Proceedings of the IASTED International Conference
Biomedical Engineering (BioMed 2012)
February 15 - 17, 2012 Innsbruck, Austria

MULTI-ORGAN SEGMENTATION OF CT IMAGES USING STATISTICAL REGION MERGING

Gobert Lee
Flinders University
School of Computer Science, Engineering and Mathematics
Bedford Park, Adelaide, Australia
gobert.lee@flinders.edu.au

Mariusz Bajger
Flinders University
School of Computer Science, Engineering and Mathematics
Bedford Park, Adelaide, Australia
mariusz.bajger@flinders.edu.au

Martin Caon
Flinders University
School of Nursing
Bedford Park, Adelaide, Australia
martin.caon@flinders.edu.au

ABSTRACT
Segmentation is one of the key steps in the process of developing anatomical models for calculation of safe medical dose of radiation for children. This study explores the potential of the Statistical Region Merging segmentation technique for tissue segmentation in CT images. An analytical criterion allowing for an automatic tuning of the method is developed. The experiments are performed using a data set of 54 images from one patient, demonstrating the validity of the proposed criterion. The results are evaluated using the Jaccard index and a measure of border error with tolerance which addresses, application-dependant, acceptable error. The outcome shows that the technique has a great potential to become a method of choice for segmentation of CT images with an overall average boundary precison, for six representative tissues, equal to 0.937.

KEY WORDS
Voxel model, image segmentation, statistical region merging.

1 Introduction

Computed tomography (CT) is a relatively high dose radiological procedure but accurate estimates of the dose to an individual are difficult to obtain. Models of human anatomy are required in order to allow the calculation of organ doses and effective dose received from diagnostic radiography procedures such as computed tomography [1]. The cross-sectional (tomographical) medical images produced by medical imaging are a source of data for constructing models of human anatomy. The models are called voxel tomographic models as the pixels of an image are extrapolated into volume elements by extending them into the third dimension - the space between adjoining images [2].

These computational models, when coupled to a Monte Carlo radiation transport code, can be used to calculate the energy deposited by individual x-ray photons in each voxel. The energy deposited in all the voxels assigned to an organ or tissue are summed to obtain the organ dose of interest [3]. A motivation for accurately calculating radiation dose to organs from CT, is to reduce the dose (once known), by more appropriate choice of exposure factors for the CT examination. Modifying the exposure factors to make them appropriate for children of different size is necessary, as x-rays are attenuated by different amounts depending on the distance that they travel through tissue. The dose delivered to children who undergo CT procedures is of particular interest as children have a long lifetime ahead of them during which they may develop a cancer, one cause of which may be their exposure to medical x-rays. In the case of calculating dose to children, it is most useful to have on hand a range of anatomical models that span the size range of children from newborn to about 15 years of age. Then by choosing the anatomical model whose size most closely matches that of the child undergoing a procedure, a realistic dose to that child may be calculated.

Two impediments to the development of a range of anatomical models are: it is difficult to obtain an extensive data set of medical images (one that covers the entire anatomy of a person from head to toe); and it is very time-consuming to segment the entire internal anatomy of the several hundred images that constitute an extensive CT examination. Consequently, few such voxel models of children exist [4], [5].These models are created using labour intensive manual image segmentation with various degrees of software assistance. In the case of [5] the head, torso and limbs from different individuals were scaled and joined together to form composite models.

The first impediment is to some extent overcome by the availability of images from combined PET/CT machines where the entire body may be imaged with low-dose CT prior to positron emission tomographical scanning. Unfortunately the resulting CT images have lower resolution when compared to the images from a normal CT scanner. Attempts to overcome the second impediment have generated various semi-automated or supervised methods of segmenting images by using PC image processing software

DOI: 10.2316/P.2012.764-052

and some basic functions and facilities such as B-spline functions, threshold numbers and region growing to manually draw organ boundaries (e.g. [5], [6], [7]). Nevertheless these existing methods are still time-consuming to apply to the hundreds of images that span a human body. There remains a need to develop semi-automated image processing techniques capable of segmenting all organs and tissues visible in images.

Ideally, the desirable segmentation method will be highly accurate, time and memory efficient, require minimum user interaction and allow for an intuitive tuning for a precise delineation of organs. Statistical Region Merging (SRM) [8] is an image segmentation algorithm which has a potential to fulfill all these requirements.

