

Multi-organ Abdominal CT Segmentation Using Hierarchically Weighted Subject-Specific Atlases

Robin Wolz¹, Chengwen Chu², Kazunari Misawa³,
Kensaku Mori^{2,4,*}, and Daniel Rueckert^{1,*}

¹ Imperial College London, London, UK

² Department of Media Science, Nagoya University, Nagoya, Japan

³ Aichi Cancer Center, Nagoya, Japan

⁴ Information and Communications Headquarters, Nagoya University, Japan

Abstract. A robust automated segmentation of abdominal organs can be crucial for computer aided diagnosis and laparoscopic surgery assistance. Many existing methods are specialised to the segmentation of individual organs or struggle to deal with the variability of the shape and position of abdominal organs. We present a general, fully-automated method for multi-organ segmentation of abdominal CT scans. The method is based on a hierarchical atlas registration and weighting scheme that generates target specific priors from an atlas database by combining aspects from multi-atlas registration and patch-based segmentation, two widely used methods in brain segmentation. This approach allows to deal with high inter-subject variation while being flexible enough to be applied to different organs. Our results on a dataset of 100 CT scans compare favourable to the state-of-the-art with Dice overlap values of 94%, 91%, 66% and 94% for liver, spleen, pancreas and kidney respectively.

1 Introduction

The accurate segmentation of organs like the liver, pancreas and kidneys on abdominal computed tomography (CT) scans form an important input to computer aided diagnosis (CAD) systems and to laparoscopic surgery assistance. The detailed segmentation and rendering of such structures can crucially assist clinicians in surgery planning and navigation. Further applications include cancer detection and staging, especially of pancreatic cancer [1]. Most previous work is based on either statistical shape models or probabilistic atlases learned on a training set and applied in combination with post-processing steps based on image intensities and morphology that are often specialized to a particular organ [2–5]. However, the high inter-subject variability in shape and location of abdominal organs, especially the pancreas, poses a challenge to generalized, population-based models and requires more subject-specific prior knowledge. In [1], a subject-specific segmentation model based on statistical shape models is presented. While this method achieves good segmentation results, it is specific

* Supported by RS-JSPS Research Cooperative Program, MEXT/JSPS KAKENHI.

and limited to pancreas segmentation. Linguraru et al.[5] propose a 4D graph model that incorporates patient-specific data for multi-organ segmentation.

Subject-specific, automated segmentation methods based on multi-atlas registration have been pioneered and are now widely used for brain segmentation of magnetic resonance (MR) imaging [6],[7]. Other application areas include cardiac segmentation on CT [8]. With multi-atlas registration, high accuracies are achieved for the segmentation of cortical and subcortical brain structures [6],[7]. A comparatively low variation between subjects in overall brain shape generally allows a good global alignment, providing atlas-patient correspondences that allow the generation of target specific segmentation priors [7]. The high shape and position variability in abdominal organs as well as the differences in global abdominal shape and potentially differences in field of view, forms a significant challenge to the pairwise image registration of abdominal scans. This hampers correspondence estimation on a localized voxel level but also the organ level which is a key requirement for atlas-based propagation strategies.

Patch-based segmentation [9] has recently been proposed as an alternative to multi-atlas registration in brain imaging. In this approach, a coarse alignment is sought between the target and atlas images, and patch-comparison between target and atlases is carried out in a local window to obtain the target labelling.

Previous work for abdominal segmentation based on spatial atlases employed general, population-based probabilistic models, e.g., [2]. Here, we propose an approach that is more sensitive to inter-subject variation. We generate a subject-specific probabilistic atlas for an unlabelled subject by selecting and aligning suitable atlases from a database, combining multi-atlas registration and patch-based refinement. We have evaluated the proposed method on a set of 100 abdominal CT scans for the segmentation of the liver, spleen, kidneys and pancreas and achieve a segmentation accuracy that compares favourably to existing methods.

2 Method

In the proposed approach, the final segmentation is obtained by establishing correspondences between every target voxel and multiple, manually labelled atlas images. Atlases are weighted on three scales: on a global level, the organ level and the voxel level. In the first step, the set of most suitable atlases from a database is selected for a new subject by measuring global image appearance. After aligning all pre-selected atlases with the target image, a local atlas weighting is carried out on an organ by organ basis. Finally, a patch-based segmentation refinement is applied to identify atlas labels at the voxel level.

