Multiatlas Segmentation Using Robust Feature-Based Registration

Citation for the original published paper (version of record):
Fejne, F., Landgren, M., Alvén, J. et al (2017). Multiatlas Segmentation Using Robust Feature-Based Registration. , Cloud-Based Benchmarking of Medical Image Analysis: 203-218.
http://dx.doi.org/10.1007/978-3-319-49644-3_12

Chapter 12
Multiatlas Segmentation Using Robust Feature-Based Registration

Frida Fejne, Matilda Landgren, Jennifer Alvén, Johannes Ulén, Johan Fredriksson, Viktor Larsson, Olof Enqvist and Fredrik Kahl

Abstract This paper presents a pipeline which uses a multiatlas approach for multiorgan segmentation in whole-body CT images. In order to obtain accurate registrations between the target and the atlas images, we develop an adapted feature-based method which uses organ-specific features. These features are learnt during an offline preprocessing step, and thus, the algorithm still benefits from the speed of feature-based registration methods. These feature sets are then used to obtain pairwise non-rigid transformations using RANSAC followed by a thin-plate spline refinement or NiftyReg. The fusion of the transferred atlas labels is performed using a random forest classifier, and finally, the segmentation is obtained using graph cuts with a Potts model as interaction term. Our pipeline was evaluated on 20 organs in 10 whole-body CT images at the VISCERAL Anatomy Challenge, in conjunction

---

F. Fejne and M. Landgren—The authors assert equal contribution and joint first authorship.

F. Fejne · J. Alvén · O. Enqvist · F. Kahl
Department of Signals and Systems, Chalmers University of Technology,
Gothenburg, Sweden
e-mail: fejne@chalmers.se

J. Alvén
e-mail: alven@chalmers.se

O. Enqvist
e-mail: olof.enqvist@chalmers.se

F. Kahl
e-mail: fredrik.kahl@chalmers.se

M. Landgren (✉) · J. Ulén · J. Fredriksson · V. Larsson · F. Kahl
Centre for Mathematical Sciences, Lund University, Lund, Sweden
e-mail: matilda@maths.lth.se

J. Ulén
e-mail: ulen@maths.lth.se

J. Fredriksson
e-mail: johanf@maths.lth.se

V. Larsson
e-mail: viktorl@maths.lth.se

A. Hanbury et al. (eds.), *Cloud-Based Benchmarking of Medical Image Analysis*, DOI 10.1007/978-3-319-49644-3_12

with the International Symposium on Biomedical Imaging, Brooklyn, New York, in April 2015. It performed best on majority of the organs, with respect to the Dice index.

12.1 Introduction

Segmentation of anatomical structures is a fundamental task in medical image analysis. It has several applications such as localization of organs, detection of tumours or other pathological structures, and the results can, for example, serve as input to computer-aided diagnosis (CAD) systems. Multiorgan segmentation is useful, e.g. in radiotherapy planning [25], where not only the location of the tumour is of most interest, but also the location of the surrounding (vital) organs. Furthermore, it can also be used in the preparation of and during computer-assisted surgery [30]. Automated methods are preferable due to the time-consuming task to do the segmentations manually and the need of a skilled expert.

In this paper, we propose a pipeline that uses a multiatlas approach for an automatic multiorgan segmentation for CT images. The segmentation of each organ is independent of the others and we show that very reliable organ localization can be obtained using (i) robust optimization techniques for registration, (ii) learned feature correspondences and (iii) refinement with a random forest classifier and graph cut segmentation.

12.1.1 Related Work

Multiatlas methods for segmentation, which were first introduced in [11, 18, 26], have become a very popular choice in medical image analysis due to their excellent performance. The methods have been extensively used on brain MR images [6, 12, 31], on cardiac CTA data [17], for thoracic CT segmentation [10, 33] and multiorgan segmentation in CT images [35]. Multiatlas methods generally produce robust results but rely on multiple image registrations as each atlas image is registered to the target image. Image registration can be divided into two different approaches: feature-based and intensity-based registrations, see the surveys [16, 28]. Feature-based methods are generally very fast, but may have a risk of failing due to many outlier correspondences between the images. The intensity-based methods are on the contrary capable of producing accurate registrations but may be slow and are sensitive to initialization.

Multiatlas segmentation is a further development from single-atlas segmentation [21] and works as follows. In order to capture more anatomical variations and reduce the effect of registration errors, several single-atlas segmentations are combined in the multiatlas approach. At first, pairwise registrations are computed between each atlas image and the target image, and thereafter, the single-atlas segmentation labels are transferred to the target image according to the registrations. Next, a segmentation proposal is obtained by fusing the transferred labels. The label fusion can be performed using several different methods, whereof the simplest one is majority voting for each voxel, see [11, 18, 26]. However, there exist more sophisticated methods, for instance weighted voting [32], probabilistic reasoning using, e.g. the STAPLE algorithm [34] and different types of machine-learning approaches [27]. The fused segmentation proposal can be further refined into a final segmentation by using graph cut [3, 22] or random forest-based methods [9]. For a comprehensive survey of multiatlas segmentation methods and their applications, see [13].

