

# 3D Segmentation for Multi-Organs in CT Images

Mariusz Bajger<sup>1,2</sup>, Gobert Lee<sup>1,2</sup> and Martin Caon<sup>1,3</sup>

<sup>1</sup> *Medical Devices Research Institute, Flinders University, Australia*

<sup>2</sup> *School of Computer Science, Engineering and Mathematics, Flinders University, Australia*

<sup>3</sup> *School of Nursing and Midwifery, Flinders University, Australia*

Received 22nd Oct 2012; accepted 15th Dec 2012

---

## Abstract

The study addresses the challenging problem of automatic segmentation of the human anatomy needed for radiation dose calculations. Three-dimensional extensions of two well-known state-of-the-art segmentation techniques are proposed and tested for usefulness on a set of clinical CT images. The new techniques are 3D Statistical Region Merging (3D-SRM) and 3D Efficient Graph-based Segmentation (3D-EGS). Segmentations of eight representative tissues (lungs, stomach, liver, heart, kidneys, spleen, bones and the spinal cord) were tested for accuracy using the Dice index, the Hausdorff distance and the  $H_t$  index. The 3D-SRM outperformed 3D-EGS producing the average (across the 8 tissues) Dice index, the Hausdorff distance, and the  $H_2$  of 0.89, 12.5 mm and 0.93, respectively.

*Key Words:* Voxel model, image segmentation, statistical region merging, efficient graph-based segmentation, full-body CT

---

## 1 Introduction

The radiation protection guidelines aim to keep dose "as low as reasonably achievable" (the ALARA principle). To evaluate the detriment it is necessary to calculate the effective dose, and this requires knowledge of the amount of energy deposited in each of the specific organs and tissues (the organ doses). For this purpose it is critical to be able to build accurate models of human anatomy. The development of those models faces two major problems: difficulties in obtaining image data sets that span the entire anatomy from head to toe, and speed and accuracy in segmenting large number of images (300 – 400) in such data sets.

---

Correspondence to: <Mariusz.Bajger@flinders.edu.au>

Recommended for acceptance by <Georgios Triantafyllidis>

ELCVIA ISSN:1577-5097

Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

So far, several models of anatomy suitable for dosimetry exist such as ADELAIDE [1], a torso, BABY and CHILD [2] and the University of Florida series of phantoms [3], [4]. The latter series of models have the limbs, heads and torsos from different individuals scaled and assembled into composite models. Additional anatomical models that span various body sizes and shapes of humans of all ages are required for dosimetry purposes. Existing anatomy models were constructed using semi-automatic or supervised segmentation methods. However, the process of segmenting all of the tissues in hundreds of images is extremely time-consuming.

In the 2007 recommendations of the International Commission on Radiological Protection [5], all organs/tissues in the body were classified into three groups in relation to radiation dose calculation. In this study, we focused on eight organs/tissues which were particularly selected in this study as they cover all three groups. Our objective is to evaluate the full potential, towards both accuracy and robustness of the results, of the automatic image analysis system, utilizing either the 3D-SRM or 3D-EGS method, introduced in our recent seminal work [6]. As far as we know, it is the first attempt to use region merging segmentation techniques in the 3D setting of full-body CT segmentation.

Two major research directions on segmentation of CT images are full-body CT segmentation, where contours of several organs are of interest (e.g. [7], [8], [9]) and CT segmentation focusing on selected organs, needed for an early diagnosis or an investigation of a pathology (e.g. [10], [11], [12] (liver, kidneys and spleen), [13] (lungs), [14] (esophagus), [15] (liver)).

Automatic segmentation of all of the tissues needed for dose calculation as recommended by [5] was not reported so far, although some results on the segmentation of multiple organs/tissues (in full-body or targeted region CT) were reported.

Linguraru et al. [10] segmented liver, spleen and kidneys using 4D extension of the well-known graph-cuts technique with shape priors and a probabilistic atlas. The study used a contrast-enhanced two phase abdominal CT (as opposed to non-contrast CT). In another study, Okada et al. [11], multi-organ hierarchically organized atlases were used to improve segmentation of the liver and some peripheral organs. Iglesias et al. in [7] segmented six organs (heart, liver, kidneys, lungs, spleen and pelvis) utilizing random forest classification. Their probabilistic classifier required an extensive training and tuning of several parameters.

Montillo et al. in [8] segmented eight tissues (heart, liver, aorta, lungs, femurs, pelvis, kidneys). The method achieved very high voxel accuracy but required a significant number of training cases to teach the classifier. In Seifert et al. [9] a database-guided segmentation is performed on five organs: heart, liver, kidneys, spleen, bladder and prostate using organ specific landmarks.

It is worth observing that the mentioned studies reported notably good results for the selected organs but also indicated high complexity of the steps with significant amount of data needed for training classifiers, significant time required for training, and necessity of setting several parameters experimentally.

As pointed out in Bajger et al. [6] the Statistical Region Merging (SRM) technique, introduced in [16], and the Efficient Graph-based Segmentation (EGS) method, proposed in [17] are time efficient and robust segmentation methods with a potential to succeed in CT segmentation.

Successful medical image segmentations using these methods were reported in several studies (e.g. [18], [19], [20], [21]), [22]), despite the fact that the methods assume some homogeneity property for components, which often fails in medical images.

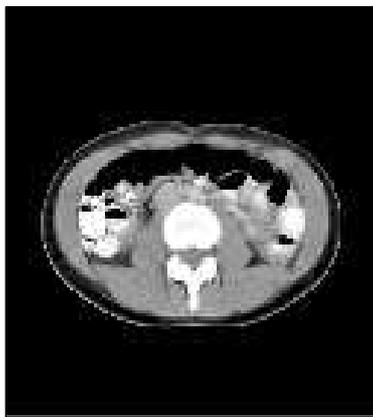
To compensate for lack of the homogeneity property, pre and post-processing techniques are often used. In Ma et al. [18] an active contour model is used to refine the final contour, in Bajger et al. [19] the level-set technique is applied and in Celebi et al. [20] morphological dilation and a majority filtering method are utilized upon the segmented regions.

A statistical atlas or a database-driven classification can then be used to accomplish the tissue classification process, as proved in the above mentioned studies on CT segmentations.

