both CT and MRI scans reliably. The method is briefly reviewed in the next section

---

*Copyright © by the paper's authors. Copying permitted only for private and academic purposes.*

In: O. Goksel (ed.): Proceedings of the VISCERAL Anatomy Grand Challenge at the 2015 IEEE International Symposium on Biomedical Imaging (ISBI), New York, NY, Apr 16<sup>th</sup>, 2015 published at <http://ceur-ws.org>

and the experimental setting detailed thereafter. The results on both training and test datasets are discussed in Sec. 4 and compared to some state-of-the-art approaches.

## 2 Method

Discrete optimisation can capture large motions by defining an appropriate range of displacements  $\mathbf{u}$ . It enables a flexible choice of different similarity terms, since no derivative is required. We use the framework presented in [HJBS13], which defines a graphical model with nodes  $p \in \mathcal{V}$  (with spatial location  $\mathbf{x}_p$ ) that correspond to control points in a uniform B-spline grid. For each node, the hidden labels  $f_p$  (from a large quantised set  $\mathcal{L}$ ) are defined as potential 3D displacements  $f_p = \mathbf{u}_p = \{u_p, v_p, w_p\}$  between a control point  $p$  in the fixed image  $F$  and moving image  $M$ . Edges between nodes used for inference of the pair-wise regularisation costs  $R(f_p, f_q)$  ( $p, q \in \mathcal{E}$ ) are modelled by a minimum spanning tree (MST) for computational efficiency. The displacement field is regularised using the squared differences of the displacements of neighbouring control points:

$$R(f_p, f_q) = \sum_{(p,q) \in \mathcal{E}} \frac{\|\mathbf{u}_p - \mathbf{u}_q\|^2}{\|\mathbf{x}_p - \mathbf{x}_q\|} \quad (1)$$

For the image similarity (data term) self-similarity descriptors are used [HJP<sup>+</sup>13]. The self-similarity context is based on local patch distances within each image and invariant to contrast change, robust to noise and modality independent. The dissimilarity metric  $D$ , the  $L_1$  norm between 64 bit binary descriptor representations  $SSC_F$  (for fixed image) and  $SSC_M$  (for moving image) at two locations  $\mathbf{x}$  and  $\mathbf{x} + \mathbf{u}$ , can be efficiently calculated in the Hamming space:

$$D(\mathbf{x}_p, \mathbf{u}_p) = 1/|\mathcal{P}| \sum_{y \in \mathcal{P}} \Xi\{SSC_F(\mathbf{x}_p + y) \oplus SSC_M(\mathbf{x}_p + \mathbf{u}_p + y)\} \quad (2)$$

where  $\oplus$  defines an exclusive OR,  $\Xi$  a population count and  $y \in \mathcal{P}$  the local patch coordinates. The combined energy function with regularisation parameter  $\alpha$  becomes:  $E(f) = \sum_{p \in \mathcal{V}} D(f_p) + \alpha \sum_{(p,q) \in \mathcal{E}} R(f_p, f_q)$ . Belief propagation [FH06] on the MST (our relaxed graphical) is employed to find the global minimum without iterations in only two passes.

Prior to the deformable registration, a block-matching based linear registration using also the SSC metric is employed as detailed in [HPSH14].<sup>1</sup>

## 3 Experiments

The deformations between different anatomies make a large number of degrees of freedoms necessary. As pre-processing the images are resampled to an isotropic resolution of  $1.8 \times 1.8 \times 1.8$  mm<sup>3</sup> and padded or cropped to have same dimensions. For the affine pre-registration, three scales of control-point grids with spacings of [9, 8, 7] voxels are used. The displacement label space is defined by two parameters: number of steps  $l_{\max}$  and quantisation step  $q$ , which together define the label space  $\mathcal{L} = q \cdot \{0, \pm 1, \dots, \pm l_{\max}\}^3$  voxels. We used  $l_{\max} = [6, 5, 4]$  and  $q = [5, 4, 3]$  voxels. For the deformable registration four scale levels with spacings of [8, 7, 6, 5], numbers of steps of  $l_{\max} = [6, 5, 4, 3]$  and quantisations of  $q = [4, 3, 2, 1]$  voxels were used. The number of random samples and the regularisation weight were left at their default parameters 50 and 2. Inverse consistent is improved by employing a symmetric calculation of deformations (see [HJP<sup>+</sup>13]).