As mentioned previously, multiatlas methods are often used to perform multiorgan segmentation. In the work of Wolz et al. [35], a hierarchical atlas is refined at three levels: global, organ and voxel levels. Instead of utilizing all the available atlases, the authors choose the most suitable ones. Similar to our paper, their final segmentation is obtained using graph cuts, and the evaluation of the algorithm results in relatively high Dice index on the liver, kidneys, pancreas and spleen.

Another example of multiorgan segmentation is the probabilistic multiatlas used by Chu et al. [4] for abdominal segmentation. They divide the image space into N subspaces and compute weights for the probabilistic atlas at both for a global level ($N = 1$) and for each subspace. The organ segmentation is then obtained by a maximum a posteriori estimation and graph cuts. The method is evaluated on different numbers of subspaces, where the best performance is obtained for $N = 64$.

Furthermore, another variant of multiatlas segmentation of abdominal organs was used in a recent paper by Xu et al. [36]. Atlas selection and label fusion were done using a reformulation of the selective and iterative method for performance-level estimation (SIMPLE) method. The authors developed a method for atlas selection, which was regularized by a Bayesian prior, learnt from context information. When evaluating the proposed method, it outperformed the compared methods, among them [35], on 11 out of 12 organs.

In [23], Okada et al. present an approach to multiorgan segmentation that uses conditional shape-location combined with unsupervised intensity priors. In their work, an organ correlation graph is used to steer the order for which the organs are segmented by utilizing spatial correlation. In addition, the authors also developed a method for modelling conditional shape-location priors.

12.1.2 Our Approach

In this paper, we propose a pipeline for segmentation of 20 different organs in whole-body CT images. The algorithm uses standard multiatlas segmentation for initial spatial localization. For the pairwise registrations between the target image and all the atlas images, we use an adapted feature-based method that has been designed to reduce the risk of establishing incorrect point-to-point correspondences between the image pairs. The main contribution of this paper is a method to identify reliable organ-specific feature points among the atlas images. The speed of general feature-based registration methods is beneficial to the algorithm, especially since this identification is done in an offline preprocessing step. We fuse the transferred labels by training a random forest classifier, and the final segmentation is then obtained with graph cuts. The pipeline is described in detail in Sect. 12.2. In Sect. 12.3, we present the results from the VISCERAL Anatomy Challenge [15], as well as a detailed evaluation of the different steps in the pipeline for some of the organs.

12.2 Methods

Our pipeline for multiorgan segmentation contains the following three main steps:

1. *Pairwise Registration.* For a particular organ, the atlases are registered to the target image in two steps: first, *subsets* of the features in each atlas image are selected and matched to the features in the target image. Next, a non-rigid transformation between the atlases and the target image is estimated using RANSAC followed by a thin-plate spline (TPS) refinement or an intensity-based free-form deformation using NIFTYREG. See Sect. 12.2.1.
2. *Label Fusion with a Random Forest Classifier.* The pairwise registrations give us a rough estimate of the location of the target organ. However, the accuracy of the solution after the registration can be further improved by taking the local appearance surrounding the target organ into account. In order to do this, we train a random forest classifier that is used to fuse the transferred atlas labels after the registration. See Sect. 12.2.2.
3. *Graph Cut Segmentation with a Potts Model.* The segmentation is further refined by encouraging spatial smoothness between neighbouring pixels. For this, we formulate the labelling problem as an optimization problem and solve it using graph cuts. See Sect. 12.2.3.

These steps of our pipeline will be presented in detail in the following sections. Each organ is segmented individually in our multiatlas approach.

12.2.1 Pairwise Registration

12.2.1.1 Determination of Organ-Specific Feature Sets

For each organ and atlas image $I_i \in I = \{I_1, \ldots, I_n\}$, a subset of features that is designed to produce a reliable registration of the organ of interest is determined. The basic idea is to evaluate how well the extracted feature points in an atlas image match to other feature points in the remaining atlas images. In order to quantify the quality of a matched feature correspondence, we first establish the so-called golden transformations between the two atlas images around the organ of interest, based on precomputed (ground truth) landmark correspondences. Hence, if a feature point is matched to a feature point in another image, then this point-to-point correspondence should be consistent with the golden transformation, provided the correspondence is correct. Otherwise, it is likely to be an outlier. Feature points that always form inlier correspondences are good candidates for reliable registration, and these points determine the organ-specific feature sets.

Establishing golden transformations. For each atlas image, we compute golden transformations between I_i and the other atlas images in $I \setminus I_i$. This is done by applying TPS to precomputed landmark correspondences using the method proposed in [5]. The landmark correspondences are computed through accurate non-rigid registrations between a randomly chosen reference atlas in the atlas set and each of the remaining atlases. The registration uses two channels, the image intensity and the ground truth mask of the organ of interest, and maximizes the similarity between these according to the normalized mutual information (NMI) measure using NiftyReg [1, 24]. With the obtained displacement field, the mesh points of a triangular mesh of the ground truth surfaces are transformed to the coordinate system of the reference atlas. For each mesh point (landmark) of the triangulation of the reference ground truth surface, the closest point on each transformed triangulated surface is found using an algorithm based on [8] and chosen as the corresponding landmark.