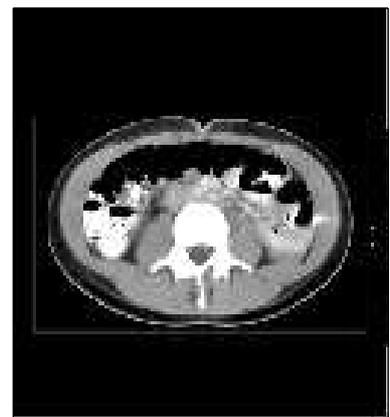
The 3D approach utilized in this study allows for a simultaneous tracking of boundaries of objects spanning across several slices. Figure 1 shows an example where 2D (slice) segmentation cannot determine the organ (spinal cord) border due to strong similarity to the surrounding tissue. In the neighbouring CT slices the spinal cord is entirely enclosed within a bone tissue and, hence, easy to segment. This information is readily incorporated into our 3D technique and the spinal cord is perfectly segmented (Figure 1 (f)). In the dataset used in this study, 41 slices included the spinal cord and four of them suffered from the above mentioned 'tissue leaking' problem when 2D segmentation was performed. The 3D techniques were able to recover the spinal cord shape successfully in all these cases.



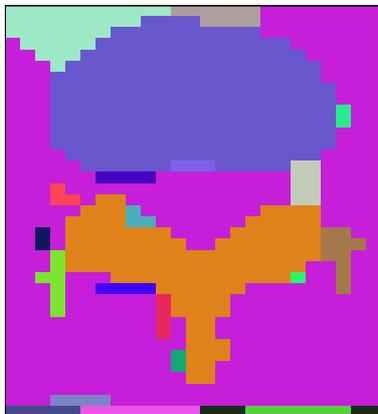
(a) Slice 67



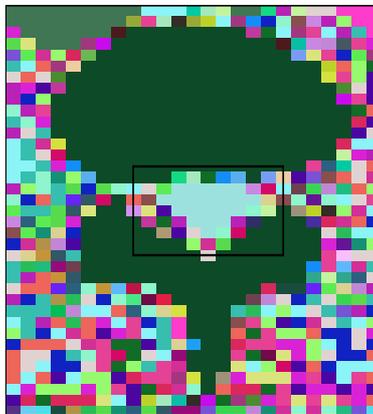
(b) Slice 68 (tissue leaking)



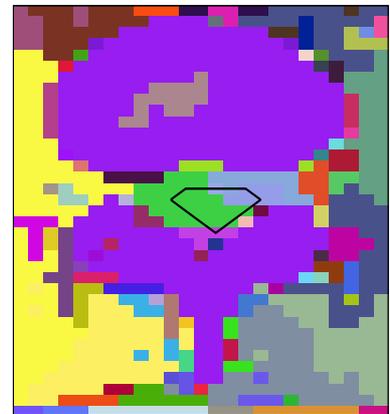
(c) Slice 69



(d) 2D segmentation of slice 68



(e) Expert segmented slice 68



(f) Our 3D segmentation of slice 68

Figure 1: Spinal cord segmentation in slice 68 using standard (2D) SRM method and our new 3D version of the method. The figure shows (a) the original CT slice 67 (resized to  $256 \times 256$ ) with the spinal cord fully confined by the bone (b) the original CT slice 68 with the spinal cord not entirely surrounded by bone tissue (c) the original CT slice 69, where the spinal cord is again well within the bone tissue (d) 2D segmentation of slice 68, which fails to segment the spinal cord from the soft tissue surrounding the vertebra (e) the expert segmented spinal cord in slice 68 (within the rectangle) (f) the slice 68 segmented using our extended 3D-SRM technique with the ground truth contour imposed; the spinal cord overlaps well with the expert contour (Dice index = 0.69)

## 2 Theoretical Background

### 2.1 Statistical Region Merging Segmentation Technique

The Statistical Region Merging technique (SRM) introduced by Nielsen and Nock ([16], [23]) considers the given image  $I$  as an observation of a true (statistical) scene  $I^*$ . The statistical region  $I^*$  is unknown and is a subject of recovery from the given image  $I$ . It is assumed that pixels in  $I^*$  are represented by a set of  $Q$  distributions from which color levels in  $I$  are sampled. Specifically, the intensity of each pixel in  $I$  is realized as a sum of  $Q$  independent random variables. Thus, in case of grey level images, these variables take values in the interval  $[0, 256/Q]$ .

The sets of  $Q$  distributions can differ between statistical pixels but it is assumed that the statistical regions in  $I^*$  are 4-connected components and that the expected value of intensity is the same for components belonging to the same true region, and it differs between statistically different adjacent regions.

The parameter  $Q$  can be considered as a measure of statistical complexity of  $I^*$ . Smaller  $Q$  values result in more general models which are hard to segment. Visually, small  $Q$  often gives undersegmented scene, while high  $Q$  values may result in oversegmentation. The general theory outlined in [16] shows that the probability of undersegmentation is high, thus it is critical to select  $Q$  in such a way that objects of interest are well-segmented.

In some specific applications like masses segmentation in mammograms ([19]) or segmentation of individual CT slice images ([21]), it is possible to develop an analytical criterion helping in optimizing the  $Q$  value. That is, one can estimate the smallest value of  $Q$  sufficient for successful segmentation of regions of interests e.g. specific tissues or masses.

SRM algorithm starts with sorting, in an increasing order, of pairs of pixels  $p, p'$  of  $I$  according to a function  $f(p, p')$ . In this study  $f(p, p')$  is the difference in intensity values in  $p$  and  $p'$ . Initially single pixels, then regions  $R, R'$ , are merged if the following predicate holds true.

$$P(R, R') = \begin{cases} true, & \text{if } |\bar{R} - \bar{R}'| \leq \sqrt{b^2(R) + b^2(R')}, \\ false, & \text{otherwise} \end{cases} \quad (1)$$

where

$$b(R) = g \sqrt{\frac{1}{2Q|R|} \ln \frac{2}{\delta}}. \quad (2)$$

$|R|$  denotes the number of pixels in the region  $R$ ,  $0 < \delta \leq 1$ ,  $\bar{R}$  stands for the average intensity across the region  $R$  and  $g$  is the number of image intensity levels.