To asses the impact of the similarity metric, we additionally performed experiments using mutual information (MI) and normalised gradient fields (NGF) [HM05]. For a more detailed comparison of the optimisation, we also applied the popular continuous-optimisation based framework NiftyReg [MRT<sup>+</sup>10] (which uses a B-spline parameterisation) with an affine initialisation [ORPA00].

<sup>1</sup>Our software is publicly available for download at [www.mpheinrich.de](http://www.mpheinrich.de) (deedsRegSSC)

Table 1: Experimental results for training dataset of VISCERAL Anatomy 3 challenge. Dice volume overlap for the 7 most common organs (psoas major muscles are abbreviated by pmm) in abdominal and thorax scans when using majority voting. The results of [GSG14] are from a different subset of the challenge (Anatomy 2), so there are not directly comparable.

<b>method</b>	liver	spleen	bladder	r kidney	l kidney	r pmm	l pmm	<b>avg</b>
deeds+SSC CT-CT	0.92	0.84	0.82	0.91	0.91	0.85	0.84	<b>0.872</b>
deeds+MI MR-CT	0.77	0.66	0.16	0.52	0.85	0.69	0.64	<b>0.610</b>
deeds+NGF MR-CT	0.77	0.74	0.31	0.55	0.86	0.75	0.73	<b>0.673</b>
<b>deeds+SSC MR-CT</b>	0.82	0.78	0.44	0.62	0.88	0.80	0.79	<b>0.732</b>
NiftyReg+MI MR-MR	0.81	0.79	0.05	0.58	0.77	0.52	0.36	<b>0.554</b>
[GSG14] MR-MR	0.83	0.66	0.21	0.88	0.85	0.64		<b>0.677</b>
<b>deeds+SSC MR-MR</b>	0.80	0.82	0.63	0.55	0.88	0.79	0.76	<b>0.744</b>
proposed <b>Test</b> MR-MR	0.79	0.71	0.36	0.78	0.83	0.76	0.78	<b>0.714</b>

## 4 Results

Our results are summarised in Table 1 for a subset of 10 training scans of the contrast enhanced (ce) abdominal MRI modality (or thorax/abdominal ceCT) and a leave-one-out validation. It can be seen that MRI segmentation is substantially more challenging yielding average results of Dice overlap for 7 organs of at most 0.744, while the results for the same setting for CT scans are  $\approx 0.13$  higher. Either of the two compared discrete optimisation strategies, by Gass et al. [GSG14] and our framework [HJBS13], outperforms the continuous optimisation approach of [MRT<sup>+</sup>10]. Using SSC as similarity metric improves the segmentation by 0.12 compared to MI and by 0.06 compared to NGF within the same framework. The multi-modal segmentation, for which we used MRI scans as fixed and CT scans as moving atlas scans, shows nearly identical accuracy to using same modality priors. This is an interesting finding, which could be employed for generating synthetic CT scans from MRI scans, e.g. for MR-PET reconstruction [HSS<sup>+</sup>08]. Due to time limitations only preliminary results for the hidden test datasets could be computed (last row of Table 1), for which we employed only three atlas scans each. We anticipate further improvements for our final results, which will subsequently be published on the VISCERAL leaderboard. The run-time of our algorithm on the virtual machine was on average 4 minutes per registration, which can be reduced with an optimised CPU implementation to less than a minute.

## 5 Conclusion

We have demonstrated that deformable registration using discrete optimisation enables accurate automatic MRI organ segmentation. Choosing both a robust similarity metric and optimisation strategy has been found to be important for achieving high overlap. Local similarity-weighted atlas performance estimation and advanced label fusion [AL13] may further improve the results. While machine learning techniques alone may not achieve the same accuracy as registration-based approaches for MRI segmentation, the combination of both can boost the performance. In initial experiments, we found that an RDF trained with both atlas-based priors and intensity features [MWG<sup>+</sup>15] improves the segmentation overlap of liver, spleen and kidneys by  $\approx 0.06$ .

## References

- [AL13] Andrew J Asman and Bennett A Landman. Non-local statistical label fusion for multi-atlas segmentation. *Medical Image Analysis*, 17(2):194–208, 2013.