Feature extraction. For each atlas image, $I_i \in I$, we calculate sparse features according to the method proposed by Svärm et al. [29]. A feature point is denoted $\mathbf{f} = (i, \mathbf{x}, \mathbf{d})$ where i is the index of the image, and $\mathbf{x}$ and $\mathbf{d}$ are the coordinates and the description vector for the point, respectively. Only the features that lie sufficiently close to the organ are considered; that is, we keep features with distance to the organ, δ, less than a predefined threshold, $D_{\max}$. For the whole atlas I, we thus obtain $F = \{F_1, \ldots, F_n\}$, where F_i is the set of feature points for I_i. For each atlas image I_i, the points in F_i are matched to the other feature sets in $F \setminus F_i$ using a symmetric neighbour approach, thus establishing point-to-point correspondences between I_i and the other atlas images.

Computation of organ-specific feature sets. Next, we proceed by applying the golden transformations to the feature points in $F \setminus F_i$ that have been matched to points in F_i, in order to transform them into the same coordinate system. Furthermore, we calculate the residuals between the coordinates of the feature points for I_i and the corresponding feature points for the other atlas images after the transformation. If feature point $\mathbf{f}_k \in F_i$ is matched to $\mathbf{f}_{\tilde{k}} \in F_j$, the residual is defined as

$$r_{\tilde{k}} = \|\mathbf{x}_k - \hat{\mathbf{T}}^G_{j,i} \circ \mathbf{x}_{\tilde{k}}\|_2, \tag{12.1}$$

where $\hat{\mathbf{T}}^G_{j,i}$ denotes the golden transformation between I_i and I_j. Each feature in F_i receives a score that is a weighted sum of the normalized residuals for all the corresponding points in the other atlas images and the normalized distance from the organ. More precisely, the score for $\mathbf{f}_k \in F_i$ is

$$\text{Score}[\mathbf{f}_k] = s_\delta + \omega_r \sum_{\tilde{k} \in \mathbf{R}} s_{\tilde{k}}, \tag{12.2}$$

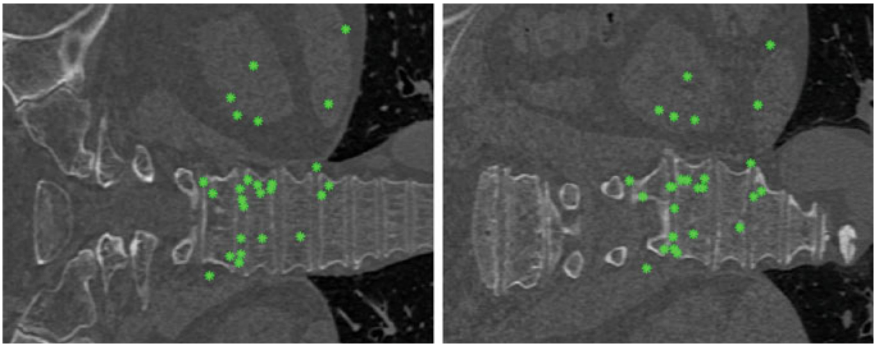

Fig. 12.1 Two CT slices of a target image (*left*) and an atlas image (*right*) with corresponding features after RANSAC for lumbar vertebra 1

where

$$\begin{aligned} s_\delta &= (D_{\max} - \delta_k)/D_{\max}, \\ s_{\tilde{k}} &= \max(T - r_{\tilde{k}}, 0)/T. \end{aligned} \tag{12.3}$$

Here, ω_r is the importance weight for the residuals; T is a predefined threshold; and $\mathbf{R}$ is the set of feature points for the atlas images $I \setminus I_i$, which have been matched to $\mathbf{f}_k$. The features are ranked according to their score, and those with the highest scores are kept and used in the registration step. This procedure is relatively time-consuming, but it is an offline process and thus only done once.

12.2.1.2 Pairwise Registration with RANSAC

At run-time, we estimate pairwise affine transformations between the target image I_t and all the images in the atlas set, using the feature sets obtained in the previous section. We apply RANSAC with the truncated l_2 norm as a cost function for outlier removal. An example of inlier feature correspondences for the lumbar vertebra 1 is illustrated in Fig. 12.1. Furthermore, for about half of the organs, the final coordinate transforms between I and I_t are obtained by TPS interpolation between the remaining correspondences. For the rest of the organs, we find that the best transformations are obtained by using NIFTYREG, with the affine transformation as initialization. These transformations are then used to transfer the labels of the atlas images into the same coordinate system as the target image, see Fig. 12.2.

12.2.2 Label Fusion with a Random Forest Classifier

We use the pairwise registrations to obtain a rough estimate of where the organ is located. This is done by fusing the transferred labels from each atlas into a so-called