The predicate (1) is based on the assumption that the two regions  $R$  and  $R'$  should merge if they come from the same true (statistical) region, that is, if  $\mathbb{E}(\bar{R} - \bar{R}') = 0$ , where  $\mathbb{E}(R)$  is the expectation over all corresponding statistical pixels of  $I^*$  of their sum of expectations of their  $Q$  random variables for their intensity values. Using some probabilistic concentration theory results it was proved in [16] that with probability  $p \geq 1 - O(|I|\delta)$ , the resulting segmentation of  $I$  is an undersegmentation of  $I^*$ . On the other hand,  $\delta$  must be kept small to facilitate merging. In this study we follow [16] and set  $\delta = \frac{1}{6|I|^2}$ , thus we may expect an undersegmentation effect with a high probability.

## 2.2 Efficient Graph-based Segmentation Technique

There are several graph-based segmentation methods in the literature based on the minimum spanning tree of the graph starting from the early work of Zahn [24] through the works of Urquhart [25] and Felzenszwalb et al. [26]. Felzenszwalb et al. [17] further developed the minimum spanning tree based approach into the well acknowledged Efficient Graph-based Segmentation (EGS) technique. The EGS technique was proved to be computationally efficient. Applications and extensions of the method can be found in the literature including the most recent ones by Huang et al. ([27]) to ultrasound imaging and Bajger et al. ([6]) to CT segmentation. For the reader's convenience we briefly outline the algorithm for the efficient graph-based segmentation based on the Kruskal's method of finding minimum spanning trees in graphs ([26], [17]).

The process begins with an undirected weighted graph  $G = (V, E)$  such that  $V$  (the set of vertices) is the set of pixels in the image and  $E$  is the set of edges that connect pixels to immediate neighbors (4-connectivity is commonly used). The weight  $w$  is defined for each edge as:  $w((v_i, v_j)) = |I(v_i) - I(v_j)|$ , for  $(v_i, v_j) \in E$ , where  $I(v_i)$  is the image intensity value at  $v_i$  (the value of the pixel gray scale). The algorithm creates a new graph in such a way that the connected components of the new graph correspond to the desired segmentation of the image. Initially each vertex (image pixel) constitutes a single component. Then edges from  $E$  are traversed (in ascending order based on weights) and considered for inclusion in the new graph according to the following merging predicate.

Two components  $C_1$  and  $C_2$  are merged if

$$d(C_1, C_2) \leq \min(Int(C_1) + \tau(C_1), Int(C_2) + \tau(C_2)). \quad (3)$$

The threshold function  $\tau$  is given by  $\tau(C) = \frac{k}{|C|}$ , where  $|C|$  stands for the number of pixels in component  $C$ ,  $k$  is a constant,  $Int(C)$  is defined as the largest weight in the minimum spanning tree for  $C$  and  $d(C_1, C_2)$  is defined as the minimum weight edge connecting  $C_1$  and  $C_2$ .

It can be observed that the edges causing merging of components are the same as the edges the Kruskal's algorithm would select when constructing the minimum spanning tree for each component.

It is remarkable that there is only one parameter involved in the predicate, the constant  $k$ , which makes the method fairly adjustable to an application at hand. The parameter  $k$  controls the degree of similarity between the components and hence the final number (and average size) of segmented areas.

## 3 Proposed 3D Extension of SRM and EGS Techniques

We extend the SRM and EGS techniques to a 3D setting by considering a stack of CT slices as an input. In graph-based setting it means that we build a three-dimensional graph model where each image pixel becomes a vertex in the graph and is connected in space by an edge to 6 immediate neighbours: four from the slice it belongs to, one from the preceding slice - spatially positioned directly above it - and one from the following slice - spatially located directly below, except for those vertices coming from the very first and very last slice. Each edge has a positive weight assigned as an absolute value of the intensity difference in the corresponding pixels.

Observe that this way each vertex of the new weighted graph corresponds to the possible location of boundary points along the length of the elongated region - corresponding to a tissue. By construction, all stack slices are connected with each other and will participate effectively in the segmentation process. Thus, if there is a problem with 2D segmentation of a tissue in a single slice it might likely be

fixed when it is embedded in stack of slices and then 3D segmented. Figure 1 shows an example when the spinal cord is hard to 2D segment since it 'leaks' through the surrounding bone tissue. However, the two neighbouring slices do not suffer from this issue - the spinal cord is entirely enclosed by the bone. Thus, when embedded in 3D stack and segmented by the extended 3D method, the spinal cord can be well outlined (Figure 1(f)).

3D extension allows for tracking of components across all slices. This is also illustrated in Figure 1. The bone surrounding the spinal cord in slice 68 is segmented into two different components by 2D segmentation (shown by different colours used for both pieces in Figure 1 (d)), while the same bone considered as 3D object became one component (Figure 1(f)) when segmented by our 3D method.

It is a straightforward observation that the underlying principles of both SRM and EGS, outlined in Section 2, can readily be applied to our 3D graph setting and our proposed algorithm, for both SRM and EGS technique, can be summarized as follows.

1. Construct the weighted graph  $G = (V, E)$  from the stack of CT images.
2. Sort the edges into a sequence  $S$  in ascending order of their weights.
3. Initialize the list of components (each vertex constitutes a single component).  
**repeat**
3. Take an edge  $s$  from  $S$  and apply the merging predicate (1) (or (3)) to the components connected by  $s$ . Update the list of components if merging occurred.
- until  $S$  is exhausted**

When all edges had been traversed, in case of EGS, each component is created as a minimum spanning tree, while for SRM components are true statistical regions, retrieved by the algorithm.

## 4 Experiment

### 4.1 Data Set and Ground Truth

A dataset of the torso CT scan of a 14-year-old female (ADELAIDE [1]) was used in this study. The dataset contains a total of 55 CT slice images. The CT scan has a field of view of radius 145mm from the scanner's isocentre. This resulted in the truncation of some of the anatomy at the shoulders and hips. The images have a pixel size of  $2.53 \times 2.53$ mm and slice separation of 10mm. The dataset was annotated by one of the authors (MC) who is an expert in human anatomy. Each of the organ/tissue-of interest in this study was manually delineated by the expert. These includes heart, liver, spleen, stomach, kidneys, lungs, spinal cord, bones and others.

### 4.2 Segmentation Evaluation Methodologies

A particular tissue/organ, when segmented, consists of a small number of adjacent regions of pixels (ie components) that are assigned a different colour (see, for example, the body of the vertebra in Figure 1(f)). When these adjacent regions are merged, a true representation of the shape of the tissue in that slice should be achieved. As the goal of this study is to explore the full potential of the proposed 3D-SRM and 3D-EGS techniques, a suitable criterion for measuring the quality of the segmentation is to determine (when the relevant component are merged) how closely the union of all relevant components resembles the shape of the tissue as determined by an expert. By a relevant component it is meant a component with at least half of its area residing within the region annotated by the expert.

