Evaluation of Statistical Region Merging Segmentation for heart and spleen contour delineation in CT images

Master Thesis

By

Sami Meshal Almutairi

2109842

Supervisor: Dr.Mariusz Bajger

School of Computer Science and Engineering and Mathematics

Faculty of Science and Engineering

July 2016

"Submitted to the School of computer Science, Engineering, and Mathematics in Faculty of Science and Engineering in partial fulfillment of the requirements for the degree of Master degree at Flinders University- Adelaide Australia"

DISCLAIMER

I certify that this work does not incorporate without acknowledgment any material previously submitted for degree and diploma in any university; and that the beat of my knowledge and belief it does not contain any material previously published or written by another person expect where due reference is made in the text.

Signature  Data

14 feb 2017

ABSTRACT

Segmentation is the key step towards extracting information from medical images for further analysis. This study investigates the accuracy of organ segmentation from CT images using the Statistical Region Merging (SRM) segmentation technique. The experiments were conducted using data set of 12 consecutive CT slices for heart and 8 CT slices for spleen. The data sets were previously manually segmented by an expert in human anatomy and these sets constituted the ground truth to judge the accuracy of an automatic segmentation using SRM. The results of SRM segmentation technique depend on a parameter (Q) which value determines granularity of the outcome. A wide range of Q values was investigated in relation to the accuracy of segmented heart and spleen. The accuracy was measured using the Dice and Jaccard indexes. The results show that the SRM segmentation technique can potentially provide high accuracy segmentation of 80% and above for both heart and spleen. However, neither heart nor spleen can be segmented at this level of precision as a single component and an additional post-processing is needed to merge eligible regions into one piece corresponding to the whole organ.

ACKNOWLEDGMENTS

I would like to acknowledgment my research supervisor Dr. Mariusz Bajger for his assistance and dedicated involvement in every step throughout the process.

I would like also to thank my family who support me in finalizing this project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1. Introduction

Computed tomography (CT) scan is a new generation technological method in which beams of X-rays are passed to gather structural and systematic information of the human body for medical diagnosis. The CT image is developed through the absorption profile of the X-rays. CT is now a popular process for imaging and diagnosing problems with crucial body parts such as the brain, liver, heart and spleen (Prince and Links 2006). Image segmentation is an important process for splitting images based on similarities between the image components, such as the intensity of colour, sharpness and texture (Gonzalez and Woods 2004; Pratt 2001; Pal and Pal 1993). Virtually, in the segmentation process, images are divided in to objects. Image segmentation can be performed for several reasons including studying the anatomy of the human body; identifying regions of interest (ROI) such as locating tumours, lesions and other abnormal changes; measuring the volume of tissue in determining tumour growth; assisting in earlier treatment planning for radiation therapy; and calculating radiation doses (Sharma and Aggarwal 2010). It has been recognised that segmentation is a key issue in CT scan imaging.

To provide a well-balanced segmentation process, a large number of methods, such as amplitude segmentation-, edge- and region-based approaches, have been used in the field of CT scan imaging. Advancements in technology have prompted the development of automatic segmentation of CT scan images. Accordingly, researchers have developed various algorithms to simplify the imaging process while increasing its accuracy. Among the different algorithms, region-based segmentation methods have emerged, which are built on the principle of homogeneity, suggesting that similar characteristics of pixels can be used to cluster homogeneous region. Three types of region-based segmentation include merging regions, splitting, and splitting and merging. In the process of merging regions, one can choose seeding points to start the process and the result depends on the choice of seed. Furthermore, the regions can be extended by merging neighbouring pixels in accordance with the merging criteria. The merging process will continue until all of the pixels are clustered in their respective regions according to the merging conditions (Sharma and Aggarwal 2010).

Texture is an important phenomenon for the region-based merging process, which can be defined as a composition of mutually related elements (Sonka, Hlavac and Boyle 1999). A texture might have different structural and tone-based features, which express the fineness

and coarseness of the texture. The intensity properties of pixels relate to tone; whereas, structure refers to the relationship between the pixels and is space oriented. Three different types of approaches are used to extract texture features: statistical, structure-based and spectral-based. By applying statistical extraction to the features, which may represent a complex space vector, texture features can be determined. The level of statistics of the feature might be classified using orders such as first, second and higher depending on the image greyness (Sharma and Aggarwal 2010; Haralick 1979; Julsez 1981; Fukunaga 1990). As an emerging technique in the CT scan process, this study evaluates Statistical Region Merging (SRM) segmentation for delineating contours of CT images for the heart and spleen.