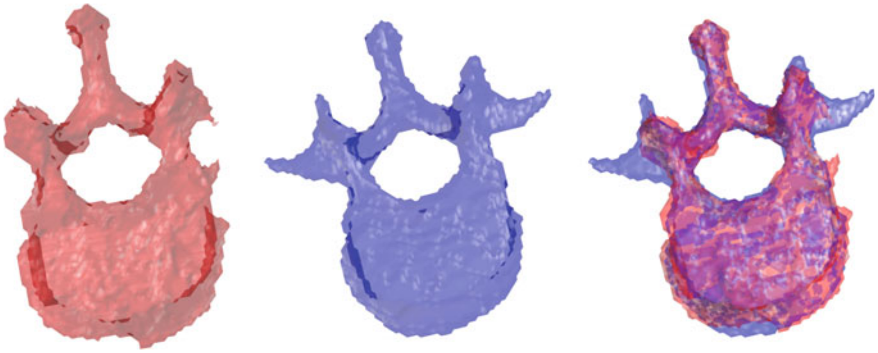

Fig. 12.2 Example of a registration of lumbar vertebra 1. *Left* the atlas ground truth mask. *Middle* the warped target mask after the registration. *Right* the masks overlaid in the same coordinate system

voxel map, P, in which each voxel can be interpreted as a measure of the likelihood of that voxel belonging to the organ, according to the pairwise registrations. More precisely, the map P is the normalized average of the warped target masks of each of the atlas images, so if half of the atlas images think that voxel i is organ, then $P(i) = 0.5$.

However, the map P largely ignores the local appearance around the target organ and in order to further improve the results, a random forest classifier is trained in an offline process. The classifier is then used to obtain a refined estimate of P, which will be denoted P_r. We implement this using Sherwood [7], which allows us to train and evaluate large random forest instances efficiently. The voxel map, P, and the target image, I, are used to compute a set of features for each voxel, which will be used as input to the random forest classifier. By smoothing I and P using a Gaussian kernel, we obtain two new volumes, which we refer to as I_s and P_s. Furthermore, for each organ we determine a threshold level, τ, for P and use this to construct a distance map, D_P, where each voxel in D_P equals the (signed) distance to the boundary surface of the binary volume $P > \tau$. For each voxel i, in each volume I, we thus obtain five features: $I(i)$, $I_s(i)$, $P(i)$, $P_s(i)$ and $D_P(i)$.

12.2.3 Graph Cut Segmentation with a Potts Model

The refined estimate, P_r, gives us a better estimation of the segmentation, but there is room for further improvement. The decision of whether voxel i should be classified as belonging to the organ or not is taken without considering the classification of neighbouring voxels, which may cause noisy and inaccurate estimates along the boundaries of the target organ. Thus, we can improve the segmentation by incorporating this information into the model, and in order to do so, we formulate our voxel labelling problem as an energy minimization problem and solve it using graph cuts [2]. More precisely, let $x_i \in \mathbf{L} = \{0, 1\}$ be a Boolean indicator variable for voxel i, that is, 1 if x_i is classified as belonging to the organ (foreground), and 0 if it belongs to the background. We are now seeking the labelling $\mathbf{x}^\star$ that minimizes the energy function of the form

$$E(\mathbf{x}^\star) = \sum_{i=1}^{n} D_i(x_i) + \sum_{i,j \in \mathbf{N}} V_{i,j}(x_i, x_j), \tag{12.4}$$

where the data term, $D_i(x_i)$, measures how well the label x_i suits voxel i, given the target image I_t, and $V_{i,j}(x_i, x_j)$ is an interaction term that regularizes the solution by assigning different costs to neighbouring voxels, which depend on the labels they take. Furthermore, n is the number of voxels in the image, and $\mathbf{N}$ defines the neighbourhood system of the voxels.

The output from the random forest classifier is used for the data term, in which voxel i is set to take the value $1/2 - P_r(i)$ if $x_i = 1$, and zero otherwise, i.e. $D_i(x_i) = x_i(\frac{1}{2} - P_r(i))$. Thus, this model makes it more likely that voxel i is classified as foreground if $P_r(i) \in [0.5, 1]$, and background if $P_r(i) \in [0, 0.5]$. As interaction term, we use Potts model, which regularizes the resulting segmentation by penalizing neighbouring voxels if they receive different labels. It assigns a cost to two neighbouring voxels, i and j, according to $\lambda[x_i \neq x_j]$, where λ is a regularization weight, and x_i and x_j are the labels for voxels i and j, respectively. This interaction cost can also be expressed as $V_{i,j}(x_i, x_j) = \lambda(x_i(1 - x_j) + x_j(1 - x_i))$.

Thus, the final segmentation, $\mathbf{x}^\star$, is obtained by solving the following minimization problem:

$$\mathbf{x}^\star = \operatorname*{argmin}_{\mathbf{x} \in \{0,1\}^n} \sum_{i=1}^{n} x_i \left(\frac{1}{2} - P_r(i)\right) + \lambda \sum_{i=1}^{n} \sum_{j \in \mathbf{N}(i)} \mu_{ij} x_i (1 - x_j), \tag{12.5}$$

where μ_{ij} is a variable that compensates for anisotropic resolution [19], and $\mathbf{N}(i)$ is the set of voxels in the neighbourhood of voxel i, which is set to be 6-connected for all the organs, i.e. each voxel that touches a side of a voxel is a neighbour. Since the cost function in (12.5) is submodular, it can be minimized efficiently using graph cuts [20] with the implementation of [14]. During the minimization, we process a smaller volume, that is, a cut-out around the zero level of the thresholded voxel map $P > \tau$, which allows us to save memory and speed up the calculations.

A comparison between the resulting segmentation of the spleen with graph cuts, using the initial voxel map, P, and the refined probability map after the random forest step, P_r, is illustrated in Fig. 12.3.