As shown in [9], SRM outperforms state-of-the art segmentation methods such as mean-shift segmentation, JSEG and the efficient graph-based segmentation in terms of processing speed. This is important as there are several hundred images in just one single CT examination. SRM is also found to be more stable with respect to different images than JSEG and the mean-shift but comparable in performance to the efficient graph-based method. However, as proved in [8], SRM is significantly better than the efficient graph-based method in handling noise.

The SRM method is backed up by solid well-established probability theory. It was proved in [8] that with a proper choice of the parameter value δ, described in Section 2.3, the probability of undermerging (oversegmentation) is very small. Furthermore, the segmentation granularity is tuned using a single intuitive parameter, Q, which, as we show later, can be well-determined allowing for an optimal tissue segmentation. Being only controlled by a single parameter the method requires very little user interaction; as shown below the parameter value can be estimated analytically prior to experiments saving users time and decreasing overall cost of the application. Finally, SRM merging strategy is linear with respect to the number of pixels and has low time and memory complexity thus making the technique a good candidate for our task.

In what follows, we explore capabilities of the method in the context of CT images. Our ultimate goal in this study is to set the ground for building an automatic image analysis system utilizing SRM method. Thus, the objective is not necessarily to make it optimal or exact in terms of organ delineation but rather evaluate its potential towards both accuracy and robustness of the results. The true measure of the method performance would be a reliable anatomical model of a child but this is an unrealizable criterion during the development stage. Thus, the approach adopted here is to compare segmented organs with those provided by an expert in human anatomy. To keep the outcome as generic as possible we do not apply any noise reduction on images (which is always highly related to a particular set of images) and allow for multiple (carefully optimized) segmented regions within organs. This way, the full potential of SRM towards accurate CT segmentation is explored.

The SRM technique, although fairly new, has already been applied in medical images on several occasions. In [10] and [11] SRM was used for an automatic diagnosis of melanoma, in [12] for an automatic detection of breast cancer, and in [13] for evaluation of the post-operative outcome of knee prosthesis implantation.

Named tissues	Remainder tissues	Other tissue
Red Bone-marrow	Adrenals	Fat
Colon	Thoracic region	Internal Gas
Lung	Gall bladder	
Stomach	Heart	Trachea
Breast	Kidneys	Soft tissue
Gonads	Lymphatic nodes	Spinal cord
Bladder	Muscle	
Oesophagus	Oral mucosa	
Liver	Pancreas	
Thyroid	Prostate	
Bone surface	Small intestine	
Brain	Spleen	
Salivary glands	Thymus	
Skin	Uterus/cervix	

Table 1: List of tissues.

2 Methods and Materials

2.1 Data Set

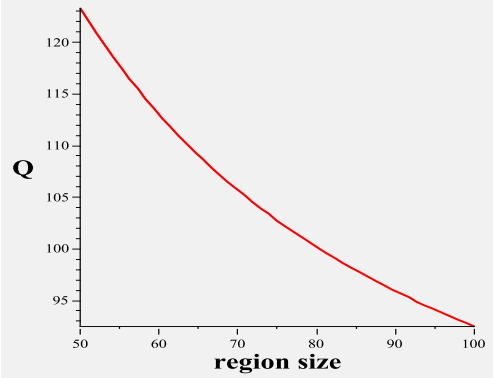
The 54 images employed in this study are of the torso of a 14-year-old female patient whose weight was about 48kg. The images were retrieved from the archive of normally scheduled procedures and patient identifying data were deleted. The images have a field of view of radius 145mm from the scanner,s isocentre. This resulted in the truncation of some of the anatomy at the shoulders and hips. The images have a pixel size of 2.53×2.53mm and slice separation of 10mm. The data set is referred to as ADELAIDE [1].

2.2 Ground Truth

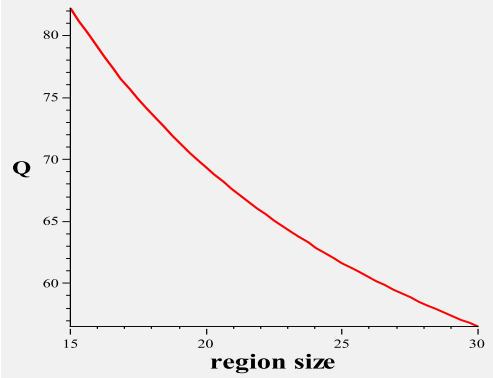
To validate the ability of our image segmenting algorithm to produce an accurate segmentation, the automatically segmented images are compared to images manually segmented by one of the authors (MC) who has 20 years of experience teaching human anatomy and physiology. For feasibility of manual segmentation the images were resized to 126×128 pixels. Subsequently the automatic segmentation with the SRM method was performed on the resized

images.

