

# Automatic 3D Multiorgan Segmentation via Clustering and Graph Cut Using Spatial Relations and Hierarchically-Registered Atlases

Razmig Kéchichian<sup>1</sup>(✉), Sébastien Valette<sup>2</sup>, Michaël Sdika<sup>2</sup>,  
and Michel Desvignes<sup>1</sup>

<sup>1</sup> Gipsa-lab: CNRS UMR 5216, Université de Grenoble,  
Saint-Martin-d'Hères, France

razmig.kechichian@gipsa-lab.grenoble-inp.fr

<sup>2</sup> Creatis: CNRS UMR 5220, Inserm U1044, Université de Lyon,  
Villeurbanne, France

**Abstract.** We propose a generic method for automatic multiple-organ segmentation based on a multilabel Graph Cut optimization approach which uses location likelihood of organs and prior information of spatial relationships between them. The latter is derived from shortest-path constraints defined on the adjacency graph of structures and the former is defined by probabilistic atlases learned from a training dataset. Organ atlases are mapped to the image by a fast (2+1)D hierarchical registration method based on SURF keypoints. Registered atlases are furthermore used to derive organ intensity likelihoods. Prior and likelihood models are then introduced in a joint centroidal Voronoi image clustering and Graph Cut multiobject segmentation framework. Qualitative and quantitative evaluation has been performed on contrast-enhanced CT images from the Visceral Benchmark dataset.

## 1 Introduction and Related Work

Clinical practice today, especially whole-body CT and MRI scanning, often generates large numbers of high-resolution images, which makes tasks of efficient data access, transfer, analysis and visualization challenging, especially in distributed computing environments which have seen growing use of handheld terminals for interactive data access and visualization of anatomy. Therefore, there is great interest for efficient and robust medical image segmentation algorithms for the purposes of creating patient-specific anatomical models, clinical applications, medical research and education, and visualization of full-body anatomy [2].

Traditionally single-object or pathology oriented, recent image processing methods [7, 9–11, 14, 16, 17] have made the analysis and segmentation of multiple anatomical structures increasingly possible. However, CT and MR images have intrinsic characteristics that render its automatic segmentation challenging. They are commonly degraded by various noise sources and artifacts due to limited acquisition time and resolution, and patient motion which all reduce the

prominence of intensity edges in images. Regardless of the imaging modality and related artifacts, many anatomically and functionally distinct structures, especially those corresponding to soft tissues, have similar intensity levels in images and, furthermore, blend into surrounding tissues which have intensities close to their own. It is impossible to identify and segment such structures automatically on the basis of intensity information only. Hence, most advanced segmentation methods exploit some form of prior information on structure location [14, 18] or interrelations [7, 10, 16, 17] to achieve greater robustness and precision.

Graph Cut methods, which have been widely applied to single-object segmentation problems [3], rely on a maximum-flow binary optimization scheme of a discrete cost function on the image graph. For a particular class of cost functions which frequently arises in segmentation applications [12], these methods produce provably-good approximate solutions in multiobject [4] and global optima in single-object segmentation. In addition, simultaneous multiobject segmentation approaches are superior to their sequential counterparts in that they raise questions neither on the best segmentation sequence to follow nor on how to avoid the propagation of errors on individual segmentations [7].

We propose a generic method for automatic multiple-organ segmentation based on multilabel Graph Cut optimization which uses location and intensity likelihoods of organs and prior information of their spatial configuration. The spatial prior is derived from shortest-path pairwise constraints defined on the adjacency graph of structures [10], and the organ location likelihood is defined by probabilistic atlases learned from the Visceral Benchmark training dataset [8]. We register organ atlases to the image prior to segmentation using a fast (2+1)D registration method based on SURF keypoints [1]. Registered atlases are also used to derive organ intensity likelihoods. Prior and likelihood models are then introduced in a joint centroidal Voronoi image clustering and Graph Cut multi-object segmentation framework. We present the results of qualitative and quantitative evaluation of our method on contrast-enhanced CT images from the Visceral Benchmark dataset.