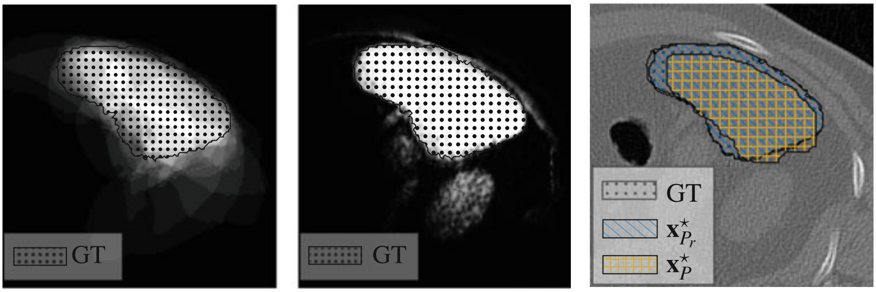

Fig. 12.3 Example of the resulting probability estimates and segmentation of the spleen for one CT slice; in each image, the ground truth (GT) is indicated. *Left* the initial probability, P. *Middle* the probability given by random forest, P_r. *Right* the resulting segmentation $\mathbf{x}^{\star}_{P}$ using P and $\mathbf{x}^{\star}_{P_r}$ using P_r overlaid on the original image

12.3 Experimental Evaluation

The different steps in the pipeline involve some tuning parameters and these have been set as follows. The 20 whole-body CT images that are available in the challenge are split into one training and validation (atlas) set consisting of the first 15 images, while the remaining five serve as a test set. For the registration, the parameters are determined by leave-one-out cross-validation of the atlas images, while the remaining ones are used as validation images for the random forest classification and graph cut segmentation.

For the computation of the organ-specific feature sets, the same parameter settings are used for all organs. At first, around 8,000–10,000 features are extracted from a whole-body CT image in less than 30 s. When ranking the features, the maximal distance is set to $D_{max} = 100$ mm, the threshold $T = 15$, and the importance weight for the residuals, ω, is set to 10. We have found empirically that the 300 features with the highest score can be used to provide robust and reliable registration. RANSAC is run 500,000 iterations, and the truncation threshold for the l_2 cost function is set to 30 mm. The value of the standard deviation, σ, for the Gaussian kernel in the smoothing of P for the random forest classifier is 1. Table 12.1 lists parameters and settings for each individual organ. Note that in the segmentation of some of the organs, we do not use a random forest classifier. This is either because the organ has a very large volume, which makes the computations heavy, or because the classifier does not improve the results at all. Furthermore, we do not use the learned features for the lungs. In our experience, very simple methods yield accurate segmentations of the lungs and that is also the case when we use ordinary features.

The single most time-consuming online part of the algorithm is the registration and the time needed strongly depends on the size of the organ and the choice of registration method. NiftyReg takes around 100–200 s per registration compared to TPS for which a registration takes less than 10s regardless of the organ type. However, we have found empirically that NiftyReg performs a lot better for ten of

Table 12.1 Parameters used in the pipeline

Organ	Registration method	RANSAC threshold	Random forest	τ	λ
Trachea	NiftyReg	20	Yes	0.25	0.35
Right lung	TPS	30	No		0.35
Left lung	TPS	30	No		0.35
Pancreas	TPS	20	Yes	0.15	0.25
Gall bladder	TPS	25	No		0.15
Urinary bladder	TPS	50	Yes	1.45	0.35
Sternum	NiftyReg	20	Yes	0.05	0.30
Lumbar vertebra 1	NiftyReg	7	Yes	0.05	0.30
Right kidney	NiftyReg	25	Yes	1.00	0.35
Left kidney	NiftyReg	20	Yes	1.50	0.35
Right adrenal gland	TPS	30	No		0.20
Left adrenal gland	TPS	20	No		0.20
Right psoas major	NiftyReg	40	Yes	0.45	0.40
Left psoas major	NiftyReg	30	Yes	0.30	0.40
Muscle body of right rectus abdominis	NiftyReg	40	Yes	0.05	0.30
Muscle body of left rectus abdominis	NiftyReg	40	Yes	0.05	0.25
Aorta	NiftyReg	30	Yes	0.20	0.35
Liver	TPS	50	Yes	0.90	0.40
Thyroid gland	TPS	20	Yes	0.10	0.25
Spleen	TPS	30	Yes	0.85	0.30

the organs, see Table 12.1. If more images are added to the training set, the process of determining the organ-specific features, the registration of the atlas images in order to obtain the voxel map, P, and the training of the random forest classifier would have to be run again from start. The only difference for the online process is that we would have to perform one extra registration per added image.