To calculate effective dose, the organ dose to 14 named tissues and 14 remainder tissues must be known [14]. Consequently these tissues (see Table 1) must be identified by the segmentation. Other tissues that were manually segmented are listed in the third column of the table.



(a)



(b)

Figure 1: Estimation of the parameter Q for SRM method. (a) Estimation for the heart, liver and kidney. (b) Estimation for the bones, spinal cord and the lungs.

2.3 Statistical Region Merging

The statistical region merging technique proposed in [8] considers image segmentation as an inference problem. The image itself is considered as an observed instance of some unknown perfectly segmented image. The true (or statistical) regions of the perfect image are to be reconstructed. One of the important advantages of the technique is that it is backed up by probabilistic concentration theory and as a result, as shown in Section 2.4, its outcome can be estimated prior to the experiment.

Let I be an observation of a true image I^*. Suppose that each pixel in I^* is represented by a family of distributions from which the observed intensity is sampled. The optimal (statistical) regions in I^* possess the homogeneity property: *all pixels have the same expectation across the region and the expectations of adjacent regions differ.* Thus, I is obtained from I^* by sampling statistical pixels for the observed intensity. More precisely, the intensity of each pixel in I is realized as a sum of Q independent random variables, each taking values in $[0, g/Q]$, where g is the number of image intensity levels. The observation was made in [8] that the parameter Q can be seen as a measure of statistical complexity of the image I^*. Higher values of Q result in a finer segmentation.

SRM comprises of two components: a merging predicate and the order of testing the predicate for growing regions. To develop the predicate authors of [8] prove the following: *For any fixed couple (R, R') of regions of I and any fixed $0 < \delta \leq 1$, the probability is no more than δ that*

$$\left|(\bar{R} - \bar{R}') - \mathbb{E}(\bar{R} - \bar{R}')\right| \geq g\sqrt{\frac{1}{2Q}\left(\frac{1}{|R|} + \frac{1}{|R'|}\right)\ln\frac{2}{\delta}}, \tag{1}$$

where $\bar{R}$ denotes the average intensity across the region R and $\mathbb{E}(R)$ is the expectation over all corresponding statistical pixels of I^* of their sum of expectations of their Q random variables for their intensity values. $|\cdot|$ denotes cardinality. Assuming that regions R, R' should be merged if $\mathbb{E}(\bar{R} - \bar{R}') = 0$ formula (1) yields the merging predicate

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

where

$$b(R) = g\sqrt{\frac{1}{2Q|R|}\ln\frac{2}{\delta}}. \tag{3}$$

Note that in the setting of 4-connectivity (which is utilized in this paper) the number of merging tests N for adjacent regions is bounded above, that is, $N < 2|I|$ for an image I. Thus the predicate will be satisfied with the high probability $p \geq 1 - N\delta$ for N merging tests assuming δ is sufficiently small (in the sequel we follow [8] and use the value $\delta = \frac{1}{6|I|^2}$).

The order of merging satisfies the invariant which implies that if two parts of the true regions are tested then all tests inside each of those regions have already being done. Let S_I be a set containing all pairs of adjacent pixels in I (based on 4-connectivity) and let $R(p)$ be the region containing pixel p. The algorithm first sorts those pairs in increasing order according to a function $f(p, p')$. Then the order is traversed one time with the merging performed for regions $R(p)$ and $R(p')$ if the predicate $P(R(p), R(p'))$ holds true. A common choice for function f (utilized in this study) is to use the pixel intensity difference

$$f(p, p') = |p_{int} - p'_{int}|\,. \tag{4}$$

2.4 SRM for CT images

Despite the solid mathematical base and clarity of the condition, the predicate (2) was found to result in overmerging in [8] and replaced by a more sophisticated version for

natural scene images. Our experiments revealed that the original version (2) works well for CT images. Moreover, its simplicity allows for an insight into the connection between the statistical complexity of the image and the size and texture of segmented objects, which allows for a reliable a priori estimation of Q.