- [CSB09] Antonio Criminisi, Jamie Shotton, and Stefano Bucciarelli. Decision forests with long-range spatial context for organ localization in CT volumes. *MICCAI workshop on Probabilistic Models for Medical Image Analysis*, pages 69–80, 2009.
- [FH06] Pedro Felzenszwalb and Daniel Huttenlocher. Efficient belief propagation for early vision. *International Journal of Computer Vision*, 70:41–54, 2006.
- [GPKC12] Ben Glocker, Olivier Pauly, Ender Konukoglu, and Antonio Criminisi. Joint classification-regression forests for spatially structured multi-object segmentation. In *ECCV 2012*, pages 870–881. Springer, 2012.
- [GSG14] Tobias Gass, Gabor Szekely, and Orcun Goksel. Multi-atlas segmentation and landmark localization in images with large field of view. In *Medical Computer Vision: Algorithms for Big Data*, pages 171–180. Springer, 2014.
- [HJBS13] Mattias P. Heinrich, Mark Jenkinson, Sir Michael Brady, and Julia A. Schnabel. MRF-based deformable registration and ventilation estimation of lung CT. *IEEE Transactions on Medical Imaging*, 32(7):1239–1248, 2013.
- [HJP<sup>+</sup>13] Mattias P. Heinrich, Mark Jenkinson, Bartłomiej W. Papież, Sir Michael Brady, and Julia A. Schnabel. Towards realtime multimodal fusion for image-guided interventions using self-similarities. In *MICCAI, LNCS*, pages 187–194. Springer, 2013.
- [HM05] Eldad Haber and Jan Modersitzki. Beyond mutual information: A simple and robust alternative. In *Bildverarbeitung für die Medizin 2005*, pages 350–354. Springer, 2005.
- [HPSH14] Mattias P. Heinrich, Bartłomiej W. Papież, Julia A. Schnabel, and Heinz Handels. Multispectral image registration based on local canonical correlation analysis. In *MICCAI, LNCS*, pages 202–209. Springer, 2014.
- [HSS<sup>+</sup>08] Matthias Hofmann, Florian Steinke, Verena Scheel, Guillaume Charpiat, Jason Farquhar, Philip Aschoff, Sir Michael Brady, Bernhard Schölkopf, and Bernd J. Pichler. MRI-based attenuation correction for PET/MRI: a novel approach combining pattern recognition and atlas registration. *Journal of Nuclear Medicine*, 49(11):1875–1883, 2008.
- [MRT<sup>+</sup>10] Marc Modat, Gerard R Ridgway, Zeike A Taylor, Manja Lehmann, Josephine Barnes, David J Hawkes, Nick C Fox, and Sébastien Ourselin. Fast free-form deformation using graphics processing units. *Computer Methods and Programs in Biomedicine*, 98(3):278–284, 2010.
- [MWG<sup>+</sup>15] Oskar Maier, Matthias Wilms, Janina von der Gablentz, Ulrike M Krämer, Thomas F Münte, and Heinz Handels. Extra tree forests for sub-acute ischemic stroke lesion segmentation in MR sequences. *Journal of Neuroscience Methods*, 240:89–100, 2015.
- [ORPA00] Sébastien Ourselin, Alexis Roche, Sylvain Prima, and Nicholas Ayache. Block matching: A general framework to improve robustness of rigid registration of medical images. In *MICCAI 2000*, pages 557–566. Springer, 2000.

# Efficient and fully automatic segmentation of the lungs in CT volumes

Yashin Dicente Cid  
yashin.dicente@hevs.ch

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

Adrien Depeursinge  
adrien.depeursinge@hevs.ch

Henning Müller  
henning.mueller@hevs.ch

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

## Abstract

The segmentation of lung volumes constitutes the first step for most computer-aided systems for lung diseases. CT (Computed Tomography) is the most common imaging technique used by these systems, so fast and accurate methods are needed to allow early and reliable analysis. In this paper, an efficient and fully automatic method for the segmentation of the lung volumes in CT is presented. This method deals with the initial segmentation of the respiratory system, the posterior extraction of the air tracks, and the final identification of the two lungs with 3 novel approaches. The system relies only on anatomical assumptions and was evaluated in the context of the VISCERAL Anatomy3 Challenge, achieving one of the best results.

## 1 Introduction

The first step of most computer-aided decision support systems for lung diseases is to segment the lungs. Moreover, an accurate segmentation of the two lungs can help the localization of other organs such as the liver or the heart that are closely related. X-ray computed tomography (CT) is considered to be the gold standard for pulmonary imaging. In the literature standard approaches for segmenting the respiratory system by thresholding the gray level images can be found in [IM03, HHR01, EBF02, LNC07]. The approaches are based on knowledge of the air gray-level in CT scans as CTs are based on tissue density. However, the gray range in the lungs regions can be affected by the radiation applied to acquire the CT and the possible change of the organ due to diseases (such as Fibrosis).