1.1 Background of the project

Several studies (Tesar et al., 2008; Prasad et al., 2008; Koss et al., 1999) have found that the individual segmentation methods have some specific limitations. For example, the spectral frequency-based approach is less efficient than other approaches, while one of the key limitations of the model-based method is that it requires manual interaction to initiate a model and appropriate parameters need to be selected. Another drawback of the model-based method is its poor capacity to exhibit convergence to concave boundaries (Pham, Xu and Prince 2000). Although the atlas-based segmentation approach has a third generation algorithm, its performance is limited by factors including the registration procedure and the manual tracing protocol. In addition, one may find it very challenging to segment complex structures in respect to the variability of their size, shape and properties. The atlas-based segmentation method also requires expert knowledge to build the database (Wolz et al. 2013; Ding, Leow and Wang n.d.). It has been observed that the statistical approach is useful for random patterns or textures. In particular, in SRM segmentation, each pixel has a unique region at the beginning, where the merging of adjacent pixels can occur in the neighbouring areas. A labelled image is created by merging segments where no label exists to create a connected region with other regions (Dvorak, Bartusek and Gescheidtova 2014). During the process of the SRM method, it is important to conduct an experiment to measure its performance.

1.2 Significance of the project

The SRM technique originated from the work of Nock and Nielsen in 2004. The image segmentation algorithm is often validated by comparing the automated image with the manually segmented image. In this process, manually segmented images are resized to 126×128 pixels so that the automatically generated segment can work on the resized images using three dimensional (3D) SRM and 3D EGS. Effective dose calculation needs to identify body tissue including the key organs such as heart, liver, stomach, spleen and spinal cord. Experiments have shown that the 3D SRM and 3D EGS approaches are feasible and effective (Bajger, Lee and Caon 2012). CT is related to a high dose radiological process, which might be harmful for patients. To reduce the harmful effects, it is essential to estimate the accurate dose for an individual, which is often difficult to obtain. An anatomical model is required to calculate effective doses for CT procedures involving human organs (Caon, Bibbo and Pattison 1999). A human anatomy model is developed from the cross-sectional medical images that originate from medical imaging. In this context, the results of this research project will provide evidence on applicability the SRM technique in the segmentation of two selected organs: the heart and the spleen.

1.3 Aim

The aim of this project is to evaluate the SRM segmentation technique for the heart and the spleen in CT images and find the optimal value of the parameter Q (see Section 2.2.2.1) yielding with a high level of accuracy and minimum eligible components.

CHAPTER 2: LITERATURE REVIEW

2. Literature review

Before starting the literature review, the problem questions have been formulated in line with the aim of the project. The key problem questions help to determine the research gaps, and the deliverables will be conveyed in accordance with those research gaps.

2.1 Problem questions

The following problem questions aim to establish the existing research gaps:

(1) What segmentation methods are used in the CT scan imaging process and how effective are they for the growing demand of CT scans?
(2) What constraints limit the performance of the existing segmentation processes used in CT scans?
(3) How effective and efficient is the performance of SRM segmentation for human organs in the context of CT scans?

Based on the above research questions, a reasonable number of relevant and recent articles have been reviewed to obtain insights into the segmentation processes and their relevant concepts.

2.2 Existing segmentation techniques in CT imaging

The existing literature describes and discusses a range of segmentation techniques which have different inherent limitations. Image segmentation is used for analysing the image and pattern recognition (LaValle and Hutchinson 1993). From a semantic viewpoint, image segmentation represents objects rather than a set of pixels. From a practical viewpoint, image segmentation provides the scope by which to estimate robust parameters and descriptors. In brief, image segmentation is used to present the images in a semantically meaningful way, making them easier to analyse (Ballard and Brown 1982).

Srinivasan and Shobha (2007) reviewed the segmentation process from existing literature and expressed that image segmentation is a primary step to analyse the images from the object detection, appreciation and identification. The segmentation method allows digital images to be partitioned into a number of regions. Finally, the entire image can be covered or a set of counters can be extracted through the image segmentation process. Pixels from a particular region have unique characteristics or computed properties including colour, intensity or texture. However, pixels belonging to adjacent regions have largely different properties. Srinivasan and Shobha (2007) described different types of techniques and methods in their review including edge detection, region splitting, growing methods and multi-scale segmentation.

Furthermore, image segmentation is important for dividing images into homogenous regions, which can be used for high level image processing including analysis, pattern recognition, object classification and change detection (Li et al. 2009). In addition, multi-spectral and multi-scale images are essential for conducting analyses because they contain all kinds of information and object characteristics that are required. Two key types of segmentation techniques were prominent in the existing literature: the atlas-based and SRM approaches.

2.2.1 Atlas-based approach

2.2.1.1 General principles

The atlas-based approach is a popular method of segmenting complex CT and MR images. It combines closed contours of the entire human body, liver, stomach and spleen, which have been segmented from a reference CT image (Fleckenstein and Jensen 1993).