Thus, for a given annotated region  $T$ , A component  $C$  is considered relevant if  $|C \cap T| > |C \setminus T|$ , where  $|A|$  indicates the number of pixels in the component  $A$ .

The segmentation results are evaluated using three different measures. Each measure is employed to assess a different attribute of the segmentation results. In principle, evaluations should be performed on 3D segmented object as the segmentation technique is 3D-based. This means, when considering, for instance, the commonly used Dice index described below, that the overlapped volume rather than overlapped area should be measured. However, due to the anisotropic nature in the dataset to avoid big interpolation errors, the 3D segmentations are assessed on slice-by-slice basis (as 2D projections of segmented volumes).

Accuracy of segmentation is often measured by quantifying the agreement of two sets with one representing the segmentation result and the other standing for the ground truth or the reference. A number of indices can be used to quantify that measure with the Dice index being one of the most widely accepted. Let  $A$  and  $B$  be the two sets to be assessed, the Dice index is defined as

$$D(A, B) = \frac{2|A \cap B|}{|A| + |B|}.$$

The index ranges between 0 and 1. A value of 1 indicates that the two sets agreed perfectly and a value of 0 indicates that no agreement was found between the two sets.

Another common measure used in evaluating segmentation, particularly of medical images, is the Hausdorff distance (see e.g. [28]). Let  $A$  and  $B$  be two sets where  $A = \{a_1, a_2, \dots, a_m\}$  and  $B = \{b_1, b_2, \dots, b_n\}$  with  $m$  and  $n$  number of elements, the Hausdorff distance is defined as

$$H(A, B) = \max(h(A, B), h(B, A)),$$

where

$$h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\|$$

and  $\|\cdot\|$  is some underlying norm on the points of  $A$  and  $B$ . In essence, the Hausdorff distance provides indication of the worst deviation (largest distance) between the segmented results and the ground truth.

Specific to the evaluation of segmentation of medical images is the uncertainty in the ground truth. In the area of medical image segmentation, ground truths are typically obtained involving manual delineation of boundaries within images by human experts. An intrinsic issue with that is inter- and intra-observer variability. The above indices take the ground truth as an absolute ground truth with no room for human errors. This does not address the uncertainty in the ground truth. The  $H_t$  metric ([29], [30]), on the other hand, addresses the uncertainty in the ground truth and measures the border accuracy with tolerance. Let  $N_A$  and  $N_B$  denote the number of pixels in boundary  $A$  and  $B$ , respectively. Then the  $H_t$  is given as

$$H_t(A, B) = \frac{1}{2} \left( \frac{N_{A_t}}{N_A} + \frac{N_{B_t}}{N_B} \right),$$

where the parameter  $t$  is an interval of tolerance, and  $N_{A_t}$  and  $N_{B_t}$  are the number of pixels in boundary  $A$  and  $B$  correctly identified with a tolerance  $t$ . The  $H_t$  metric increases monotonically with  $t$ , and converges to 1. For two borders that are exactly the same, the  $H_t$  is equal to unity with  $t$  set to zero. Briefly, the  $H_t$  metric is the sum of equally weighted fractions of border  $A$  and border  $B$  correctly identified within a certain tolerance.

Table 1: Evaluation segmentation of eight tissue/organs based on 3D-SRM and 3D-EGS methods. The Dice index, the  $H_t$  measure and the Hausdorff Distance measure are shown. Table entries are (mean  $\pm$  standard deviation) over the relevant CT slices for the specific tissue/organ.  $H_1$  and  $H_2$  are the  $H_t$  measure with the tolerance parameter  $t$  set to 1 and 2 (pixels), respectively.

Tissue	Method	Dice Index	$H_t$		Hausdorff Distance (mm)
			$H_1$	$H_2$	
Lungs	3D-SRM	$0.95 \pm 0.06$	$0.96 \pm 0.03$	$0.98 \pm 0.02$	$8.65 \pm 7.19$
	3D-EGS	$0.96 \pm 0.05$	$0.97 \pm 0.04$	$0.98 \pm 0.03$	$7.95 \pm 7.34$
Heart	3D-SRM	$0.93 \pm 0.06$	$0.86 \pm 0.09$	$0.93 \pm 0.07$	$10.61 \pm 5.65$
	3D-EGS	$0.93 \pm 0.06$	$0.85 \pm 0.11$	$0.92 \pm 0.09$	$11.74 \pm 7.88$
Liver	3D-SRM	$0.92 \pm 0.05$	$0.80 \pm 0.15$	$0.87 \pm 0.13$	$18.30 \pm 16.36$
	3D-EGS	$0.87 \pm 0.09$	$0.70 \pm 0.18$	$0.79 \pm 0.15$	$28.49 \pm 32.18$
Kidneys	3D-SRM	$0.93 \pm 0.04$	$0.88 \pm 0.12$	$0.95 \pm 0.07$	$10.45 \pm 8.06$
	3D-EGS	$0.88 \pm 0.06$	$0.77 \pm 0.14$	$0.86 \pm 0.10$	$16.02 \pm 9.24$
Spleen	3D-SRM	$0.87 \pm 0.10$	$0.79 \pm 0.14$	$0.88 \pm 0.10$	$18.02 \pm 13.80$
	3D-EGS	$0.76 \pm 0.23$	$0.72 \pm 0.14$	$0.83 \pm 0.10$	$21.83 \pm 19.90$
Stomach	3D-SRM	$0.72 \pm 0.20$	$0.74 \pm 0.05$	$0.82 \pm 0.06$	$17.20 \pm 6.85$
	3D-EGS	$0.69 \pm 0.21$	$0.74 \pm 0.07$	$0.83 \pm 0.06$	$19.32 \pm 10.81$
Spinal Cord	3D-SRM	$0.83 \pm 0.09$	$0.95 \pm 0.06$	$1.00 \pm 0.02$	$3.67 \pm 1.20$
	3D-EGS	$0.79 \pm 0.13$	$0.91 \pm 0.09$	$0.98 \pm 0.06$	$3.65 \pm 1.21$
Bones	3D-SRM	$0.89 \pm 0.04$	$0.98 \pm 0.02$	$0.99 \pm 0.01$	$12.93 \pm 15.91$
	3D-EGS	$0.88 \pm 0.05$	$0.96 \pm 0.02$	$0.98 \pm 0.02$	$17.90 \pm 16.48$

Other than measuring the fidelity of the segmentation results, oversegmentation or undersegmentation is another criterion in judging a segmentation algorithm. Ideally, the whole organ is segmented as one single region/component. But this is seldom achieved. Often, either undersegmentation results, that is, a segmented region covers more than one true organ or oversegmentation happens, when a true organ is segmented into multiple components. The former is reflected in the Dice index, while the latter is explicitly measured by reporting the number of segmented regions required to approximate the coverage of the segmentation result in this paper.