In our case, the parameter Q needs to be selected in a way that small segmented regions corresponding to various tissues are not merged with other comparable in size and texture regions. Otherwise, those tissues would be lost.

Using the value of spatial resolution of images, one can estimate that the minimum size of regions corresponding to various organs varies approximately from 25 (for small organs like the spinal cord) to 50 pixels for larger organs like kidneys or heart. Further inspection of CT images reveals that the difference of intensity across the organs can be as small as 15. Armed with these numbers we can use the predicate (2) to estimate the optimal value for Q as follows.

Let $T_0 = \left|\bar{R} - \bar{R}'\right|$ and assume that the merging predicate (2) is true for the regions R, R', where $\bar{R}$ denotes the average intensity across the region R. Solving (2) for Q yields

$$Q \leq \frac{g^2}{2T_0^2}\left(\frac{1}{|R|} + \frac{1}{|R'|}\right)\ln\frac{2}{\delta}. \qquad (5)$$

This gives an upper bound on the value of Q such that two regions with average intensity difference T_0 (or less) will merge. Thus, to prevent merging of regions R, R' such that $|R \cup R'| \approx 100$ and $T_0 = 15$, we get from (5), that Q must be at least equal 123. For $T_0 = 30$ and the objects of total size about 50 the Q value must be at least equal 62 to allow for separation of these objects.

The graphs in Figure 1 show the behaviour of Q as a function of size of one of the regions, with the other region size fixed to 50 or 25 pixels, and threshold T_0 set to 15 and 30, respectively. Note that bigger Q values will result in oversegmentation. Thus, we consider the smallest Q value allowing for region separation being optimal.

In this study we focused on six representative organs: lungs, the spinal cord and bones - with an optimal Q set to 64; and the heart, liver and kidneys - with the optimal Q value chosen as 128, in line with the analytical estimation illustrated by Figure 1.

Figure 2 (a), (c), (e) shows the heart segmented into three different components (left ventricle, right ventricle and aorta) while the expert annotated heart area constitutes of only one component. A close look at the original image reveals that the heart indeed composes of clearly distinguishable three areas of varying texture. Thus, a prior knowledge must be used to guide the segmentation in this case. The goal of this study is to explore full potential of the SRM itself. Accordingly, a suitable criterion for measuring the quality of the segmentation of tissues is to compute how well the union of all relevant components which overlap the annotated tissue fits the annotation. By a relevant component is meant a component with at least half of its area residing within the annotated region. Thus a component C is said to overlap the annotated region T if $|C \cap T| > |C \setminus T|$, where $|A|$ indicates the number of pixels in the component A. The accuracy of the segmentation is evaluated by using Jaccard index (Section 2.5.1) and the H_t metric (Section 2.5.2) by taking the union of all relevant components for the given tissue.

2.5 Evaluation

Manual segmentation by experts is considered as the ground truth in medical image segmentation task. Evaluation on machine segmentation results is made against the ground truth. The following sections describe two evaluation measures that were used in this study. These two measures highlight different attributes of the 'goodness' of the machine segmentation.

2.5.1 Jaccard index

One of the most widely used measures for accuracy assessment is the Jaccard index [15]. It is based on sets agreement and, in this project, is used to quantify the similarity between the machine segmentation and the ground truth (the expert segmentation). The Jaccard index takes on a value between 0 and 1. A value of 1 means that the segmentation result and the ground truth are identical. A value of 0 indicates that the two sets have no common elements. The Jaccard index is defined as

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

and is simply the number of elements common to both sets divided by the total number of elements in the two sets.

2.5.2 Border errors with tolerance (H_t metric)

Border positioning errors are usually taken as the average of the pixel deviation from the true border over the entire length of the border. This gives an overall error measurement but washes out the local deviation information. A more sophisticated measure that retains the local deviation information and takes tolerance into consideration is the H_t metric ([16], [17]). The H_t metric is the average of the fractions of border A and border B correctly identified within a certain tolerance. It 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 t is the tolerance (in pixels), N_A and N_B are the number of pixels in boundary A and B, respectively, 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 the unity with t set to zero.