Two types of atlases are commonly used to segment images, such as 3D and 2D atlases. While 3D atlases are comparatively complex in nature as they are a combination of multiple parameters, 2D atlases are easily controllable because of their simplicity. In 3D CT/MR images, segmentation of the surfaces of the anatomical structures are performed using 3D atlases (Thurfjell et al. 1993; Dawant et al. 1999; Hartmann et al. 1999; Rueckert et al. 2002; Cuadra et al. 2004; Lorenzo-Vald'es et al. 2004); thus, individual anatomical structures need

individual 3D atlas-based algorithms. In other words, the algorithms must be anatomically structure specific. In general, atlas-based registration techniques are used for processing atlas-based segmentation. Furthermore, the boundaries of the segmented anatomical structures are determined from the registered atlas contours (in the case of 2D) and surfaces (in the case of 3D).

The atlas-based segmentation algorithm has two key steps: global transformation and local transformation. Global transformation aligns the algorithm with the target image, and the local transformation or deformation of the atlas accurately registers the algorithm with the corresponding image features (Ding, Leow and Wang n.d.).

2.2.1.2 Application of the atlas approach in the CT images

Ding, Leow and Wang (n.d.) conducted an experiment for segmenting abdominal CT volumes into multiple slices utilising a single 2D atlas approach. The results of the test showed that their algorithm was successful and accurate in segmenting abdominal CT images into 34 CT slices of 1 mm thickness. The experiment outcomes indicate the flexibility of the developed algorithms, making them compatible for segmenting CT images for different patients. This study contributed to solving the problems of segmenting images of the human body.

In 2D atlas-based method, good initialisation is needed to ensure the local deformation avoids affecting the control system by extracting the wrong subject boundaries. However, to compute the optimal transformation parameters, an iterative optimization algorithm needs to be applied.

Linguraru et al. (2010) investigated normalised probability-based atlases in segmenting and quantifying livers and spleens to extract imaging biomarkers with volume and height. The study developed a clinical tool to segment the liver and spleen in relation to normal and pathological contexts, where the volume and height errors were 2.2% and 2.8%, respectively. This segmentation method a combined statistics related to the enhancement, shape and location of the liver and the spleen. The authors found in the segmentation process that the process of initialising the models is simultaneous, whereas the segmentation process is

sequential. Linguraru et al. (2010) concluded that automated computer-assisted tools, which can generate the liver and spleen biomarkers, have the capacity to assist in detecting abdominal disorders, which are independent of routine clinical data analysis.
Wolz et al. (2013) introduced a method of segmentation combining aspects from multi-atlas registration and patch-based segmentation. Their findings showed that the accuracy level was 92% for the spleen. Another probabilistic atlas for modelling organ appearances such as the liver, kidney and spinal cord was proposed by Park, Bland and Meyer et al. (2003). It showed the accuracy of the improved segmentation method using an unsupervised approach. Incorporating spatial priori knowledge, atlases were applied for different abdominal organs. In addition, Okada et al. (2012, pp. 173–180) applied atlases by combining inter-organ spatial relationships to form a model shape for performing multi-organ segmentation. Wang et al. (2012) developed organ segmentation in mice using the probabilistic atlas model. This multi-atlas registration model showed a high variability in the global appearance of the brain. However, the main advantage of this approach was that it does not need any specialisation and prior parameter settings.

Although many studies (e.g. Isgum et al. 2009; Commowick and Malandain 2007; Han et al. 2008; Ramus, Commowick and Malandain 2010; Rikxoort et al. 2010; Artaechevarria, Munoz-Barrutia and Solorzano 2009) have employed strategies to overcome the challenges of the multi-atlas segmentation approach, some limitations remain, including issues pertaining to accuracy.

2.2.2 SRM approach

2.2.2.1 Principles of SRM

The SRM technique is an algorithm used for segmenting medical imaging to estimate the values with a region and group these values based on the pre-established merging criteria (Calderero and Marques, 2010).

Nock and Nielsen (2004) undertook a detailed study that offers theoretical and empirical evidence on the SRM technique and its applicability to medical imaging. The model

developed by Nock and Nielsen (2004) is a prototype of an image I, which is an observation of a statistical (true) scene denoted by I*. Further, Nock and Nielsen (2004) explained that the statistical region denoted by I* is uncertain, or rather, unknown, and that it is the subject of extraction from the given image I. Their study also postulates that the pixels in observations I* are represented by set of distributions Q, from which the colour levels in I are sampled. This implies that the intensity of each of the pixels in I can be derived by summing up all the Q independent random variables, and in a scenario where a grey-level image exists, these Q values would lie with in the interval of [0, 256/Q] (Nock and Nielsen 2004).