## 5 Experimental Results

Using the proposed 3D-SRM and 3D-EGS methods, eight representative tissue/organs in CT images were segmented. The eight tissue/organs were lungs, heart, liver, kidneys, spleen, stomach, spinal cord and bones (including all bones). According to their statistical complexities and sizes, these eight tissue/organs can be largely grouped into two groups with the lungs, heart, liver, kidneys and the spinal cord in one group, and spleen, stomach and bones in the other. This is reflected in the choice of the values of the parameter in both methods. For the SRM-based method, the  $Q$  value was set to 128 for the first group (except for the spinal cord where due to small size it was set to 2000 to avoid undersegmentation) and 512 for the other group. For the EGS-based method, the  $k$  value was set to 3000 for lungs, heart, liver and the spinal cord, 2000 for kidneys, and 1000 for spleen and bones.

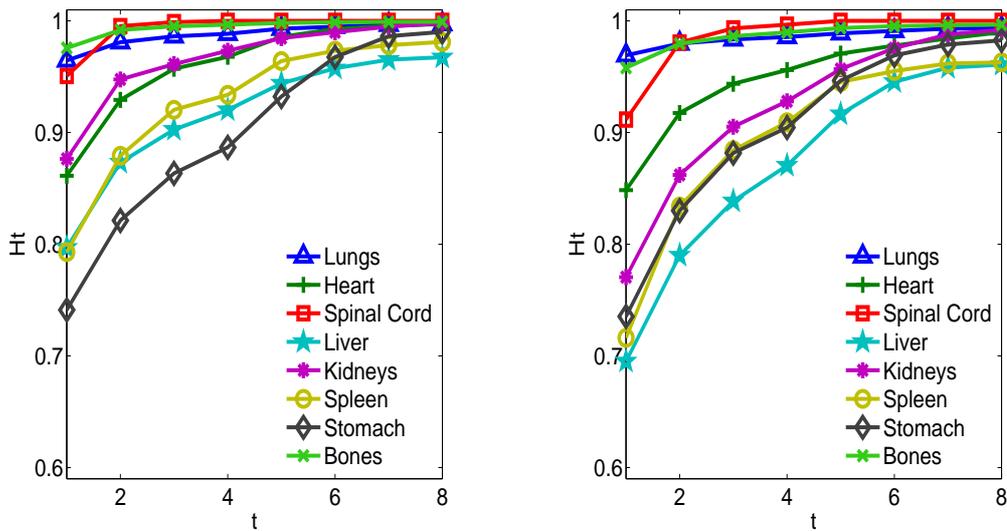


Figure 2:  $H_t$  plots for eight tissue/organs segmented using 3D-SRM (left) and 3D-EGS (right) methods. The tolerance parameter  $t$  of the  $H_t$  metric ranges from 1 to 8 (pixels). The 1st and 2nd markers of each  $H_t$  profile are the  $H_1$  and  $H_2$  reported in Table 1.

Table 1 shows the evaluations of segmentation using 3D-SRM and 3D-EGS methods. The Dice index,  $H_t$  measure and Hausdorff distance for each of the eight tissue/organs were depicted. These indices were computed by taking the average over all relevant CT slices that a particular tissue/organ was considered. For example, the lungs span over 22 CT slices. The, say, 3D-SRM based, segmented lungs in each CT slice were evaluated and resulted in individual Dice values. The average (and the standard deviation) of these 22 Dice values is the first entry in the column labelled 'Dice Index' in Table 1. For lungs, heart, liver, kidneys and bones, the mean Dice indices are close to 0.90 and over (0.89 to 0.95) for 3D-SRM and ranges from just below 0.90 to 0.96 (0.87 to 0.96) for 3D-EGS. This shows that the performance of the segmentations based on 3D-SRM and 3D-EGS for this group of organs/tissues is high. For spleen and stomach, the Dice indices are 0.87 and 0.72 using the 3D-SRM method and 0.76 and 0.69 using the 3D-EGS method. The performance for this group is also relatively high but slightly lower than that in the previous group, as spleen and stomach are typically difficult to distinguish from the neighbouring soft tissues in the abdominal area (even to the human expert's eyes). For the spinal cord, the Dice index is 0.83 and 0.79 for 3D-SRM and 3D-EGS, respectively. The performance for this group is also relatively high but slightly lower than that in the first group. This is because the spinal cord is represented by only a few pixels and the Dice index is an area measurement. The inclusion or exclusion of one extra pixel will have a large impact in the percentage change. Overall, the segmentation results for 3D-SRM in all groups are higher than that for 3D-EGS.

For the  $H_t$  measure, only  $H_1$  and  $H_2$  with tolerance of 1 and 2 pixels, respectively, were shown in Table 1. Taken into consideration the uncertainty in the ground truth (produced by manual delineation of the borders), allowing a 2 pixels tolerance in the agreement between the segmentation and the ground truth is practical. Table 1 shows that the segmentation performance measured by border accuracy using the  $H_2$  index is high for both 3D-SRM and 3D-EGS in group 1 organs/tissues, around 0.95 or above for 3D-SRM and ranges from 0.86 to above 0.95 in group 1 except for liver. For spleen and stomach, the  $H_2$  index is also relatively high with 0.88 and 0.82 for 3D-SRM and 0.83 and 0.83 for 3D-EGS. The  $H_2$  values for spinal cord based on both segmentation methods are very close to 1. This is because for a reasonably accurate segmentation of the small size spinal cord, a tolerance of 2