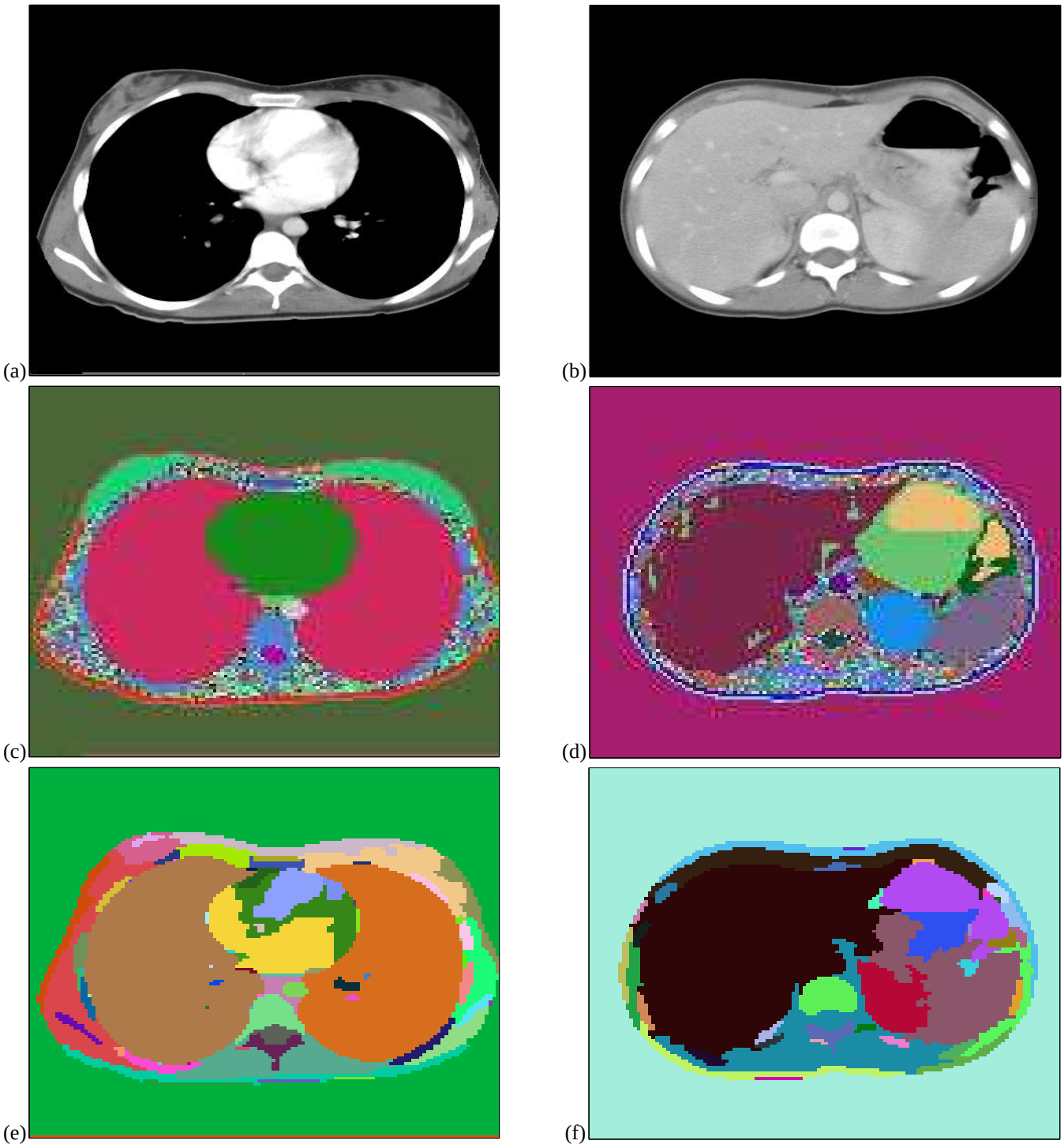


Figure 2: Comparison of an expert and software segmentation. For illustration random colors are used to show segmented regions. Panels in the first column show a slice image containing the heart with (a) the original image, (c) the expert segmented heart (heart area in green) and (e) the SRM segmented heart (yellow, green and blue components). Panels in the second column show a slice image containing the spinal cord with (b) the original image, (d) the expert segmented spinal cord (in green) and (f) the SRM segmented spinal cord (which spills all across the lower part of the image).