Nock and Nielsen (2004) noted that the sets of Q distributions are subject to variation when analysed between statistical pixels and therefore hypothesised that the statistical regions in I* constitute the four connected segments. The study also hypothesised that the expected value of intensity is similar to elements that belong to a similar true/statistical region, which varies between adjacent regions that are statistically dissimilar. Thus, the Q values can be treated as a measure of the statistical complexity of I*. However, this implies that when used in theoretical models that are more general, smaller Q parameters would yield pixels that are difficult to segment. Accordingly, Nock and Nielsen (2004) reached the assumption that small Q values would more likely yield under-segmentation, while high Q values would lead to over-segmentation. Given these results, it is advisable to pick Q values in a manner that allows objects of interest to be well segmented.

The SRM algorithm developed by Nock and Nelson (2004) begins by rearranging the pixels in increasing order, from the lowest to the highest pixel pairs *p* and *p'* of I based on the function *f(p,p')*; *f(p,p')* denotes the variation of the intensity values in *p* and *p'*. Further, the study hypothesised that if the object has single pixels, then the regions R and R' can be merged if the following equation holds:

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

where

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

where, |R| represents the number of pixels within region R, $0 < \partial \leq 1$, and Ŕ represents the average intensity with region R and g in the number of image intensity levels.

In equation (1), the predicate (I) is based on the postulation that the two regions denoted by R and Ŕ should merge if they are by-products of similar statistical (true=) regions: if
$\bar{E}$ (Ŕ – Ŕ') = 0, where $\bar{E}$(Ŕ) is the expectation over all neighbouring statistical pixels of I* of their total Q random variables' expectations based on their intensity values.
The SRM technique is a member of the family of the region growing techniques that use statistic-based tests for fusing regions. In SRM, grouping is considered an inferential problem, where original imagery creates observation imagery by applying sampling techniques. Simultaneously, the segmentation imagery can be developed through image observation and the regeneration of the homogeneous boundary of region. Multi-scale segmentation algorithms for higher resolution (HR) imagery include region growing methods, where pixels can be taken as seeds and the growing region can be considered based on certain homogeneity criteria.

Calderero (2010) described other SRM techniques based on the principles of the parametric probability region model, which may show properties of colour homogeneity (LaValle and Hutchinson 1993; Gies and Bernard 2004) or texture assumptions. Calderero (2010) explained the generality, accuracy and efficiency of the region models and concluded that under-segmentation error is more important than over-segmentation error. To overcome under-segmentation error, Calderero (2010) proposed the first-order Markov model, which is a slightly more complex statistical model. To reduce over-segmentation error, he suggested using area-unweighted criteria to combine the scale-oriented merging order (Calderero 2010).

The segmentation process of the SRM technique comprises three steps: pre-processing, segmentation and segmentation merging. The general pre-processing algorithms can be used in MRI processing steps. Subsequent segmentation steps are conducted by applying the SRM algorithm. During the segmentation, particular tissues in several locations are merged to the regions. The pre-processing stage examines tissues such as the definition of the ROI, reduction of noise and compensation of inhomogeneity. Q parameters measure the statistical complexity of the image, the model generality and the statistical task hardness; the smaller the size of the segments, the larger the Q values. The advantage of the SRM technique is that boundary detection is possible in the areas where the clear borders of the active contours are invisible.

Li et al. (2009) emphasised that the SRM segmentation is constrained by some factors such as the border object and border shape. They also observed that the SRM method has the ability to (1) provide robust and accurate segmentation regarding spectral and shape information, (2) adapt to noise and corruption, and (3) handle occlusions (Li et al. 2009). However, one limitation of SRM, which applies the minimum heterogeneity rule (MHR), is that the similarity of radiometric and semantic forces the users to set non-intuitive parameters including the scale, weight and stop criterion. Li et al. (2009) also mentioned some investigation issues such as required improvements to the sort function, the merge predicate, the evaluation index study to estimate the results of the estimation, parameter determination for various classes and comparisons with other segmentation techniques.

Lee, Bajger and Caon (2012) observed that segmentation is a key step in developing anatomical models that help to calculate medical radiation doses for children. Their study the SRM segmentation technique with respect to tissue segmentation in CT images and introduced a criterion, which allows the method to be tuned automatically. Evidence suggests that the SRM has a great potential for CT image segmentation with a high level of overall boundary precision (93.7%).

This study finds that developing a range of anatomical models faces two key challenges: the unavailability of an ample data set of medical images with ground truth over label and the significant time requirement for segmenting the entire internal anatomy form their corresponding images. Composite models can be developed by joining the head, torso and limbs from different individuals (Lee, Bajger and Caon 2012).