12.3.1 Challenge Results

In our contribution to the VISCERAL Anatomy Challenge, all the 20 images available for training formed the atlas set in the final submission. The algorithm was evaluated on a test set consisting of 10 new whole-body CT images that only were available to the organizers of the challenge, and the evaluation took place at ISBI 2015. The results are measured using the Dice index, which is defined as Dice(S, G) $= 2|S \cap G|/(|S| + |G|)$, where S and G are the computed segmentation and ground truth, respectively. Thus, a perfect segmentation would yield Dice index 1, while a segmentation with no ground truth overlap would receive a Dice index of

Table 12.2 Final results measured in Dice index for whole-body CT images. Here, "*" means that no segmentation was provided

Organ	Our	CMIV	HES-SO	SIAT
Left kidney	**0.934**	0.896	0.784	*
Right kidney	**0.915**	0.796	0.790	*
Spleen	0.870	**0.910**	0.703	0.874
Liver	0.921	**0.936**	0.866	0.923
Left lung	**0.972**	0.961	**0.972**	0.952
Right lung	**0.975**	0.970	**0.975**	0.957
Urinary bladder	**0.763**	0.713	0.698	*
Muscle body of left rectus abdominis	**0.746**	*	0.551	*
Muscle body of right rectus abdominis	**0.679**	*	0.519	*
Lumbar vertebra 1	**0.775**	*	0.718	*
Thyroid	0.424	*	**0.549**	*
Pancreas	0.383	*	**0.408**	*
Left psoas major muscle	**0.861**	0.828	0.806	*
Right psoas major muscle	**0.847**	0.817	0.787	*
Gall bladder	0.190	*	**0.276**	*
Sternum	**0.847**	*	0.761	*
Aorta	**0.830**	*	0.753	*
Trachea	**0.931**	*	0.92	*
Left adrenal gland	0.282	*	**0.373**	*
Right adrenal gland	0.220	*	**0.355**	*
Average	**0.718**	*	0.678	*

0. Our results are reported in Table 12.2 together with the results from the strongest participants:

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

In summary, our algorithm provides the best results for 13 of the 20 organs.

12.3.2 Detailed Evaluation

In this section, we evaluate what kinds of benefits specific parts of the pipeline provide for lumbar vertebra 1, the left kidney and the spleen. Dice scores after

Table 12.3 Results for three organs after different steps in the pipeline, measured in Dice index. Here, OF indicates that ordinary feature sets were used instead of the organ-specific ones. $\mathbf{x}_P^\star$ is the map P, thresholded at τ and $\mathbf{x}^\star$ is the graph cut segmentation after the random forest step

Organ	$\mathbf{x}_P^\star$ (OF)	$\mathbf{x}^\star$ (OF)	$\mathbf{x}_P^\star$	$\mathbf{x}^\star$
Lumbar vertebra 1	0.730	0.682	0.858	0.884
Left kidney	0.891	0.901	0.891	0.896
Spleen	0.688	0.814	0.686	0.808

different steps in the pipeline, when using both organ-specific and ordinary feature sets, are presented in Table 12.3. We denote the segmentation that can be obtained after fusing the transferred labels from the registration with $\mathbf{x}_P^\star$ and it is the map P, thresholded at τ. The values in the table are the average results for the five images that formed the test set when using the 15 first as atlas set as described in the introduction to this section. Note that these values are not comparable with the ones in Table 12.2 since those were obtained with a different amount of training data and evaluated on images that are not accessible to us. The results in Table 12.3 should be compared among themselves in order to determine the contribution of different parts of the pipeline for the selected organs.

Clearly, the organ-specific features improve the results significantly for the segmentation of lumbar vertebra 1, which is demonstrated in Fig. 12.4. The left and the middle pictures in Fig. 12.5 show the segmentation of the spleen for CT image 19 before and after the graph cut part in the pipeline, respectively. It is clear from both the figures and the table that the random forest classifier improves the results a lot. However, note that the segmentation in the middle picture of Fig. 12.5 is not a valid spleen shape. The right picture of Fig. 12.5 illustrates the final segmentation for the left kidney. This segmentation is obtained using organ-specific features, and the result is quite accurate. However, the random forest classifier and the graph cut segmentation do not seem to further improve the Dice index for the left kidney according to the results in the table. Note that this does not necessarily mean that the segmentation is not improved. It is still possible that the produced solution is more regularized

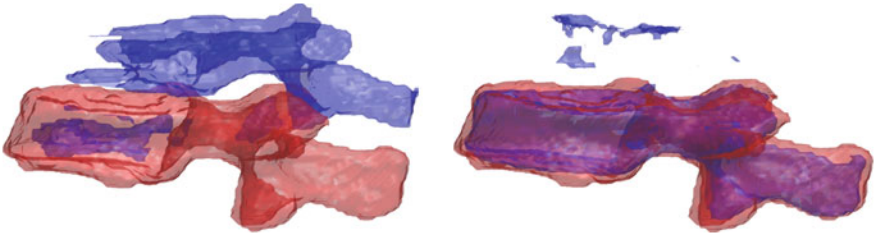

Fig. 12.4 Segmentation of lumbar vertebra 1. Here, the ground truth is *red* and our segmentation is *blue*. *Left* segmentation using ordinary features. The registration fails to find the correct vertebra as it is confused by a nearby vertebra. *Right* segmentation using organ-specific features. Now the correct vertebra is located