---

*Copyright © by the paper's authors. Copying permitted only for private and academic purposes.*

In: O. Goksel (ed.): Proceedings of the VISCERAL Anatomy Grand Challenge  
at the 2015 IEEE International Symposium on Biomedical Imaging (ISBI), New York, NY, Apr 16<sup>th</sup>, 2015  
published at <http://ceur-ws.org>

This work presents a novel and fully automatic approach for segmenting the lungs. We first apply a K–Means [Mac67] clustering of the CT intensities with a fixed number of clusters equal to 2 for segmenting the respiratory system. In the second step, the air tracks are removed from the initial segmentation. A novel technique is presented based on the mass–distribution of the lung volumes. The final step consist of identifying the right and left lung and refining the final mask by mathematical morphological operations in 3D. The separation of right and left lungs is challenging when both lungs seem to be connected. In this case, a bidirectional process across the 2D axial slices is applied. It allows to reduce the splitting error due to the information propagated between slices. Once both lungs are identified, a refinement in 3D is applied to each lung mask. The entire approach is completely unsupervised and provides an accurate and fast fully automatic segmentation of the lungs.

## 2 Database used

VISCERAL<sup>1</sup> Anatomy3 is the benchmark used in the VISCERAL Challenge at ISBI 2015. This benchmark contains a set of medical image series with annotated structures from various modalities. We evaluated our method for segmentation of right and left lung in the modalities of CT and with and without contrast agent (CTce). A total of 20 training patients in each modality were provided to optimize parameters.

The methods proposed by the participants were executed by the organizers of the challenge in the cloud and tested on a dataset of 10 patients per modality. The test set is not accessible by the participants to avoid possible overfitting of the methods. Despite the challenge offering a training set, the method proposed in this work was set up based on anatomical assumptions and no training was required. Patients from other datasets were used to define these assumptions, leaving the training set of this challenge for verification purposes.

## 3 Methods

The method presented is composed of three parts: an initial clustering of the CT values for segmenting the complete respiratory system (lungs, trachea and primary bronchi); a process to remove the trachea and primary bronchi; and finally, the identification of right and left lung with a refinement of each lung mask (see Fig. 1). Some steps of the process are performed in 2D following the axial dimension of the CT volume, i.e. going through the axial slices.



Figure 1: Pipeline of the proposed method for segmenting the lung volumes in CT.

### 3.1 Respiratory system segmentation

The proposed method for segmenting the respiratory system is based on the assumption that the latter is the biggest 3D connected air region inside the body. The first step is to fill the holes in the axial slices by a filling operation [Soi03], where a hole is defined as an area of dark pixels surrounded by lighter pixels. The resulting image contains a dense–body (see Fig. 2b). Then the absolute difference between the original and the dense–body image is computed. The resulting

<sup>1</sup><http://www.visceral.eu/>, as of 30 March 2015

image contains values that are clearly larger than 0 in the air regions inside the body, and close to 0 in the other regions (see Fig. 2c). In this new image, a K-Means [Mac67] algorithm with  $K = 2$  is performed, which yields a binary mask (see Fig.2d). Artificial objects in the CT containing air, such as the plastic bed, may be selected in the clustering, but are removed by analyzing the aspect ratio of the corresponding bounding boxes. Finally, the biggest connected 3D region is used as the initial lung mask. This region showed to include either both lungs connected by the trachea, or only one lung in the case of not being connected by the air track. To deal with this case, the process also selects all air blocks in the same axial slice-range, i.e., in the same slices where the largest 3D region is present and removes the regions that can not be easily connected to the lungs.

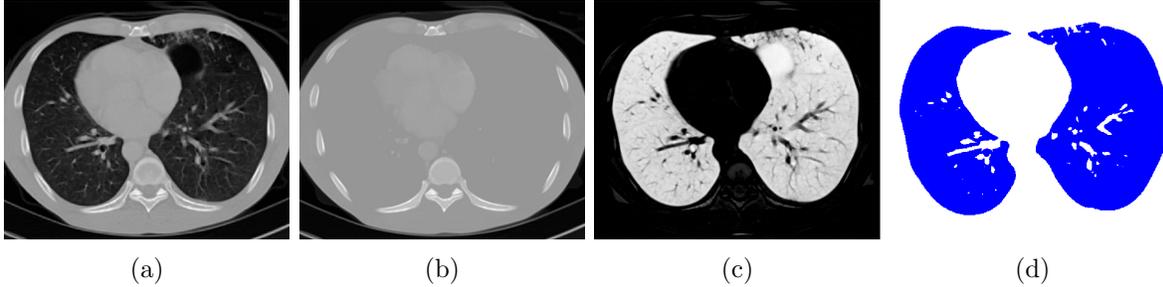


Figure 2: CT pre-processing and posterior clustering for segmenting the respiratory system inside the body. (a): Original CT. (b): Dense-body after filling holes. (c): Absolute difference between (a) and (b). (d): Mask achieved by 2-Means clustering over (c).