Lee, Bajger and Caon (2012) also found that the SRM technique is robust and offers a great level of accuracy in the automatic segmentation of CT images. The study demonstrated this finding by selecting six different representative tissues such as the lungs, heart, kidneys, liver, spinal cord and bone. The optimal region was found to be either 64 or 128 due to a minimum Q value required. This study suggests that active contour models (Kass, Witkin & Terzopoulos 1987; Xu & Prince 1998) can improve the accuracy of the SRM technique in CT images.

It is observed that parameter Q is a factor of the statistical complexity of the image. As a high Q value results in under merging, the expectation is to keep the Q value as small as possible, which is the requirement of region separation (Bajger, Lee and Caon 2012).

Finally, Dvorak, Bartusek and Gescheidtova (2014) stressed the need to improve the segment merging process, the determination of the tissue and the 3D extension. The key advantage compared to other emerging methods is that the method is independent of the number of segmented tissues.

2.2.2.2 Statistic region merging as a technique on detection

1. Application of SRM on mammography

Bajger et al. (2010) used the SRM technique to detect mammographic masses using an automatic method. Images were taken from both the local database mammograms and the Digital Database for Screening Mammography (DDSM). The speciality of the SRM technique is that the segmentation process is considered as an inference problem. The SRM technique provides some advantages in that no prior statistical model is needed for diseased or normal tissues. The only required parameter is the potential to control the complexity of the segmentation process, and the result of the segmentation is close to optimal, although an over-merging error exists (Nock and Nielsen 2004).

To overcome the issue of over merging in the SRM technique, Nock and Nielson (2004) introduced a more sophisticated version to create the scene of the images naturally. Experiments carried out by Bajger et al. (2010) show that the original version (2) works better. The simplicity of version (2) is in its simple results that give an insightful connection between the statistical image complexity and the size of segmented objects. Furthermore, it has flexibility, as the object size can vary from 110 to 140 pixels with the spatial value image resolution. Bajger et al. (2010) concluded that this finding was true for any datasets whether they are obtained from the DDSM or the Local Database of Mammograms, as the spatial image resolution of the datasets was very similar. Nock and Nielsen (2004) endorsed that the SRM technique is superior to EGS for handling noise in the natural scenes of images. This

study also revealed the differences between how SRM and EGS relate to mammogram segmentation (Ma et al. 2007; Bajger, Ma and Bottema 2009).

This study confirms the applicability of the SRM technique on mammograms. The experiment also found that the method produces accurate regions, which are sufficient for using seed components to refine the other region or other methods of merging. It has also been observed that mass regions are detected successfully by applying the features of lesions. The evidence from this study also demonstrates the improvement of the reports created by Bajger, Ma and Bottema (2009) and Ma, Bajger and Bottema (2009). Anyone can apply analytical judgement regarding the segment size based on the image characteristics. The findings of this study also confirm that determining the segmentation parameters does not affect the important lesions that are lost in the process because of over merging and suggests that an efficient process can make the SRM segmentation technique more sophisticated for detecting and analysing the image regions.

Several studies have tested the SRM method (Nock and Nielsen 2004; Felzenszwalb and Huttenlocher 2004) and found that it significantly better handles noise. This technique is based on the probabilistic concentration theory, which is predictable prior to any experiments.

2. Application of SRM technique on minimum heterogeneity rule MHR

Li et al. (2009) used SRM successfully for object merging using the principles of initial segmentation and MHR. In this experiment, Li et al. (2009) applied the QuickBird imageries approach in demonstrating the SRMMHR segmentation method. The SRM method includes spectral, shape and scale information, an ability to adapt to noise and corruption, and handles occlusions well. This integrated SRM approach also has the ability to eliminate redundant objects and avoid the limitations of the Fractal Net Evolution Approach (FENA). The authors concluded that the SRMMHR technique provides accurate, efficient and high resolution imagery with a multi-scale segmentation approach (Li et al. 2009).

3. Application of SRM on dermoscopy

As skin-imaging technology is advancing, there is a growing demand for computer-based diagnoses of melanoma. Celebi et al. (2007) introduced a quick and unsupervised method to detect borders of skin lesions with pigments by applying the SRM algorithm to dermoscopy images. Their study tested 90 dermoscopy images and showed that it is possible to obtain fast and accurate border detection. The SRM algorithm is flexible and adaptable for testing any dermoscopy images because it is simple, efficient and well-performed on a wide range of image domains. A preselected set of borders is used as the ground-truth. A metric developed by Hance et al. (1996) can be used to quantify the border detection error.

Furthermore, Celebi et al. (2008) expressed that SRM is a colour image segmentation technique, which is based on the principles of region growing and merging. The SRM model originated from an unknown theoretical image in which statistic-based regions needed reconstruction. This process was simple, efficient and independent of quantization or colour space transformation. The pre-processing phase entails removing the black frame and smoothing the images. The segmentation phase involved pixel couple ordering and region merging. The post-processing phase included eliminating the background skin regions, removing isolated regions and expanding the initial border through morphological dilation to obtain an accurate result.