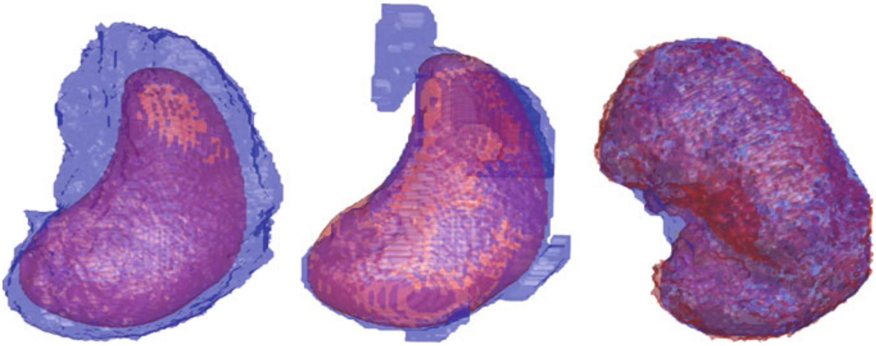

Fig. 12.5 Ground truth (*red*) and computed segmentation (*blue*) for some of the organs after different steps in the pipeline. Organ-specific features are used for all the segmentations. *Left* warped ground truth mask of the spleen after the registration. *Middle* graph cut segmentation of the spleen. *Right* graph cut segmentation of the *left* kidney

and accurate, just not according to the Dice index. Furthermore, the Dice scores for the left kidney and the spleen are more or less the same when using organ-specific features compared to when using ordinary features. In our experience, with a lot of training data the organ-specific features generally perform a little better than the ordinary features although that may not be the case for all the organs. It is, however, worth mentioning that there are certain limitations regarding the performance of the organ-specific features. It requires accurate landmark correspondences in order to establish reliable golden transformations, see Sect. 12.2.1. If this is not the case, it may not be advantageous to use organ-specific features.

12.4 Conclusions

In this paper, we have described an algorithm that uses a feature-based approach to multiatlas segmentation of organs in whole-body CT images. The results clearly demonstrate that this method manages to locate and segment the organs with the state-of-the-art results, and our approach outperforms the participants at the VISCERAL Anatomy Challenge on segmentation at ISBI 2015 for 13 out of 20 organs.

However, some parts of the algorithm could benefit from further work. For instance, incorporating prior information about the organ shapes into the pipeline would help to guarantee that the algorithm produces feasible organ shapes. A more thorough evaluation of the organ-specific feature sets could help determine individual parameter settings for the organs, as well as help in explaining why the method improves the results for some organs but not for all. Furthermore, adding additional features to the random forest classifier would likely yield better results. Moreover, the calculations could be sped up, e.g. by considering alternative registration methods.

References

1. Alvén J, Norlén A, Enqvist O, Kahl F (2016) Überatlas: fast and robust registration for multi-atlas segmentation. Pattern Recognit Lett. doi:10.1016/j.patrec.2016.05.001
2. Boykov Y, Veksler O, Zabih R (2001) Fast approximate energy minimization via graph cuts. IEEE Trans Pattern Anal Mach Intell 23(11):1222–1239
3. Candemir S, Jaeger S, Palaniappan K, Musco J, Singh R, Xue Z, Karargyris A, Antani S, Thoma G, McDonald C (2014) Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. IEEE Trans Med Imaging 33(2):577–590
4. Chu C, Oda M, Kitasaka T, Misawa K, Fujiwara M, Hayashi Y, Nimura Y, Rueckert D, Mori K (2013) Multi-organ segmentation based on spatially-divided probabilistic atlas from 3D abdominal CT images. In: Mori K, Sakuma I, Sato Y, Barillot C, Navab N (eds) MICCAI 2013. LNCS, vol 8150. Springer, Heidelberg, pp 165–172. doi:10.1007/978-3-642-40763-5_21
5. Chui H, Rangarajan A (2000) A new algorithm for non-rigid point matching. In: IEEE conference on computer vision and pattern recognition, vol 2, pp 44–51
6. Chupin M, Gérardin E, Cuingnet R, Boutet C, Lemieux L, Lehéricy S, Benali H, Garnero L, Colliot O (2009) Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from adni. Hippocampus 19(6):579–587
7. Criminisi A, Shotton J, Konukoglu E (2011) Decision forests: a unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. Found Trends® Comput Graph Vis 7:81–227
8. Eberly D (2008) Distance between point and triangle in 3D. Geometric Tools, LLC. http://www.geometrictools.com/
9. Han X (2013) Learning-boosted label fusion for multi-atlas auto-segmentation. In: Wu G, Zhang D, Shen D, Yan P, Suzuki K, Wang F (eds) MLMI 2013. LNCS, vol 8184. Springer, Cham, pp 17–24. doi:10.1007/978-3-319-02267-3_3
10. Han X, Hoogeman MS, Levendag PC, Hibbard LS, Teguh DN, Voet P, Cowen AC, Wolf TK (2008) Atlas-based auto-segmentation of head and neck CT images. In: Metaxas D, Axel L, Fichtinger G, Székely G (eds) MICCAI 2008. LNCS, vol 5242. Springer, Heidelberg, pp 434–441. doi:10.1007/978-3-540-85990-1_52
11. Heckemann RA, Hajnal JV, Aljabar P, Rueckert D, Hammers A (2006) Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. NeuroImage 33(1):115–126
12. Heckemann RA, Keihaninejad S, Aljabar P, Rueckert D, Hajnal JV, Hammers A (2010) Improving intersubject image registration using tissue-class information benefits robustness and accuracy of multi-atlas based anatomical segmentation. Neuroimage 51(1):221–227
13. Iglesias JE, Sabuncu MR (2015) Multi-atlas segmentation of biomedical images: a survey. Med Image Anal 24(1):205–219
14. Jamriška O, Sýkora D, Hornung A (2012) Cache-efficient graph cuts on structured grids. In: IEEE conference on computer vision and pattern recognition, pp 3673–3680
15. Jiménez-del-Toro OA, Müller H, Krenn M, Gruenberg K, Taha AA, Winterstein M, Eggel I, Foncubierta-Rodríguez A, Goksel O, Jakab A, Kontokotsios G, Langs G, Menze B, Fernandez TS, Schaer R, Walleyo A, Weber M, Cid YD, Gass T, Heinrich M, Jia F, Kahl F, Kechichian R, Mai D, Spanier AB, Vincent G, Wang C, Wyeth D, Hanbury A (2016) Cloud-based evaluation of anatomical structure segmentation and landmark detection algorithms: VISCERAL anatomy benchmarks. IEEE Trans Med Imaging 35(11):2459–2475
16. Khalifa F, Beache G, Gimel'farb G, Suri J, El-Baz A (2011) State-of-the-art medical image registration methodologies: a survey. In: El-Baz AS, Rajendra Acharya U, Mirmehdi M, Suri JS (eds) Multi modality state-of-the-art medical image segmentation and registration methodologies. Springer, Heidelberg, pp 235–280
17. Kirisli HA, Schaap M, Klein S, Neefjes LA, Weustink AC, van Walsum T, Niessen WJ (2010) Fully automatic cardiac segmentation from 3D CTA data: a multi-atlas based approach. In: SPIE medical imaging, San Diego, USA