Given a target image \mathcal{I} , the segmentation problem is formulated by assigning each voxel $x_i \in \mathcal{I}$, a label $l \in \{l_0, l_1, \dots, l_L\}$ with L anatomical labels and a background label l_0 . The labelling procedure is defined as a weighted fusion of expert votes defined on a database of N atlases $\mathbf{A} = \{\mathcal{A}_1, \dots, \mathcal{A}_N\}$:

$$l(x_i) = \frac{\sum_{n=1}^N w_n^g \sum_{x_j \in \mathcal{A}_n} w_j^o w_j^v l(x_j)}{\sum_{n=1}^N w_n^g \sum_{x_j \in \mathcal{A}_n} w_j^o w_j^v} \quad (1)$$

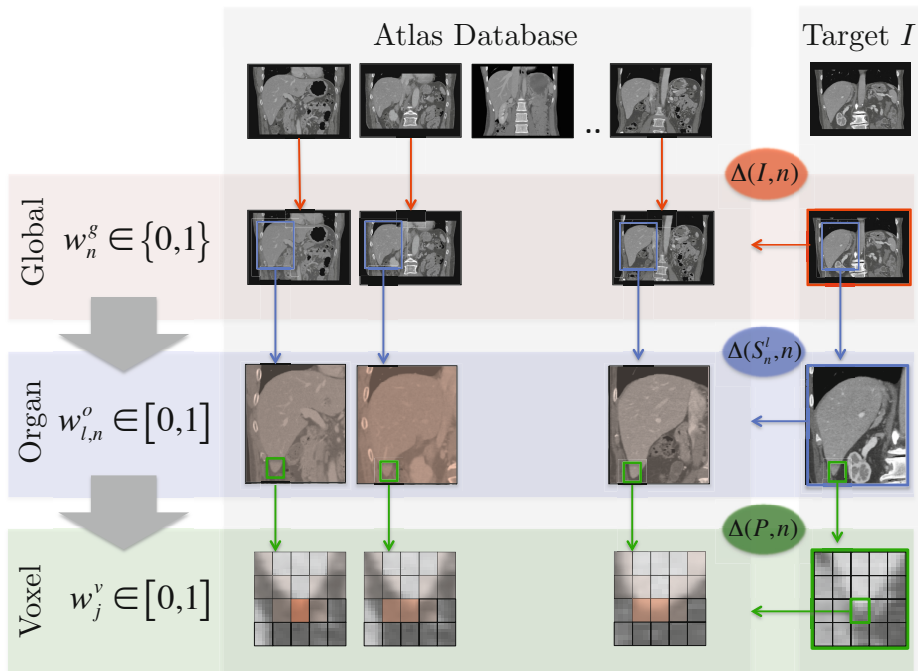


Fig. 1. In the proposed method, atlases are weighted on the global, organ, and voxel level. This scheme specifies the expert knowledge used during the segmentation in a coarse to fine approach, allowing an improved correspondence estimation. On the global level, a binary weighting is applied while a continuous weighting is used in organ- and voxel levels (indicated in red tone).

with weights w_n^g defined on the global atlas level and local weights w_j^o and w_j^v that are assigned on an organ- and voxel-level respectively.

Based on this hierarchical weighting scheme, the target segmentation is inferred from all available atlases. The proposed method is illustrated in Fig. 1: at the **global** stage, a set of atlases is selected based on overall appearance. After pairwise alignment, an individual atlas weighting is defined for every **organ**. Finally, the labelling at the most localized level is inferred after non-linearly aligning atlases on an organ level and by evaluating image-atlas similarities on a voxel-by **voxel** basis.

Weights on all levels are based on the sum of squared intensity differences $\Delta(R, n)$ between image \mathcal{I} and atlas \mathcal{A}_n , defined over a region of interest R measuring image appearance on the relevant level of locality:

$$\Delta(R, n) = \sum_{j \in R} \|\mathcal{I}(x_j) - \mathcal{A}_n(x_j)\|^2 \quad (2)$$

In Sections 2.1, 2.2 and 2.3, the atlas weighting schemes applied at the global-organ- and voxel levels are described respectively.

2.1 Global Atlas Weighting

For a given target image \mathcal{I} , binary weights w_n^g are assigned on a global atlas level. This weight defines a pre-selection of atlases based on global appearance in order to deal with significant differences in body size and field of view. No pairwise registration is performed before computing global atlas weights. By doing so, local minima that may arise from a mis-registration of very different images does not influence the measurement. The ROI R in Eq. 2 spans the whole CT scan and weights are defined as follows:

$$w_n^g = \begin{cases} 1 & \text{if } \Delta(\mathcal{I}, n) < \Delta_\delta \\ 0 & \text{, otherwise.} \end{cases} \quad (3)$$

where Δ_δ is a threshold on the global image distance.