The SRM technique is applied to detect skin cancer from images of skin lesions. Lau and Al-Jumaily (2009) observed that the SRM process is a combination of growing and merging regions. SRM works on the principle of statistical tests in relation to region merging. Lau and Al-Jumaily (2009) described SRM as a more accurate process of segmentation. SRM has the potential to include a more useful area, and its outer region also forms a part of the ROI. However, SRM has one key limitation; it is less accurate than neural networks. A classification test revealed that the threshold of a back-propagation neural network (BNN) is 0.5% more accurate than SRM.

4. Application of SRM on MRI

Dvorak, Bartusek and Gescheidtova (2014) conducted an experiment to develop fully automatic segmentation of multi-contrast MRI images using the SRM technique. They developed an algorithm based on Nock and Nielsen's approach (2004). The obtained results demonstrated the effectiveness of SRM at segmenting multi-contrast MRI images. The findings confirm that the method is independent of the number of segmented tissues and the developed algorithm is flexible for any parts of body and including healthy.

2.2.2.3 Constraints of the SRM technique in different applications

Although SRM is a fairly new technique, Celebi et al. (2007) and Zortea et al. (2011) have already used it to diagnose melanoma, detect breast cancer and examine post-operative outcomes of implementing a knee prosthesis (Celebi et al. 2007; Zortea et al. 2011; Bajger et al. 2010; Battiato et al. 2006). When using this method, care needs to be taken when merging various tissues with other regions in a small-segmented region to avoid tissue loss. Selecting the appropriate Q value is important in this regard. The minimum size of the regions can be varied from 25 to 50 pixels: 25 for small organs such as the spinal cord and 50 for larger organs such as kidneys or heart. The intensity difference between all organs can be no more than 15.

The above SRM applications indicate that over merging is only one source of error. It was also found that multiscale segmentation ensures a hierarchy of segmentations at different scales, as scale is controlled by the Q value. Therefore, a smaller Q parameter increases the difficulty of the statistical estimation task and decreases the regions in the final segmentation.

2.3 Summary of the review

The findings from the above literature review are clear-cut and straight-forward. The SRM technique has the potential to sequent a mammography, dermoscopy, MRI and SRMMHR images, while maintaining accuracy and simplicity. An analysis of the review in line with the research questions justifies conducting a study to evaluate the SRM approach in the

segmentation of the heart and the spleen. To fill the research gap, this research project evaluates the effectiveness and suitability of the SRM technique for segmentation of the heart and spleen in CT scan images. This project also identifies the limitations of the SRM technique.

CHAPTER 3: METHODOLOGY

3. Materials

3.1 Data set

This project used the 12 CT images of the heart and 8 CT images of the spleen from a female patient aged 14 years and weighing approximately 14kg. These images were taken from the hospital's archive with the patient's identification data removed. The images had a pixel size of 2.53 X 2.43 mm and a slice separation of 10 mm, and the field of view of radius was 145 mm from the scanner.

3.2 Ground truth

The ground truth images consist of organs that were segmented manually by an expert in human anatomy. The size of the ground truth images was 126 X 128 pixels. Figure 1 shows an example of an expert-segmented image.