### 3.2 Removing trachea and primary bronchi

In order to remove the trachea and primary bronchi, the process defines a plane that divides the 3D image into two parts, leaving an equivalent number of mask-voxels on each side. This process uses the center of mass of the mask obtained in Section 3.1. The plane is used as the reference axis in each slice and the Euclidean distance from every pixel to this axis is computed (see Fig. 3a). Finally, each conected 2D component (CC) is assigned to the maximum distance found among all its constituting voxels (see Fig. 3b). The regions with a maximum distance to the central axis below a threshold are considered part of the air track and removed. This threshold is dynamically defined for each slice and patient according to the size of the mask.

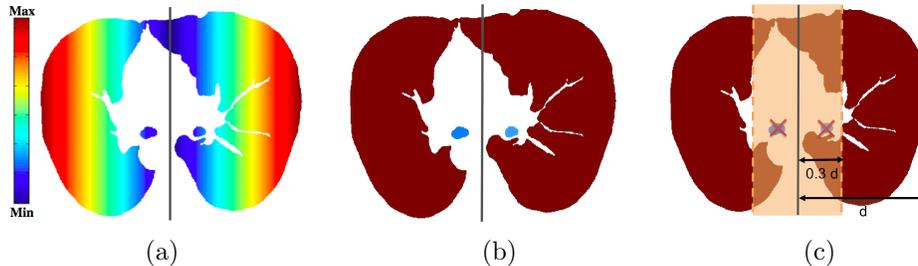


Figure 3: (a): Distance image to the reference axis (in gray). (b): Connected components labeled with the maximum distance found in their pixels. (c): Dynamic threshold to remove air tracks.

### 3.3 Right-left lung identification and mask refinement

After removing the trachea and the primary bronchi, two scenarios are present: either the lungs were already 3D-disconnected or they seemed to be merged by the parenchyma, resulting in a single

connected 3D component. An algorithm going through the sorted slices was designed to predict the best boundary in those slices where the lungs were connected. First, an initialization of the right (R) and left (L) labels is performed in the first slice presenting two significant CC. Then, the following slices with two CC (so-called 2-CC slices) are consistently labeled by projecting the labels from the previous slice. For the slices presenting only one CC (so-called, 1-CC slices) (see Fig. 4a), the algorithm applies a dilation on the labeled regions from the previous 2-CC slice, and projects them into the region of the current CC. The resulting labeled region contains pixels with one label (R or L), and with two labels (both R and L). This process propagates a boundary assumption to the current slice depending on the previous slice. This propagation results in a different labeling if the slices are selected in ascending or in descending order. Hence, the process is executed in both directions and the results are fused. The pixels with double label and the pixels with different label due to the double execution define a region of conflicts, as it is shown in Figure 4b. Then, a K-nearest neighbor algorithm [DHS01] in 3D is applied to decide the best label for each pixel of this region. Other small regions with no label after the procedure are labeled using the adjacent slices. Once both lungs are identified, the holes and the cavities are filled for each lung mask independently, achieving the result presented in Figure 4c.