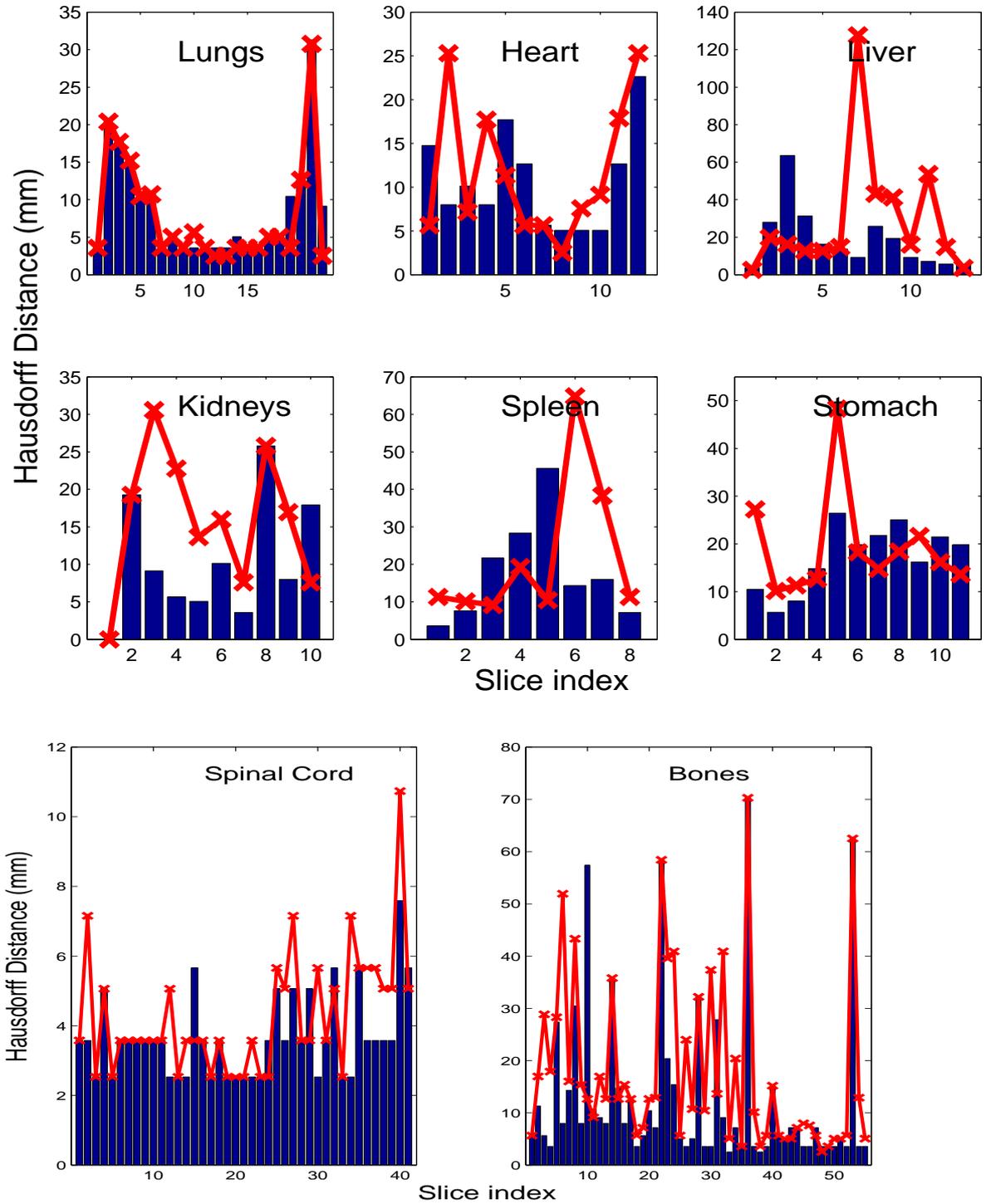


Figure 3: Hausdorff distance (in millimeters) for each slice for all eight organs/tissues. Bars indicate values for 3D-SRM while (red) crosses are used to indicate outcomes for 3D-EGS technique.

pixels almost certainly will capture the ground truth. Overall, the  $H_t$  index for 3D-SRM in all groups is higher than that for 3D-EGS. The  $H_t$  measures with the tolerance parameter values  $t$  ranging from