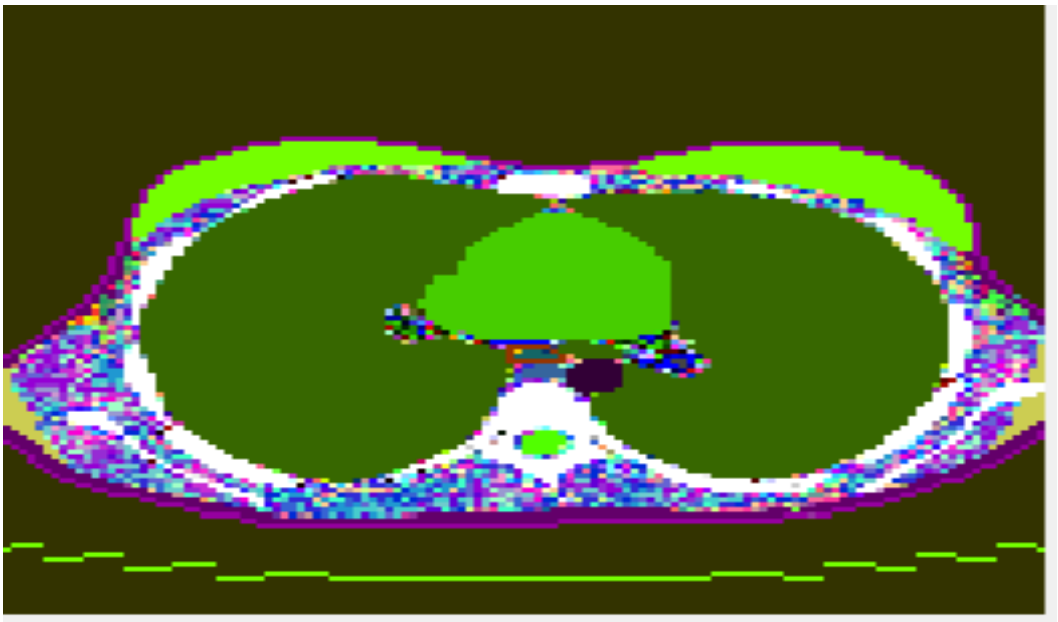

Figure 1: An example of abdominal part, an expert-segmented image.

3.3 Unions of intersections

The original images were automatically segmented using SRM with Q values ranging from 8 to 512 in steps of 8 (i.e. Q={8, 16, 24,… 512}). The output is markedly different for each Q value. For example, Figure 2 shows the segmented images for Q = 16, Q = 64 and Q = 256.

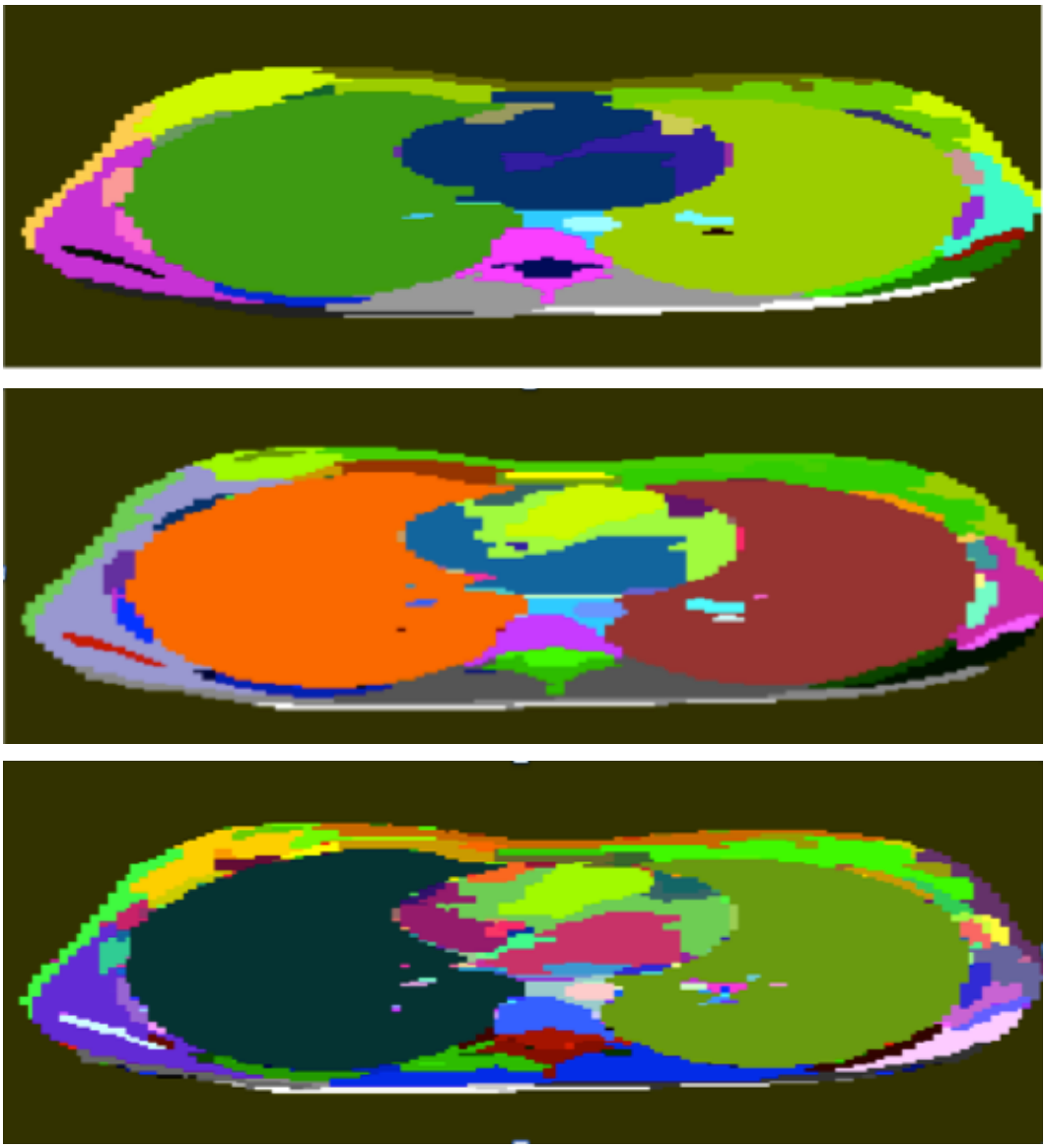

Figure 2: Segmented images of different colour using the SRM technique with different Q values: Q = 16, Q = 64 and Q = 256.