2.2 Organ Level Atlas Weighting

Atlases \mathcal{A}_n that show a high similarity with the target image \mathcal{I} at a global scale are thus weighted with $w_n^g = 1$. These atlases are then aligned with the target image to perform atlas selection on the localized organ, and voxel levels. After pairwise alignment, the atlas label maps S_n^l for label l from atlas n are transformed to the target space. The organ-wide atlas weight $w_{l,n}^o$ is the product of a term $w_{l,n}^{\mathcal{I}}$ that is defined by atlas-target similarities as well as a term $w_{l,n}^S$ that is defined by the agreement between individual atlas label maps:

$$w_{l,n}^o = w_{l,n}^{\mathcal{I}} w_{l,n}^S \quad (4)$$

The similarity-based term $w_{l,n}^{\mathcal{I}}$ is defined after affine alignment of \mathcal{A}_n to \mathcal{I} and measures similarities over the transformed organ label:

$$w_{l,n}^{\mathcal{I}} = \exp\left(-\frac{\Delta(S_n^l, n)}{h}\right) \quad (5)$$

where h is a user-set variable that defines the number of atlases supporting the segmentation of organs with label l . $w_{l,n}^S$ measures the agreement between transformed label maps after applying a non-rigid atlas-target alignment:

$$w_{l,n}^S = \frac{1}{N-1} \sum_{k \in N, k \neq n} \text{JI}(S_k^l, S_n^l) w_k^g \quad (6)$$

where JI is the Jaccard index.

This term is based on a similar principle to the one proposed in the STAPLE algorithm [10] where raters that show good agreement with others are weighted higher. The main difference being that in STAPLE the confidence vote is evaluated on the voxel-level while in our approach confidence is evaluated over the whole label of a given structure, allowing independent intensity-based corrections at the voxel level.

2.3 Local Voxel Weighting

Label weights on a voxel level are assigned based on the similarity of a patch surrounding a given voxel $x_i \in \mathcal{I}$ and patches in a local neighbourhood of all aligned atlas images \mathcal{A}_n . This nonlocal means fusion strategy was recently adapted from image-denoising to the labelling of brain MR scans [9]. A 3-D patch of size $s_p \times s_p \times s_p$ is defined around every voxel $x_i \in \mathcal{I}$. The image similarity between this patch and patches of the same size defined around the voxels in a local neighbourhood of size s_n , is evaluated in all available atlases. This results in weights defined by patches P around x_i and the voxels $x_{n,j}$ in all atlases \mathcal{A}_n and in the defined search window. Following [9], the weights are defined as

$$w_j^v(x_i, x_j) = \exp\left(-\frac{\Delta(P, n)}{h^v}\right) \quad (7)$$

where x_j is a voxel in atlas \mathcal{A}_n as defined in Eq. 1 and h^v defines the weighting in relation to patch distance.

3 Experiments and Results

Our method was evaluated on 100 3D abdominal CT scans acquired from 21 female and 79 male subjects. Subjects were aged between 26 and 83 with a mean age of 63. Scans have a resolution of 512 x 512 voxels in plane and contain between 263 and 538 slices depending on the field of view and the slice thickness. Voxel sizes range between 0.55 and 0.82 mm and the slice spacing varied between 0.4 and 0.8 mm. For each scan a manual segmentation generated by a single trained rater is available for the liver, spleen, pancreas and the kidneys.

3.1 Experiments

A leave-one-out strategy was applied, where one scan in turn was segmented by using the remaining 99 subjects as atlas database. The threshold Δ_δ was defined in a way that 30 atlases were weighted with $w_n^g = 1$ by the global similarity measure in Eq. 3. Correspondence is sought between atlases and target image before extracting organ- and voxel level similarities in Eqs. 4 and 7. A rigid and affine registration step was followed by a multi-level non-rigid registration step using free-form deformations with B-spline control-point spacings of 20 mm, 10 mm and 5 mm [11]. The registration is driven by the normalized mutual information between target and source in the relevant region of interest.

In the patch-based segmentation procedure used to obtain voxel-level weights defined in Eq. 7 a patch-size of $s_p = 5$ and a neighbourhood size of $s_n = 9$ were used and h^v was set to the minimum patch distance as proposed in [9].

The results obtained with the proposed approach are compared to results based on a direct application of atlas-registration with atlas selection as proposed in [6],[7]. For direct fusion, a set of 30 atlases is selected for each organ from the atlas database based on image similarities and non-rigidly aligned with the target. The final label at the voxel level is obtained by performing a majority voting strategy as widely used in brain segmentation [6],[7].