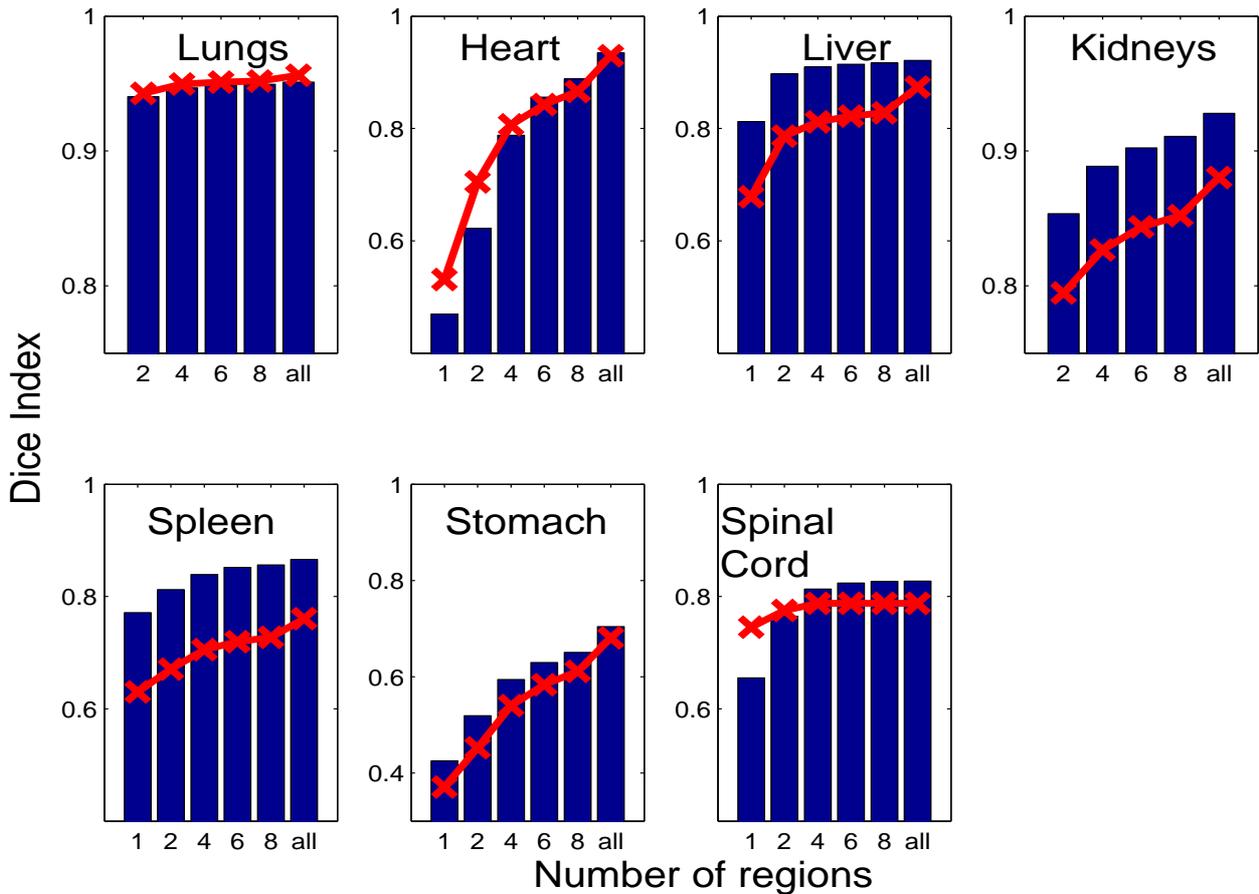


Figure 4: Efficiency of the 3D-SRM and 3D-EGS segmentation methods. For each of the seven tissue/organs shown, using only the largest 2 to 8 relevant SRM (blue bars) or EGS (red cross) regions can well approximate the coverage achieved by all relevant components as indicated by the Dice index).

1 to 8 (pixels) for all eight tissue/organs segmented using both 3D-SRM and 3D-EGS methods are shown in Figure 2. It can be observed that the  $H_t$  index increases rapidly with  $t$ , stabilising at  $t$  over 2 for lungs, heart and spinal cord and at  $t$  over 8 for the other organs/tissues. Overall, the  $H_2$  index for 3D-SRM in all groups is higher than that for 3D-EGS.

The mean Hausdorff distances for all organs/tissues are depicted in the last column in Table 1. The variation of the Hausdorff distance in each organ/tissue seems to be large as indicated by the standard deviation. Figure 3 depicts the Hausdorff distance for each slice of each organ/tissue. It can be observed that the large variation in each organ/tissue is typically due to extreme results in one or two slices.

Coverage of the 3D-SRM and 3D-EGS segmentation results is shown in Figure 4. Using the 3D-SRM or 3D-EGS methods, the segmentation result of a particular tissue/organ is, typically, represented by the union of a number of 3D-SRM regions or 3D-EGS regions. The number of the (3D-SRM or 3D-EGS) regions in the union set is significantly smaller than the number of pixels (voxels) in that tissue/organ. Figure 4 shows the coverage (of the region(s) compared to the ground truth) using only the largest 2 to 8 and all eligible regions for 3D-SRM (blue bars) and 3D-EGS (red cross) for each of seven tissue/organs. It is visible from these graphs that using only the largest 2 to 8 eligible regions

can well approximate the performance achieved by all eligible components as indicated by the Dice index (in both methods). Only seven tissue/organs were shown. Bones are not included as bones comprise a large number of entities. Naturally, bones need large number of eligible components to approximate the final coverage. This does not provide any useful insight about the efficiency of the 3D-SRM and 3D-EGS methods, thus, is not shown.

## 6 Discussion

The results show that both SRM and EGS extensions to 3D setting have a great potential for CT segmentation. It is transparent that overall 3D-SRM outperforms 3D-EGS in all aspects; accuracy - measured by the Dice index and the Hausdorff distance (Table 1), and also the granularity of segmentation - measured by accuracy vs number of components per tissue/organ (Figure 4) are better (for some organs, like spleen, liver or kidneys, significantly better) for 3D-SRM than for 3D-EGS. This can be attributed to the well-known fact that the EGS technique is significantly more sensitive to noise than the SRM method ([16]) and spleen, liver and kidneys often have large portions of their boundaries blurred and fuzzy.

In literature, multi-organ segmentation is hardly addressed so far. This is partly because of high computational complexity of the task and also difficulty in gathering data. Most published results in the area focus on segmentation of the abdominal region. [10] reported segmentation results of kidneys, spleen and liver in contrast-enhanced two phases abdominal CT scan. The results for the Dice index were 0.93, 0.91 and 0.95, respectively. Another study [7] reported segmentation of six tissue/organs (heart, liver, spleen, lungs, kidneys and pelvis) in CT scans. The results were evaluated using Dice index and Hausdorff distance but were given in small plot profiles and difficult to read accurately. Dice indices for the 6 tissue/organs were best read out as 0.7, 0.8, 0.6, 0.9, 0.5 and 0.6, respectively, and that of Hausdorff distance as 12, 14, 18, 6, 19, and 15 mm, respectively. Both our 3D-SRM and 3D-EGS results are very comparable to these results.

## 7 Conclusion

In this paper, we have shown that the novel 3D segmentation techniques introduced in [6] based on Statistical Region Merging and the Efficient Graph-based Segmentation have a great potential to become methods of choice for full body CT segmentations for the purpose of CT dose estimation. In particular, we have addressed the issue of over- and under-segmentation which is intrinsic for medical image segmentation. Both methods achieve very high accuracy of the outcome with negligible oversegmentation and also compare very favourably with limited quantitative relevant results found in the literature.

It is worth pointing out that although 3D-EGS technique produced slightly worse results it is likely to be improved when a pre-processing is incorporated into the framework. Image smoothing is likely to improve the 3D-EGS outcome since the method is known to be noise sensitive. Furthermore, contour smoothing, using for example, active contour models is also likely to improve further both results. In this study, to keep the outcome as generic as possible, we did not apply any image noise reduction techniques or contour smoothing, which are data dependant.

## Acknowledgement

This work was partially supported by the Channel 7 Children's Research Foundation Grant (ref.12468/2012).