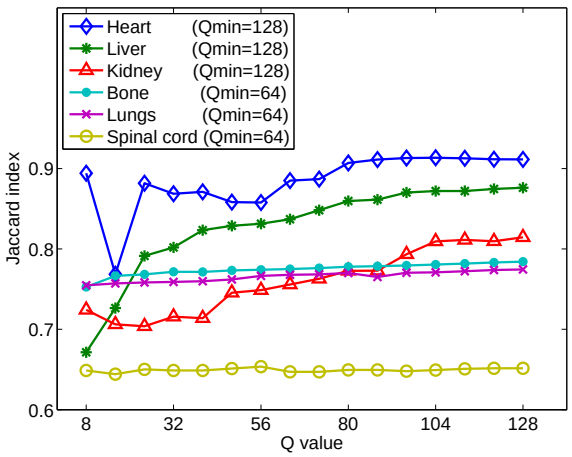


Figure 3: SRM segmentation using a range of Q values. Six different profiles are shown, one for each individual organ/tissue. For each organ, the mean Jaccard index over all relevant slices for SRM results obtained for a given Q value is plotted against that Q value. The estimated Q_{min}, the smallest Q value allowing for region separation, for an organ, is listed in the figure legend.

3 Results

Six anatomical tissues in a torso CT scan were segmented using the SRM method. The six anatomical tissues are lungs, heart, kidneys, liver, spinal cord and bone. For each of the six organ/tissue, Q_{min}, the smallest Q value allowing for optimal region separation (Section 2.4), was computed and approximated by either 64 or 128 .

Results of the SRM segmentations were assessed against the ground truth. Two different measures were used in the evaluation, the Jaccard index and the H_t metric (Section 2.5). Figure 3 shows the Jaccard index for the six organs/tissues. The SRM segmentation was performed on each organ/tissue with the Q value ranging from 8 to 128 in a step of 8. As each organ/tissue appears in more than one CT slices, the Jaccard index shown is the average Jaccard index over all relevant CT slices. It can be observed that for the group with a Q_{min} value of 64 (bone, lungs and spinal cord), the SRM segmentation has reached the optimal point at $Q = 64$. Further increase in the Q value gives no significant improvement in the segmentation, hence Jaccard index stays roughly the same with Q in the range of 64 to 128. In fact, it can be observed that when the experimental Q value is below the Q_{min}, the Jaccard index does not show any significant drop. This is because the Q_{min} is a conservative estimation. The true Q_{min} for this group of organs/tissues is likely to be lower than 64, hence the observation of no change in the experimental Q range of $[8, 128]$. On the other hand, for the group with a Q_{min} value of 128 (heart, liver and kidneys), the experimental Q value is below the optimal value of 128, hence the Jaccard index increases with the increase of Q. Table 2 shows the Jaccard index for the SRM segmentations displayed in Figure 3 with an experimental Q value of 64 and 128 only. For the group of tissues with a Q_{min} of 64, increasing Q value to 128 results in the insignificant increase of approximately 0.01 in the Jaccard value. On the other hand, for the group of tissues with a Q_{min} of 128, using a lower Q value of 64 results in the decrease of approximately 0.05 in the Jaccard value.

The H_t metric is another index used in this study, specifically for the evaluation of the border precision of the segmentation results. The parameter t prescribes the precision tolerance (in pixels). As the tolerance t increases (i.e. becomes more lenient), H_t increases. Figure 4 shows the H_t value for t in the range of [0, 6] for the two groups of organs/tissues, with a Q_{min} of 64 (Figure 4, first column) and 128 (Figure 4, second column), separately. SRM seg-

Table 2: Evaluation of the SRM segmentations using the Jaccard index. Q_{min} value is the smallest Q value for optimal region separation. Segmentation of the six organs/tissues using the SRM method with a Q value of 64 and 128 are reported. The corresponding Jaccard index (averaged over all relevant CT slices) and the standard deviation (std) are shown.

Tissue	Q_{min}	$Q = 64$		$Q = 128$	
		Jaccard	std	Jaccard	std
Lungs	64	0.768	0.053	0.775	0.054
Sp. cord	64	0.647	0.005	0.652	0.005
Bone	64	0.775	0.006	0.784	0.005
Heart	128	0.885	0.007	0.911	0.004
Liver	128	0.837	0.004	0.876	0.004
Kidneys	128	0.756	0.015	0.815	0.005

Table 3: Evaluation of the SRM segmentations using the H_t metric. Q_{min} value is the smallest Q value for optimal region separation. Segmentation of the six organs/tissues using the SRM method with a Q value of 64 and 128 are reported. The H_t metric for t has the value of 1 and 2 (pixels) is reported.