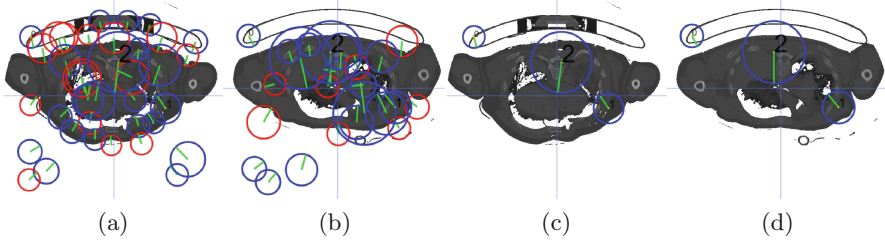
## 2 Methods

### 2.1 SURF Keypoint-Based Image Registration

We first outline our fast (2+1)D rigid registration method, which is based on keypoints detection. Features are extracted in 2D volume slices. This has the advantage of being fast and easily parallelizable. Another advantage is that medical data is usually stored in a Picture Archiving and Communication System (PACS) as volume slices instead of full volumes. Our method easily fits into such medical environments. Note that while feature extraction is done in 2D images, registration is still performed in 3D, hence the (2+1)D notation.

We currently use the SURF image descriptor [1], but our approach is generic and would work with others. To reduce computation time, we first downsample the input volume to a user-specified dimension. As a rule of thumb, we isotropically resample each volume so that its second longest dimension is equal to the

desired resolution  $R$ . For example, when  $R = 80$ , the POPI CT volume [19] of dimensions  $482 \times 360 \times 141$  and spacing  $0.97 \text{ mm} \times 0.97 \text{ mm} \times 2 \text{ mm}$  is resampled to a  $107 \times 80 \times 64$  volume with an isotropic spacing of  $4.39 \text{ mm}$ .



**Fig. 1.** Matching two slices of POPI (a) and P1 (b) volumes. (a) and (b) show all features found in both slices (a feature is represented by a circle and its orientation by its radius). (c) and (d) show the three matching features between the two slices.

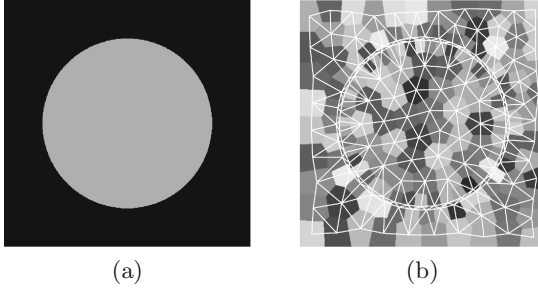
Next, we extract 2D SURF features from each slice. As these operations are completely independent, this step is carried out in a parallel manner. Figure 1 shows feature extraction on POPI and P1 CT images [15]. The number of extracted features is 1154 and 1079, respectively. Figures 1c and d show the three matching couples found in both images. Once 2D matches are found, we are able to perform volume registration. For robustness purposes, we use a simple scale + translation transformation model:

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = s \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}. \quad (1)$$

We estimate the 4 parameters  $s$ ,  $t_x$ ,  $t_y$  and  $t_z$  in similar spirit to the RANSAC method [6], which is an iterative parametric model estimation method known to be very efficient in the presence of outliers. One RANSAC iteration usually consists in randomly picking a small number of samples to estimate the model parameters, and then counting the number of data samples consistent with the model, rejecting outliers. After performing all iterations, the model providing the highest number of consistent data samples is kept as the solution.

## 2.2 Organ Atlas Construction

Using 20 contrast-enhanced CT images and ground-truth annotations thereof from the Visceral Benchmark dataset [8], we construct a probabilistic atlas for each of the following 20 structures: thyroid; trachea; sternum; liver; spleen; pancreas; gallbladder; first lumbar vertebra; aorta; urinary bladder; right and left lungs, kidneys, adrenal glands, psoas major and rectus abdominis muscle bodies. In addition, we create atlases for three additional image and body regions: background (BKG), thorax and abdomen (THAB) and a body envelope (ENV) from



**Fig. 2.** An image-adaptive CVT clustering and its dual graph for a circle image.

annotations generated automatically as follows. BKG is created by thresholding the CT image followed by morphological processing in order to isolate the body region. THAB is created as the dilated union of the aforementioned 20 structures and their bounding 3D ellipse, from which the structures are subtracted after dilation. Finally, ENV is defined as the image minus BKG and THAB. Note that ENV is a crude body envelope that comprises skin, fat, muscle and bone structures. Refer to Fig. 4c for an illustration.