Table 1. Average Jaccard index (JI), Dice overlap (similarity index, SI) and Recall/Precision for 100 subjects. Compared are reference segmentations with the proposed method and a standard multi-atlas segmentation technique.

Structure	Proposed method			Direct fusion		
	Dice	Jaccard	Rec./Prec.	Dice	Jaccard	Rec./Prec.
Liver	94.4	89.5	95.3/93.9	93.0	87.1	93.5/92.9
Spleen	90.9	84.6	92.4/90.7	78.4	68.0	77.0/82.9
Pancreas	65.5	49.6	62.9/70.7	43.6	29.3	33.6/71.9
Kidneys	94.3	88.1	93.4/93.9	88.3	79.6	84.9/92.3

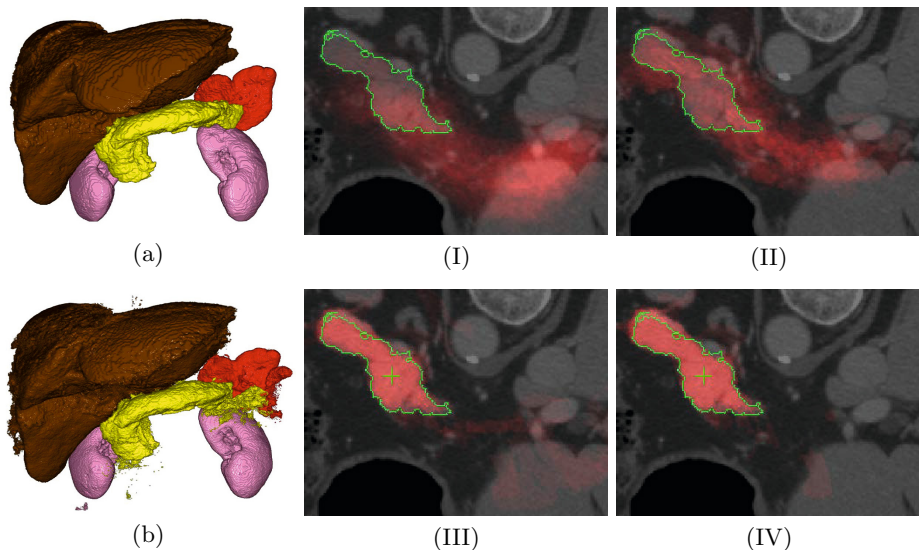


Fig. 2. Panels (a) and (b) show renderings for the liver (brown), kidneys (pink), pancreas (yellow) and spleen (red) based on the manual reference and the proposed method respectively. Panels I-IV illustrate the automated, hierarchical pancreas segmentation on an axial slice with manual labels outlined in green and atlas probabilities in red tone. Panel (I) shows the overall distribution of all atlases in the database. In panel (II), shows the subject-specific probabilistic atlas at the global scale. Panel (III) shows the refined atlas after pairwise registration and organ-wide intensity-based weighting. Panel (IV) also incorporates the proposed confidence-based weighting.

3.2 Results

Automated and manual segmentations were compared by widely used measures that all are defined by the true positive (TP) and false negative (FN) rates and that range from 0 to 100. The Jaccard index (JI), the Dice overlap (Similarity index, SI) as well as recall (REC) and precision (PRE) values between structural labels obtained with the proposed method and direct fusion as well as manual reference labels are presented in Tab. 1.

Fig. 2 shows example for the automated segmentation procedure. Panels (a) and (b) compare volume renderings obtained from automated and manual segmentations. Panels (I) - (IV) present exemplarily the different steps of the proposed hierarchical atlas weighting scheme for pancreas segmentation.

4 Discussion and Conclusion

In this paper we propose a novel, atlas-based segmentation technique for multi-organ abdominal segmentation. The strength of the presented method is its general nature which allows it to be applied robustly to multiple organs without specialisation and individual parameter settings. The method is based on multi-atlas registration, a technique widely used in brain imaging. Here, we present a hierarchical coarse-to-fine atlas weighting strategy that is designed to deal with the challenges found in abdominal segmentation. Especially, a large inter-subject variation in abdominal appearance poses a significant challenge to image registration algorithms and therefore to correspondence estimation, the essence of atlas-based segmentation techniques. We show how the implementation of the proposed hierarchical model leads to segmentation results that compare favourable to the state-of-the-art on the segmentation of a large ($N=100$) and relatively diverse image database.