Tissue	Q_{min}	$Q = 64$		$Q = 128$	
		H_1	H_2	H_1	H_2
Lungs	64	0.890	0.912	0.894	0.913
Spinal cord	64	0.876	0.974	0.879	0.980
Bone	64	0.956	0.979	0.963	0.982
Heart	128	0.813	0.902	0.861	0.944
Liver	128	0.746	0.827	0.825	0.910
Kidneys	128	0.750	0.837	0.810	0.893

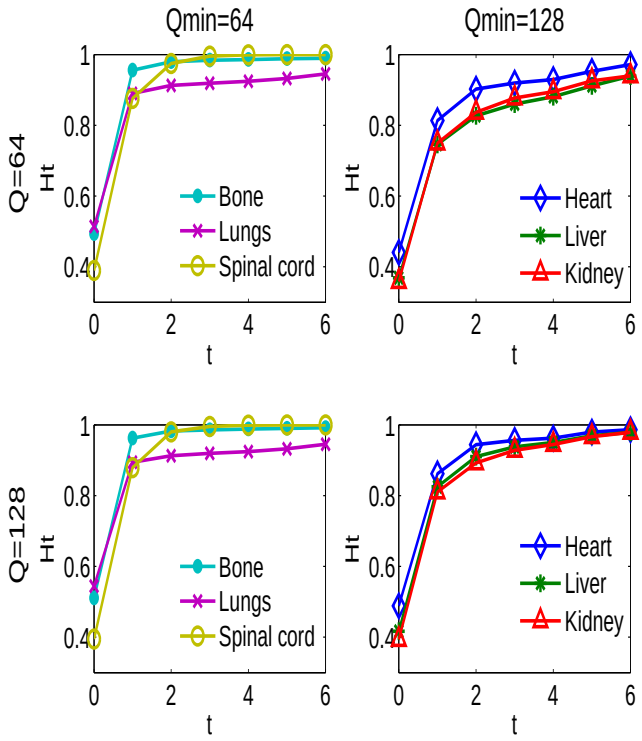


Figure 4: Evaluation of the SRM segmentations using the H_t metric. Organs/tissues depicted in the first column have a Q_{min} of 64 and those in the second column 128. Plots in the first row were obtained with an experimental Q value of 64 and those in the second row have $Q = 128$.

mentations with experimental Q value of 64 (Figure 4, first row) and 128 (Figure 4, second row) are reported. Though the H_t metric and the Jaccard index evaluate different characteristics of the SRM segmentation results, the same trend can be observed over the range of the experimental Q values. Figure 4, first column shows the H_t results for the group with a Q_{min} of 64. Again, as the SRM segmentation has reached the optimal point at Q value of 64, further increase in the Q value gives no significant improvement in the segmentation. Hence the H_t line plot for individual organ stays roughly the same with the Q value of 64 and 128. For the group with a Q_{min} of 128, using an experimental Q value lower than the Q_{min} will see a drop in the H_t value. Hence, H_t line plot for an individual organ/tissue is lower in the top panel than in the bottom panel in the second column in Figure 4. Table 3 highlights the values H_1 and H_2 in Figure 4. The H_1 and H_2 values are of particular interest as they measure the accuracy of the segmentation within 1 and 2 pixels, which can be considered as an acceptable tolerance error for the manual expert segmentation.

4 Discussion

In two CT slice images the SRM method did not segment the spinal cord. Figure 2(f) shows ones of those slice images. The expert segmentation Figure 2(d) clearly delineates the tissue based on the knowledge of human anatomy but the method, being entirely intensity-based, cannot distinguish the object from the surrounding muscle/soft tissue. The two images were not included in the evaluation of the spinal cord segmentation.

Furthermore, the SRM segments the images in a CT examination slice by slice individually. The SRM region labels used in two different images are unrelated. This means that the SRM region label assigned to, say, the liver in one CT slice is different from that assigned to the liver in the neighbouring CT slice. Segmented regions of the same organ appeared in different slice images need to have a unique label. This can be achieved by implementing algorithms identifying the connectedness of segmented regions in neighbouring images. In this study, the connectedness is identified by the user.