To create atlases, we use a representative image from the dataset as a reference and register remaining images to it via the method described in Sect. 2.1. We register each structure separately in a bounding cube of a given margin in the intensity image, defined according to the corresponding annotation image, and apply the obtained transform subsequently to the annotation image. We accumulate annotations thus registered in a 3D histogram of reference image dimensions which is normalized to produce the corresponding probability map.

### 2.3 Image Clustering

The full-resolution voxel representation is often redundant because objects usually comprise many similar pixels that could be grouped. Therefore, we simplify the image prior to segmentation by an image-adaptive centroidal Voronoi tessellation (CVT) which strikes a good balance between cluster compactness and object boundary adherence, and helps to place subsequent segmentation boundaries precisely. We have shown that the clustering step improves the overall runtime and memory footprint of the segmentation process up to an order of magnitude without compromising the quality of the result [10].

Let us define a grayscale image as a set of voxels  $\mathcal{I} = \{v \mid v = (x, y, z)\}$  and associate to each voxel  $v \in \mathcal{I}$  a gray-level  $I_v$  from some range  $I \subset \mathbb{R}$ . Given a grayscale image  $\mathcal{I}$  and  $n$  sites  $c_i \in \mathcal{I}$ , a CVT partitions  $\mathcal{I}$  into  $n$  disjoint clusters  $C_i$  associated with each centroid  $c_i$  and minimizes the following energy:

$$F(v; c_i) = \sum_{i=1}^n \left( \sum_{v \in C_i} \rho(v) (\|v - c_i\|^2 + \alpha \|I_v - I_i\|^2) \right) . \quad (2)$$

In (2),  $\rho(v)$  is a density function defined according to the intensity-gradient magnitude at voxel  $v$ ,  $\rho(v) = |\nabla I_v|$ ,  $\alpha$  is a positive scalar and  $I_i$  is the gray-level of the cluster  $C_i$  defined as the mean intensity of its voxels. Refer to Fig. 2 for an illustration in 2D. To minimize (2), we apply a variant of the clustering algorithm in [5] which approximates a CVT in a computationally-efficient manner, involving only local queries on voxels located on boundaries of pairs of clusters.

For referral in later sections, we shall define the graph of a CVT, illustrated in Fig. 2b. Denote the surface of a cluster  $C_i$  by  $\partial C_i$ . Given a CVT clustering  $\mathcal{C}$ , let the set  $\mathcal{S}$  index its clusters, and let  $\mathcal{G} = \langle \mathcal{S}, \mathcal{E} \rangle$  be an undirected graph on cluster centroids where pairs of clusters having nonzero-area common surface define the set of edges  $\mathcal{E} = \{\{i, j\} \mid i, j \in \mathcal{S}, |\partial C_i \cap \partial C_j| \neq 0\}$ . Consequently, the neighborhood of a node  $i \in \mathcal{S}$  is defined as  $\mathcal{N}_i = \{j \mid j \in \mathcal{S}, \exists \{i, j\} \in \mathcal{E}\}$ .