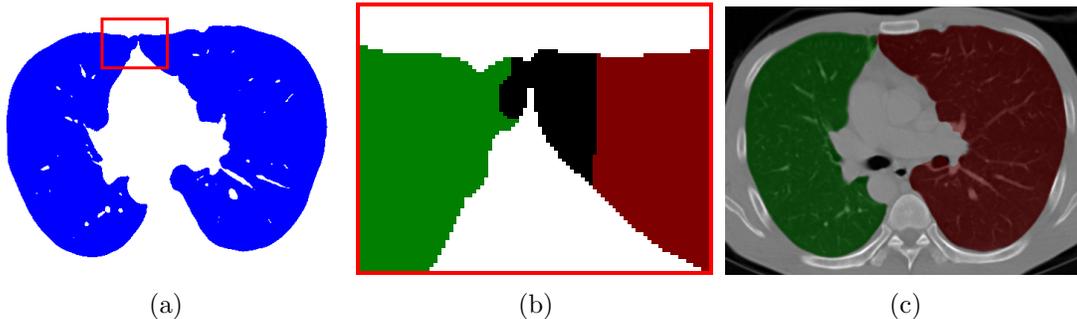


Figure 4: (a) Axial slice presenting only one connected component. The region in the red box shows where the two lungs are connected. (b) Detail of the merging zone in (a): In black, pixels with double label (R and L) due to the procedure explained in Section 3.3. (c) Final refined mask after identifying left and right lungs.

## 4 Results

The results shown in this section were provided by the organizers of the VISCERAL Grand Challenge at ISBI 2015. Table 1 shows a subset of the most relevant results. All results are published on the VISCERAL website. The evaluation was performed on the test set detailed in Section 2. The system presented in Section 3 showed to be one of the best algorithms presented in this edition, achieving a minimum Dice coefficient of 0.972 for both lungs in CT and CTce, and a maximum Hausdorff distance of 0.052.

## 5 Conclusions

The method presented in this paper introduces a new method for the extraction of the respiratory system in chest CT volumes. This initial step clearly separates the regions of interest, allowing to apply a fast K-Means clustering with a fixed number of 2 clusters. It detects the lung regions in a larger gray-level range than standard thresholding. Moreover, the extraction of the air tracks and the posterior differentiation of the lungs were done with simple geometric techniques that are computationally inexpensive. The procedures provide a fast system for segmenting the lungs in CT images that can be applied for large datasets. Furthermore, all steps rely on anatomical assumptions

Table 1: Table showing a subset of the performance measures provided by the VISCERAL Challenge. The best results for each modality and lung are highlighted in bold.

	Dice coefficient				Average Hausdorff distance			
	CT		CTce		CT		CTce	
	LL	RL	LL	RL	LL	RL	LL	RL
Our method	<b>0.972</b>	0.974	<b>0.974</b>	<b>0.973</b>	0.050	0.046	<b>0.050</b>	<b>0.052</b>
Participant 2	<b>0.972</b>	<b>0.975</b>	0.956	0.963	<b>0.043</b>	<b>0.038</b>	0.071	0.065
Participant 3	0.961	0.970	0.972	0.971	0.356	0.096	0.076	0.070
Participant 4	<b>0.972</b>	<b>0.975</b>	—	—	0.045	0.043	—	—
Participant 5	0.952	0.957	0.966	0.966	0.101	0.094	0.069	0.069

and require no training. The method showed almost perfect performance in CT and CTce. The presented segmentation can be applied directly to new CT scans with no further modifications. The participation in the VISCERAL challenge proved the reliability of this new efficient and fully automatic method, achieving an average Dice coefficient of 0.973 and an average Hausdorff distance of 0.0495.

## 6 Acknowledgments

This work was partly supported by the Swiss National Science Foundation in the PH4D project (grant agreement 320030-146804).

## References

- [DHS01] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification*. Wiley-Interscience, 2001.
- [EBFFR02] Ayman El-Baz, Aly A. Farag, Robert Falk, and Renato La Rocca. Detection, visualization, and identification of lung abnormalities in chest spiral CT scans: Phase I. In *Proceedings of the International Conf. on Biomedical Engineering*, Cairo, Egypt, 2002.
- [HHR01] Shiyong Hu, Eric A. Hoffman, and Joseph M. M. Reinhardt. Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. *IEEE Transactions on Medical Imaging*, 20(6):490–498, 2001.
- [IM03] S.G Armato III and H MacMahon. Automated lung segmentation and computer-aided diagnosis for thoracic {CT} scans. *International Congress Series*, 1256(0):977–982, 2003. {CARS} 2003. Computer Assisted Radiology and Surgery. Proceedings of the 17th International Congress and Exhibition.
- [LNC07] W. Li, S.D. Nie, and J.J. Cheng. A fast automatic method of lung segmentation in ct images using mathematical morphology. In *World Congress on Medical Physics and Biomedical Engineering 2006*, volume 14 of *IFMBE Proceedings*, pages 2419–2422. Springer Berlin Heidelberg, 2007.
- [Mac67] James MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. University of California Press, 1967.
- [Soi03] Pierre Soille. *Morphological Image Analysis: Principles and Applications*, chapter 6. Springer, second edition edition, 2003.