## References

- [1] M. Caon, G. Bibbo, and J. Pattison. An EGS4-ready tomographic computational model of a 14-year-old female torso for calculating organ doses from CT examinations. *Physics in Medicine and Biology*, 44(9):2213–2225, 1999.
- [2] M. Zankl, R. Veit, G. Williams, K. Schneider, H. Fendel, N. Petoussi, and G. Drexler. The construction of computer tomographic phantoms and their application in radiology and radiation protection. *Radiation and Environmental Biophysics*, 27:153–164, 1988.
- [3] C. Lee, J. Williams, C. Lee, and W. Bolch. The UF series of tomographic computational phantoms of pediatric patients. *Medical Physics*, 32:3537–48, 2005.
- [4] C. Lee et al. The UF family of reference hybrid phantoms for computational radiation dosimetry. *Physics in Medicine and Biology*, 55:339–863, 2010.
- [5] ICRP. The 2007 recommendations of the International Commission on Radiological Protection. *Annals of the ICRP*, 103, 2007.
- [6] M. Bajger, G. Lee, and M. Caon. Full-body CT segmentation using 3D extension of two graph-based methods: a feasibility study. In *Signal Processing, Pattern Recognition and Applications (SPPRA 2012): Proc. of the IASTED International Conference*, pages 43–50, Crete, Greece, 2012.
- [7] J.E. Iglesias, E. Konukoglu, A. Montillo, Z. Tu, and A. Criminisi. Combining generative and discriminative models for semantic segmentation of CT scans via active learning. In *Proc. of IPMI*, volume 6801 of *LNCS*, pages 25–36, 2011.
- [8] A. Montillo, J. Shotton, J. Winn, J. E. Iglesias, D. Metaxas, and A. Criminisi. Entangled decision forests and their application for semantic segmentation of CT images. In *Information Processing in Medical Imaging 2011*, volume 6801 of *LNCS*, pages 184–196, 2011.
- [9] S. Seifert, A. Barbu, S. K. Zhou, D. Liu, J. Feulner, M. Huber, M. Suehling, A. Cavallaro, and D. Comaniciu. Hierarchical parsing and semantic navigation of full body CT data. In *Medical Imaging 2009: Physics of Medical Imaging, Proc. of SPIE*, volume 7258, pages 725902–8, 2009.
- [10] M.G. Linguraru, J.A. Pura, A.S. Chowdhury, and R.M. Summers. Multi-organ segmentation from multi-phase abdominal CT via 4D graphs using enhancement, shape and location optimization. In *Proc. of MICCAI*, volume 6363 of *LNCS*, pages 89–96, 2010.
- [11] T. Okada, K. Yokota, M. Hori, M. Nakamoto, H. Nakamura, and Y. Sato. Construction of hierarchical multi-organ statistical atlases and their application to multi-organ segmentation from CT images. In *Proc. of MICCAI, Part I*, volume 5241 of *LNCS*, pages 502–509, 2010.

- [12] M.G. Linguraru, J.K. Sandberg, Z. Li, F. Shah, and R. M. Summers. Automated segmentation and quantification of liver and spleen from CT images using normalized probabilistic atlases and enhancement estimation. *Med. Phys.*, 37 (2):771–783, 2010.
- [13] A. El-Baz, G. Gimel'farb, R. Falk, and M.A. El-Ghar. Automatic analysis of 3D low dose CT images for early diagnosis of lung cancer. *Pattern Recognition*, 42:1041–1051, 2009.
- [14] J. Feulner, S. K. Zhou, M. Huber, A. Cavallaro, J. Hornegger, and D. Comaniciu. Model based esophagus segmentation from CT scans using a spatial probability map. In *Proc. of MICCAI 2010, Part I*, volume 6361 of *LNCS*, pages 95–102, 2010.
- [15] F. Heckel, O. Konrad, H. K. Hahn, and H-O. Peitgen. Interactive 3D medical image segmentation with energy-minimizing implicit functions. *Computers and Graphics*, 35:275–287, 2011.
- [16] R. Nock and F. Nielsen. Statistical Region Merging. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(11):1452–1458, 2004.
- [17] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *Int. J. Comput. Vision*, 59(2):167–181, 2004.
- [18] F. Ma, M. Bajger, J.P., Slavotinek, and M. J. Bottema. Two graph theory based image segmentation methods for identifying pectoral muscle in mammograms. *Pattern Recognition*, 40:2592–2602, 2007.
- [19] M. Bajger, F. Ma, S. Williams, and M. Bottema. Mammographic mass detection with Statistical Region Merging. In *Digital Image Computing: Techniques and Applications, Proc. of DICTA 2010*, pages 27–32, Sydney, Australia, 2010.
- [20] M.E. Celebi, H.A. Kingravi, H. Iyatomi, Y.A. Aslandogan, W. Stoecker, R. Moss, J.M. Malterers J.M. Grichnik, A.A. Marghoob, H. Rabinovitz, and S.W. Menzies. Border detection in dermoscopy images using statistical region merging. *Skin Research and Technology*, pages 347–353, 2008.
- [21] G. Lee, M. Bajger, and M. Caon. Multi-organ segmentation of CT images using statistical region merging. In *Biomedical Engineering (BioMed 2012): Proc. of the IASTED International Conference*, pages 199–206, Innsbruck, Austria, 2012.
- [22] S. Battiato, C. Bosco, G. M. Farinella, and G. Impoco. 3D CT segmentation for clinical evaluation of knee prosthesis operations. In *Proceedings of the Eurographics Italian Chapter*, 2006.
- [23] F. Nielsen and R. Nock. On region merging: the statistical soundness of fast sorting, with applications. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pages 19–27, 2003.
- [24] C.T. Zahn. Graph-theoretic methods for detecting and describing gestals clusters. *IEEE Transactions on Computing*, 20:68–86, 1971.
- [25] R. Urquhart. Graph theoretical clustering based on limited neighbourhood sets. *Pattern Recognition*, 15:3:173–187, 1982.

- [26] P. Felzenszwalb and D. Huttenlocher. Image segmentation using local variation. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pages 98–104, 1998.
- [27] Q.H. Huang, S. Y. Lee, L. Z. Liu, M. H. Lu, L. W. Jin, and A. H. Li. A robust graph-based segmentation method for breast tumors in ultrasound images. *Ultrasonics*, 52:266–275, 2012.
- [28] D.P. Huttenlocher, G.A. Klanderman, and W.J. Rucklidge. Comparing images using the Hausdorff distance. *IEEE Trans. on Pattern Analysis and Machine Intel.*, 15 (9):850–863, 1993.
- [29] O. Kubassova, R. D. Boyle, and A. Radjenovic. A novel method for quantitative evaluation of segmentation outputs for dynamic contrast-enhanced MRI data in RA studies. In *Proc. of the Joint Disease Workshop, 9th International Conference on Medical Image Computing and Computer Assisted Intervention*, volume 1, pages 72–79, 2006.
- [30] M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis, and Machine Vision*. Thomson, 2008.