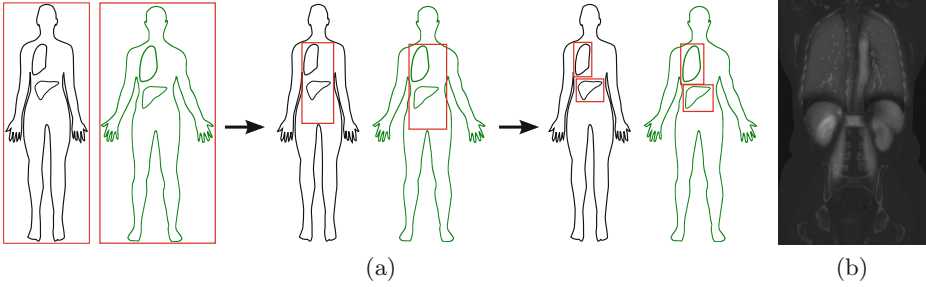
## 2.4 Multi-Organ Image Segmentation

We formulate image segmentation as a labeling problem, defined as the assignment of a label from a set of 23 labels  $L$  representing the structures to be segmented to each of the variables in a set of  $n$  variables, indexed by  $\mathcal{S}$ , corresponding to the clusters of a CVT-clustered image. Assume that each variable  $i \in \mathcal{S}$  is associated with the corresponding node in the graph  $\mathcal{G}$  of the CVT defined in Sect. 2.3. An assignment of labels to all variables is called a configuration, and is denoted by  $\ell \in \mathcal{L}$ . An assignment of a label to a single variable is denoted by  $\ell_i$ . We cast the labeling problem in a maximum a posteriori estimation framework and solve it by minimizing the following energy function of label configurations via the Expansion Moves multilabel Graph Cut algorithm [4]:

$$E(\ell) = t_1 \sum_{i \in \mathcal{S}} D_i(\ell_i) + t_2 \sum_{i \in \mathcal{S}} P_i(\ell_i) + \frac{1}{2} \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{N}_i} V_{i,j}(\ell_i, \ell_j) . \quad (3)$$

In (3),  $t_1$  and  $t_2$  are temperature hyperparameters,  $\mathcal{N}_i$  is the neighborhood of the variable  $i \in \mathcal{S}$ . The first and second sums in (3) correspond respectively to organ intensity and location (atlas) likelihood energies, and the third is the energy of a prior distribution of label configurations expressed as a Markov random field w.r.t. the graph  $\mathcal{G}$ . We shall define these terms in detail.

**Spatial Configuration Prior.** Pairwise terms of (3) encode prior information on interactions between labels assigned to pairs of neighboring variables encouraging the spatial consistency of labeling with respect to a reference model. We define these terms according to our piecewise-constant vicinity prior model proposed in [10], which, unlike the standard Potts model, incurs multiple levels of penalization capturing the spatial configuration of structures in multiobject segmentation. It is defined as follows. Let  $\mathcal{R}$  be the set of symmetric adjacency relations on pairs of distinct labels,  $\mathcal{R} = \{r \mid arb, a, b \in L, a \neq b\}$ .  $\mathcal{R}$  can be represented by a weighted undirected graph on  $L$ ,  $\mathcal{A} = \langle L, W \rangle$ , with the set of edges  $W = \{\{a, b\} \mid \exists r \in \mathcal{R}, arb, a \neq b\}$  where edge weights are defined by  $w(\{a, b\}) = 1$ , such that  $w(\{a, b\}) = \infty$  if  $\nexists r \in \mathcal{R}, arb$ .



**Fig. 3.** (a) An illustration of our hierarchical registration procedure, and (b) an example of registered organ atlases overlaid on the image.

Given the graph  $\mathcal{A}$ , we define the pairwise term in (3) as:

$$V_{i,j}(\ell_i, \ell_j) = |\partial C_i \cap \partial C_j| \varpi(a, b), \quad \ell_i = a, \ell_j = b . \quad (4)$$

where  $\varpi(a, b)$  is the shortest-path weight from  $a$  to  $b$  in  $\mathcal{A}$ . The area of the common surface of adjacent clusters  $|\partial C_i \cap \partial C_j|$  is introduced so that the sum of pairwise energies in (3)  $\forall a, b \in L$  is equal to the area of the common surface between the corresponding pair of objects multiplied by the shortest-path weight.

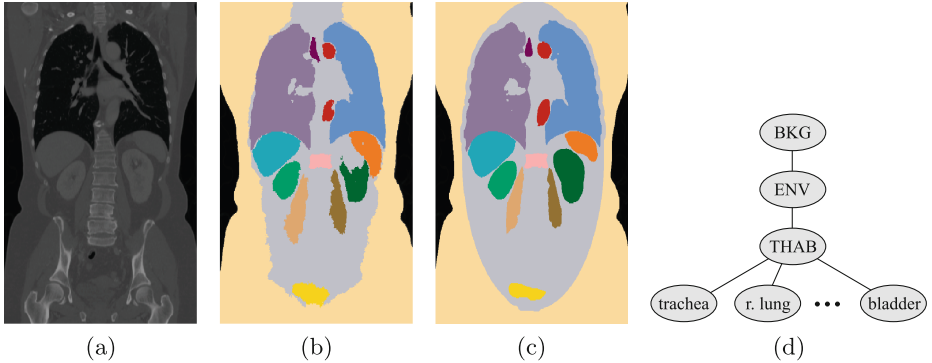
**Intensity and Location Likelihoods.** Unary terms of (3) measure the cost of assigning labels to variables. They are defined as negative log-likelihood functions derived from organ observed intensity and location probabilities:

$$D_i(\ell_i) = -\ln \prod_{v \in C_i} \Pr(I_v | \ell_i) , \quad (5a)$$

$$P_i(\ell_i) = -\ln \prod_{v \in C_i} \Pr(X_v | \ell_i) . \quad (5b)$$

In (5b),  $X_v$  denotes the object-space coordinates of the voxel  $v$ . Conditional probabilities in (5a) and (5b) correspond respectively to those of voxel intensity and location given the structure  $\ell_i$ . To estimate the conditional probability distribution  $\Pr(I | l)$  for a given label  $l \in L$ , we first register the corresponding organ atlas to the image, then estimate the conditional probability as a Gauss-smoothed and normalized intensity histogram derived from voxels in high-probability regions of the registered atlas. Conditional probability distributions  $\Pr(X | L)$  are defined directly from registered atlases. The next section details our hierarchical registration approach which maps organ atlases to the image.

**Hierarchical Registration of Organ Atlases.** We register probabilistic organ atlases, constructed as described in Sect. 2.2, to the image in a 3-step hierarchical fashion starting at the full image, then on an intermediate level



**Fig. 4.** A segmentation of 12 structures from the Visceral Benchmark dataset image 10000109\_1\_CTce\_ThAb. The coronal sections correspond to (a) the image, (b) its segmentation and (c) the associated ground-truth with additional labels for BKG, ENV and THAB. (d) illustrates the adjacency graph used to define the spatial prior model.

**Table 1.** Mean Dice figures for select organs calculated on 10 segmentations per organ.

Structures	Dice	Structures	Dice	Structures	Dice
Lung (R)	0.95	Kidney (R)	0.81	Urinary bladder	0.77
Lung (L)	0.96	Kidney (L)	0.86	Trachea	0.62
Liver	0.93	Psoas major (R)	0.71	Sternum	0.63
Spleen	0.84	Psoas major (L)	0.79	Aorta	0.57

corresponding to the THAB region, and finally on individual organs. After performing registration on each level, we apply the obtained transform to the corresponding atlas as well as to those of organs contained in the registered region. As in Sect. 2.2, we register each structure separately in a bounding cube of a given margin in the intensity image, defined according to the corresponding atlas. Figure 3 illustrates the hierarchical registration procedure and gives an example of registered organ atlases overlaid on the image to which they have been registered.

### 3 Results and Conclusions

We have carried out qualitative evaluation on several contrast-enhanced CT images from the Visceral Benchmark training dataset [8]. An example is given in Fig. 4. The number of CVT clusters is equal to 3% of image voxel count. Temperature parameters  $t_1$  and  $t_2$  are set such that intensity and location likelihood-based unary terms contribute equally to (3). The spatial prior is defined according to the adjacency graph given in Fig. 4d. Table 1 presents the results of quantitative evaluation of our method on contrast-enhanced CT images during the Visceral Anatomy 2 Benchmark. We report results corresponding to

the best setting of temperature parameters out of 5 tested settings. In addition, we give mean Dice figures only for organs for which our method successfully produced segmentations on all 10 test images.

We conclude on two remarks. (1) Even though our hierarchical approach to mapping atlases to the image relies on a rigid registration method, unlike many hierarchical methods which use non-rigid deformable registration [13], it helps localizing segmented structure boundaries quite well, because location information roughly registered atlases contribute is complemented by intensity similarity and spatial consistency criteria. (2) Full-body modeling by the introduction of annotations BKG, ENV and THAB, not only complements location information and allows for hierarchical registration, but also increases the discriminative power of the spatial prior by higher penalization of inconsistent configurations.

