

# 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<sup>+</sup>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

---

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

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<sup>st</sup>, 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<sup>+</sup>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:

$$\hat{T} = \arg \min_T E(T, X, A_n) \quad (1)$$

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 ( $\psi$ ) and pairwise ( $\Psi$ ) potentials over discrete labels as follows:

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

where  $\Omega$  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.  $\lambda$  is the pairwise weight.

For robust and smooth solution of (2), an efficient method was proposed in [GKT<sup>+</sup>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<sup>+</sup>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:

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

## 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<sup>+</sup>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  $L_S = \{1, \dots, N_S\}$  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  $p$ :

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

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 reliability 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:

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

To obtain the weights  $w$ , we use local normalized cross correlation (LNCC, [CBD<sup>+</sup>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:

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

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

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

which normalizes LNCC to the range  $[0, 1]$ .  $\gamma$  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  $\ell$  separately from the others using the two stages below. To localize the unknown voxel coordinates  $p^\ell$  of landmark  $\ell$  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  $A_n^\ell$  in the current atlas. Similarly, a box-shaped search region  $X^\ell$  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 ( $d_{\text{HW}}$ )	Max size ( $n_{\text{max}}$ )
Template image $A_n^\ell$	$p_{A_n}^\ell$	20 mm	$41^3$ voxels
Search region $X^\ell$	$p_{A_n}^\ell \cdot \frac{ X }{ A_n }$	–	$241^3$ 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  $\text{SSD}^a \cdot \text{COR}^b$  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  $p_n^\ell$  for landmark  $\ell$  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  $A_n^\ell$  centered at given landmark location  $p_{A_n}^\ell$  in atlas image  $A_n$
  - Crop a large search region  $X^\ell$  centered around a grossly approximated location in  $X$
  - Compute  $\text{SSD}(A_n^\ell, X^\ell)$  and  $\text{COR}(A_n^\ell, X^\ell)$
  - $p_n^\ell = \arg \max_i (\text{SSD}[i]^2 \cdot \text{COR}[i]^3)$
- $p^\ell = \text{median} \{p_n^\ell \mid \forall_n\}$

## 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.  $\{1,2,\dots,P\}$  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.

Modality	Anatomy1				ISBI challenge			
	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

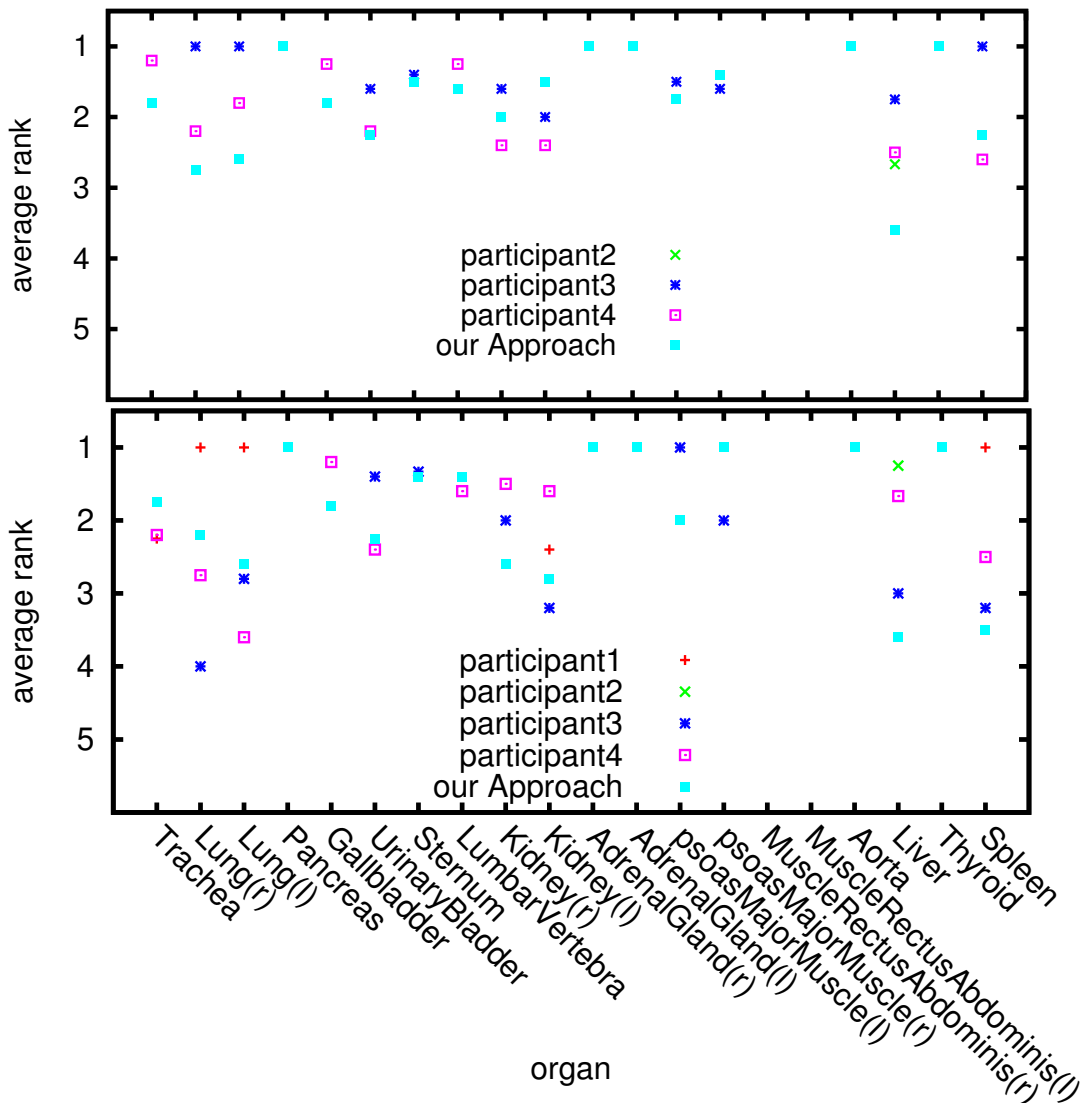


Figure 1: Average segmentation rank (by Dice coefficient) for each participant and organ for the CT (top) and CTce (bottom) modalities.

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<sup>+</sup>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<sup>+</sup>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<sup>+</sup>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<sup>+</sup>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.