The results in Tab. 1 show how the proposed hierarchical model clearly outperforms multi-atlas selection and registration schemes as originally proposed for brain segmentation [7]. Especially for highly variable structures like the pancreas, a substantial improvement is achieved. The method performs at least as good as state-of-the-art multi-organ segmentation strategies presented in recent years [3],[5]. Both methods, however, are evaluated on less structures and smaller datasets. The evaluation of the hierarchical model in [3], is restricted to the liver and the work in [5] does not provide results for the pancreas which is the most challenging structure due to a high variability in shape and also position. Shimizu et. al [1] recently describe a method based on shape models that addresses this challenge, achieving a satisfactory accuracy for pancreas segmentation on a database of 20 subjects. While our method gives stable results for pancreas segmentation, such more specialized models can improve the segmentation accuracy. Adding some shape knowledge may further improve the presented results for more challenging organs and is a direction of future work.

The run-time of our method is defined by that of the non-rigid registration step. In the current research implementation, one registration runs for approximately one hour for a whole abdominal scan, resulting in an overall runtime for all organs of around three hours on a machine with eight Intel Xeon cores clocked at 3GHz and 32GB RAM. A recent implementation [12] of the used registration algorithm [11], however, allows speed ups of around 10 fold, making the method relevant for an application in a clinical environment.

Future work needs to be done to evaluate the full potential of the proposed method. As can be seen from the volume rendering in Fig. 2, the lack of topological constraints in our voxel-level weighting may produce sometimes fuzzy and

anatomically implausible or even subdivided segmentations. A way to address this problem is to incorporate smoothness constraints into the final label assignment instead of independently thresholding the weight at the 50% level as described in the current model in Eq. 1. One way to include such constraints is via graph cuts, an optimization technique widely used for labelling problems in medical imaging [13]. Furthermore, to assess inter-observer variability, all reference segmentations will be performed by a second independent, trained rater.

References

1. Shimizu, A., Kimoto, T., Kobatake, H., Nawano, S., Shinozaki, K.: Automated pancreas segmentation from three-dimensional contrast-enhanced computed tomography. *Int. J. CARS* 5, 85–98 (2010)
2. Park, H., Bland, P., Meyer, C.: Construction of an abdominal probabilistic atlas and its application in segmentation. *IEEE TMI* 22(4), 483–492 (2003)
3. Okada, T., Yokota, K., Hori, M., Nakamoto, M., Nakamura, H., Sato, Y.: Construction of Hierarchical Multi-Organ Statistical Atlases and Their Application to Multi-Organ Segmentation from CT Images. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) *MICCAI 2008, Part I. LNCS*, vol. 5241, pp. 502–509. Springer, Heidelberg (2008)
4. Rusko, L., Bekes, G., Fidrich, M.: Automatic segmentation of the liver from multi- and single-phase contrast-enhanced CT images. *MedIA* 13(6), 871–882 (2009)
5. Linguraru, M., Pura, J., Chowdhury, A., Summers, R.: Multi-organ Segmentation from Multi-phase Abdominal CT via 4D Graphs Using Enhancement, Shape and Location Optimization. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) *MICCAI 2010, Part III. LNCS*, vol. 6363, pp. 89–96. Springer, Heidelberg (2010)
6. Heckemann, R.A., Hajnal, J.V., Aljabar, P., Rueckert, D., Hammers, A.: Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage* 33(1), 115–126 (2006)
7. Artaechevarria, X., Munoz-Barrutia, A., Ortiz-de Solorzano, C.: Combination Strategies in Multi-Atlas Image Segmentation: Application to Brain MR Data. *IEEE TMI* 28(8), 1266–1277 (2009)
8. Isgum, I., Staring, M., Rutten, A., Prokop, M., Viergever, M., van Ginneken, B.: Multi-Atlas-Based Segmentation With Local Decision Fusion - Application to Cardiac and Aortic Segmentation in CT Scans. *IEEE TMI* 28(7), 1000–1010 (2009)
9. Coupe, P., Manjón, J.V., Fonov, V., Pruessner, J., Robles, M., Collins, D.L.: Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *NeuroImage* 54(2), 940–954 (2011)
10. Warfield, S.K., Zou, K.H., Wells III, W.M.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE TMI* 23(7), 903–921 (2004)
11. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L.G., Leach, M.O., Hawkes, D.J.: Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE TMI* 18(8), 712–721 (1999)
12. Modat, M., Ridgway, G.R., Taylor, Z.A., Lehmann, M., Barnes, J., Hawkes, D.J., Fox, N.C., Ourselin, S.: Fast free-form deformation using graphics processing units. *Computer Methods and Programs in Biomedicine* 98(3), 278–284 (2010)
13. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. *IEEE PAMI* 23(11), 1222–1239 (2001)