5 Conclusion

The study shows that Statistical Region Merging can be tuned to automatically segment CT images with a great level of accuracy and robustness. Using six different representative tissues it was shown that the accuracy of an expert segmentation can be well matched. The result is particularly promising since no pre or post processing was incorporated into the process. Neither filtering nor post segmentation merging of regions was performed. The addition of those steps will certainly improve the outcome for a specific application at hand. It is expected that the obtained approximate contours will be subject of further processing using one of the numerous region-merging algorithms available in literature (e.g. [17] or [18], [19]) and active contour models (e.g. [20], [21]) to futher improve their accuracy. To keep the results applicable to a wide range of CT images those steps were not performed in this study.

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 and A. Wittman. The adult male voxel model "golem" segmented from whole-body CT patient data. *Radiation and Environmental Biophysics*, 40:153–162, 2001.

[3] M. Caon, G. Bibbo, and J. Pattison. Monte Carlo calculated effective dose to teenage girls from computed tomography examinations. *Radiation Protection Dosimetry*, 90(4):445–448, 2000.

[4] M. Zankl, R. Veit, G. Williams, K. Schneider, H. Fendel, N. Petoussi, and G. Drexler. The construc-

tion of computer tomographic phantoms and their application in radiology and radiation protection. *Radiation and Environmental Biophysics*, 27:153–164, 1988.

[5] 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.

[6] T. Nagoka, S. Watanabe, K. Sakurai, E. Kunieda, S. Watanabe, M. Taki, and W. Yamanaka. Development of realistic high resolution whole-body voxel models of japanese adult males and females of average height and weight, and application of models to radio-hyphen frequency electromagnetic-field dosimetry. *Physics in Medicine and Biology*, 49:1–15, 2004.

[7] U. Fill, M. Zankl, N. Petoussi-Henss, M. Siebert, and D. Regulla. Adult female voxel models of different stature and photon conversion coefficients for radiation protection. *Health Physics*, 86(3):253–272, 2004.

[8] R. Nock and F. Nielsen. Statistical Region Merging. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(11):1452–1458, 2004.

[9] Q. Wang and Z. Wang. A subjective method for image segmentation evaluation. In H. Zha, R. Taniguchi, and S. Maybank, editors, *Proceedings of ACCV 2009*, *LNCS*, volume 5996, pages 53–64, 2010.

[10] M. Emre Celebi, Hassan A. Kingravi, Jeongkyu Lee, William Van Stoecker, Joseph M. Malters, Hitoshi Iyatomi, Y. Alp Aslandogan, Randy Moss, and Ashfaq A. Marghoob. Fast and accurate border detection in dermoscopy images using Statistical Region Merging. In J. P. W. Pluim and J. M. Reinhardt, editors, *Proceedings of SPIE: Medical Imaging 2007*, volume 6512, pages 65123V–1, 2007.

[11] M. Zortea, S. O. Skrovseth, T. R. Schopf, H. M. Kirchesch, and F. Godtliebsen. Automatic segmentation of dermoscopic images by iterative classification. *International Journal of Biomedical Imaging*, 2011.

[12] M. Bajger, F. Ma, S. Williams, and M. Bottema. Mammographic mass detection with Statistical Region Merging. In *Proceedings of DICTA 2010: Digital Image Computing: Techniques and Applications*, pages 27–32, Sydney, Australia, 2010.

[13] 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.

[14] ICRP. The 2007 recommendations of the International Commission on Radiological Protection. *Annals of the ICRP*, 103, 2007.

[15] P. Jaccard. Nouvelles recherches sur la distribution florale. *Bull. Soc. Vaudoise Sci. Natl.*, 44:223–270, 1908.

[16] 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 *Proceedings of the Joint Disease Workshop, 9th International Conference on Medical Image Computing and Compter Assisted Intervention*, volume 1, pages 72–79, 2006.

[17] M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis, and Machine Vision*. Thomson, 2008.

[18] Kostas Haris, Serafim N. Efstratiadis, Nicos Maglaveras, and Aggelos K. Katsaggelos. Hybrid image segmentation using watersheds and fast region merging. *IEEE transactions on Image Processing*, 7:1684–1699, 1998.

[19] J. Cousty, M. Couprie, Najman. L., and G. Bertrand. Weighted fusion graphs: Merging properties and watersheds. *Discrete Applied Mathematics*, 156:3011–3027, 2008.

[20] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: active contour models. *Int. J. Comput. Vision*, 1(4):321–331, 1987.

[21] C. Xu and J. L. Prince. Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing*, 7(3):359–369, 1998.