18. Klein A, Mensh B, Ghosh S, Tourville J, Hirsch J (2005) Mindboggle: automated brain labeling with multiple atlases. BMC Med Imaging 5(1):7
19. Kolmogorov V, Boykov Y (2005) What metrics can be approximated by geo-cuts, or global optimization of length/area and flux. In: IEEE international conference on computer vision, vol 1, pp 564–571
20. Kolmogorov V, Zabin R (2004) What energy functions can be minimized via graph cuts? IEEE Trans Pattern Anal Mach Intell 26(2):147–159
21. Kurkure U, Le Y, Ju T, Carson J, Paragios N, Kakadiaris I (2011) Subdivision-based deformable model for geometric atlas fitting. In: IEEE international conference on computer vision, Barcelona, Spain
22. Lee JG, Gumus S, Moon CH, Kwoh CK, Bae KT (2014) Fully automated segmentation of cartilage from the MR images of knee using a multi-atlas and local structural analysis method. Med Phys 41(9)
23. Okada T, Linguraru MG, Hori M, Summers RM, Tomiyama N, Sato Y (2015) Abdominal multi-organ segmentation from CT images using conditional shape-location and unsupervised intensity priors. Med Image Anal 26(1):1–18
24. Ourselin S, Roche A, Subsol G, Pennec X, Ayache N (2001) Reconstructing a 3D structure from serial histological sections. Image Vis Comput 19(1–2):25–31
25. Pekar V, McNutt TR, Kaus MR (2004) Automated model-based organ delineation for radiotherapy planning in prostatic region. Int J Radiat Oncol Biol Phys 60(3):973–980
26. Rohlfing T, Brandt R, Menzel R, Maurer CR Jr (2004) Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. NeuroImage 21(4):1428–1442
27. Sanroma G, Wu G, Gao Y, Shen D (2014) Learning-based atlas selection for multiple-atlas segmentation. In: IEEE conference on computer vision and pattern recognition, Columbus, USA
28. Sotiras A, Davatzikos C, Paragios N (2013) Deformable medical image registration: a survey. IEEE Trans Med Imaging 32(7):1153–1190
29. Svärm L, Enqvist O, Kahl F, Oskarsson M (2015) Improving robustness for inter-subject medical image registration using a feature-based approach. In: International symposium on biomedical imaging
30. Taylor R, Stoianovici D (2003) Medical robotics in computer-integrated surgery. IEEE Trans Robot Autom 19(5):765–781
31. van der Lijn F, den Heijer T, Breteler MM, Niessen WJ (2008) Hippocampus segmentation in MR images using atlas registration, voxel classification, and graph cuts. NeuroImage 43(4):708–720
32. Wang H, Suh J, Dass SR, Pluta J, Craige C, Yushkevich P (2013) Multi-atlas segmentation with joint label fusion. IEEE Trans Pattern Anal Mach Intell 35(3):611–623
33. Wang L, Chen KC, Gao Y, Shi F, Liao S, Li G, Shen SGF, Yan J, Lee PKM, Chow B, Liu NX, Xia JJ, Shen D (2014) Automated bone segmentation from dental CBCT images using patch-based sparse representation and convex optimization. Med Phys 41(4):043503
34. Warfield S, Zou K, Wells W (2004) Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans Med Imaging 27(3):903–921
35. Wolz R, Chu C, Misawa K, Fujiwara M, Mori K, Rueckert D (2013) Automated abdominal multi-organ segmentation with subject-specific atlas generation. IEEE Trans Med Imaging 32(9):1723–1730
36. Xu Z, Burke RP, Lee CP, Baucom RB, Poulose BK, Abramson RG, Landman BA (2015) Efficient multi-atlas abdominal segmentation on clinically acquired CT with SIMPLE context learning. Med Image Anal 24(1):18–27