## References

1. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)
2. Blume, A., Chun, W., Kogan, D., Kokkevis, V., Weber, N., Petterson, R.W., Zeiger, R.: Google body: 3D human anatomy in the browser. In: ACM SIGGRAPH 2011 Talks, pp. 19:1–19:1 (2011)
3. Boykov, Y., Funka-Lea, G.: Graph cuts and efficient N-D image segmentation. *Int. J. Comput. Vis.* **70**(2), 109–131 (2006)
4. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(11), 1222–1239 (2001)
5. Dardenne, J., Valette, S., Siauve, N., Burais, N., Prost, R.: Variational tetrahedral mesh generation from discrete volume data. *Vis. Comput.* **25**(5), 401–410 (2009)
6. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **24**(6), 381–395 (1981)
7. Fouquier, G., Atif, J., Bloch, I.: Sequential model-based segmentation and recognition of image structures driven by visual features and spatial relations. *Comput. Vis. Image Underst.* **116**(1), 146–165 (2012)
8. Hanbury, A., Müller, H., Langs, G., Weber, M.A., Menze, B.H., Fernandez, T.S.: Bringing the algorithms to the data: cloud-based benchmarking for medical image analysis. In: Catarci, T., Forner, P., Hiemstra, D., Peñas, A., Santucci, G. (eds.) CLEF 2012. LNCS, vol. 7488, pp. 24–29. Springer, Heidelberg (2012)
9. Iglesias, J.E., Konukoglu, E., Montillo, A., Tu, Z., Criminisi, A.: Combining generative and discriminative models for semantic segmentation of CT scans via active learning. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 25–36. Springer, Heidelberg (2011)
10. Kéchichian, R., Valette, S., Desvignes, M., Prost, R.: Shortest-path constraints for 3d multi-object semi-automatic segmentation via clustering and graph cut. *IEEE Trans. Image Process.* **22**(11), 4224–4236 (2013)
11. Kohlberger, T., Sofka, M., Zhang, J., Birkbeck, N., Wetzl, J., Kaftan, J., Declerck, J., Zhou, S.K.: Automatic multi-organ segmentation using learning-based segmentation and level set optimization. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 338–345. Springer, Heidelberg (2011)



12. Kolmogorov, V., Zabih, R.: What energy functions can be minimized via graph cuts? *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(2), 147–159 (2004)
13. Lester, H., Arridge, S.R.: A survey of hierarchical non-linear medical image registration. *Pattern Recogn.* **32**(1), 129–149 (1999)
14. Linguraru, M.G., Pura, J.A., Pamulapati, V., Summers, R.M.: Statistical 4D graphs for multi-organ abdominal segmentation from multiphase CT. *Med. Image Anal.* **16**(4), 904–914 (2012)
15. Lynch, R., Pitson, G., Ball, D., Claude, L., Sarrut, D.: Computed tomographic atlas for the new international lymph node map for lung cancer: a radiation oncologist perspective. *Pract. Radiat. Oncol.* **3**(1), 54–66 (2013)
16. Okada, T., Yoshida, Y., Hori, M., Summers, R.M., Chen, Y.W., Tomiyama, N., Sato, Y.: Abdominal multi-organ segmentation of CT images based on hierarchical spatial modeling of organ interrelations. In: *Proceedings of the 3rd International Conference on Abdominal Imaging: Computational and Clinical Applications*, pp. 173–180 (2012)
17. Seifert, S., Barbu, A., Zhou, S.K., Liu, D., Feulner, J., Huber, M., Suehling, M., Cavallaro, A., Comaniciu, D.: Hierarchical parsing and semantic navigation of full body CT data. In: *SPIE Medical Imaging, Lake Buena Vista, FL, USA, Feb 2009*
18. Song, Z., Tustison, N., Avants, B., Gee, J.: Adaptive graph cuts with tissue priors for brain MRI segmentation. In: *IEEE ISBI*, pp. 762–765, Apr 2006
19. Vandemeulebroucke, J., Sarrut, D., Clarysse, P.: Point-validated pixel-based breathing thorax model. In: *International Conference on the Use of Computers in Radiation Therapy (ICCR)*, pp. 195–199. Toronto, Canada, Jun 2007