The automatic segmentation image is a labelled image. Each label in the automatic segmented image corresponds to the ROI. The overlap between the ROI and the expert-

segmented image contains identical eligible components. If the common region overlaps more than 50% of the segmented ROI (from the automatic segmented image), it is considered an eligible component, and the whole ROI is added (by binary union operation). The process is repeated for every label, and the overlap of all the labelled regions (which exceeds a threshold of 50% of the labelled region) are combined as a union of sets. Figure 3 shows five labelled components. Components 1, 2 and 3 are determined as eligible components, while components 4 and 5 are determined as ineligible components.

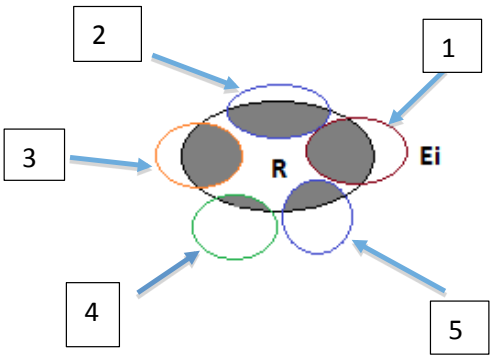

Figure 3: An example of the selection process of the eligible components.

Let E_i^R be the set of all eligible components and N be the number of eligible components. Then, the union image is formed by the union of all the eligible components as

N_R: total number of eligible compounds for the ROI of the annotated region R.

$$u_R = \cup_{i=1}^{N_R} E_i^R \,. \tag{2}$$

In other words, if the number of pixels in the overlap between the labelled ROI (from the automatic segmented image) and the annotated regions (organ tissue region in the expert-segmented image) is larger than the non-overlap regions, the labelled ROI is used to construct the organ tissue in the output image by forming a union of the overlap regions for different labels. Finally, the union image is intersected with the annotated expert organ regions to produce the final RIO corresponding to the organ,

$$S = R \cap u_R. \tag{3}$$

3.4 Evaluations

3.4.1 Dice index

To evaluate the performance of the union of the intersection technique and the objective validity of the results obtained, the Dice index is used. The Dice index is a statistical test, which is used for measuring the overlap between the expert-segmented image and the union of an intersections. The Dice index between two regions A and B is defined as

$$D(A,B) = \frac{2\,|A \cap B|}{|A| + |B|} \tag{4}$$

where $|A| + |B|$ is the total number of the pixels in the set A.

Equation (4) shows that the Dice index is the common/overlapped region divided by the sum of the individual components and results in a value between 0 and 1. The Dice index will result in zero if there is no overlap between the two regions and one if one region completely overlaps the other region identically.

3-4-2 Jaccard index

The Jaccard index measurement can also be used for comparing the similarities between the expert-segmented image and the union of an intersection image. The Jaccard index between two regions A and B is defined as

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{5}$$

Equation (5) shows that the Jaccard index is the common/overlapped region divided by the union of the individual components and results in a value between 0 and 1. The Jaccard index will result in zero if there is no overlap between the two regions and one if one region completely overlaps the other region.

CHAPTER 4: RESULTS

4. Results

This study performed two-dimensional segmentation of two organs – the heart and the spleen – in CT scans using SRM segmentation. The results of the SRM were evaluated against segmentation performed by an expert in human anatomy. The Dice and Jaccard indices (see Section 3.4) were used to evaluate the SRM segmentation technique. This chapter presents the results of the evaluation of the SRM segmentation technique for contour delineation of the heart and the spleen in CT images. Our goal is not segment the organs however, to investigate the potential of SRM technique towards object segmentation in CT images. Thus we investigate an optimal choice of SRM parameters would enable the best outcome. Ground truth is used to judge the performance of the technique for various choice of Q value while keeping the number eligible segments on Q reasonable level.

4.1 Heart

Figure 4 displays the means of the Dice and Jaccard index values for the heart. The SRM segmentation technique was performed on the heart with Q values ranging from 8 to 512 in steps of eight. As the heart organ appears in 12 slices, Figure 4 shows the means of the Dice and Jaccard index values for all related CT slices. Overall, the mean Jaccard index increases as the value of Q increases; however, there are some slight fluctuations and occasional dips for some Q values.

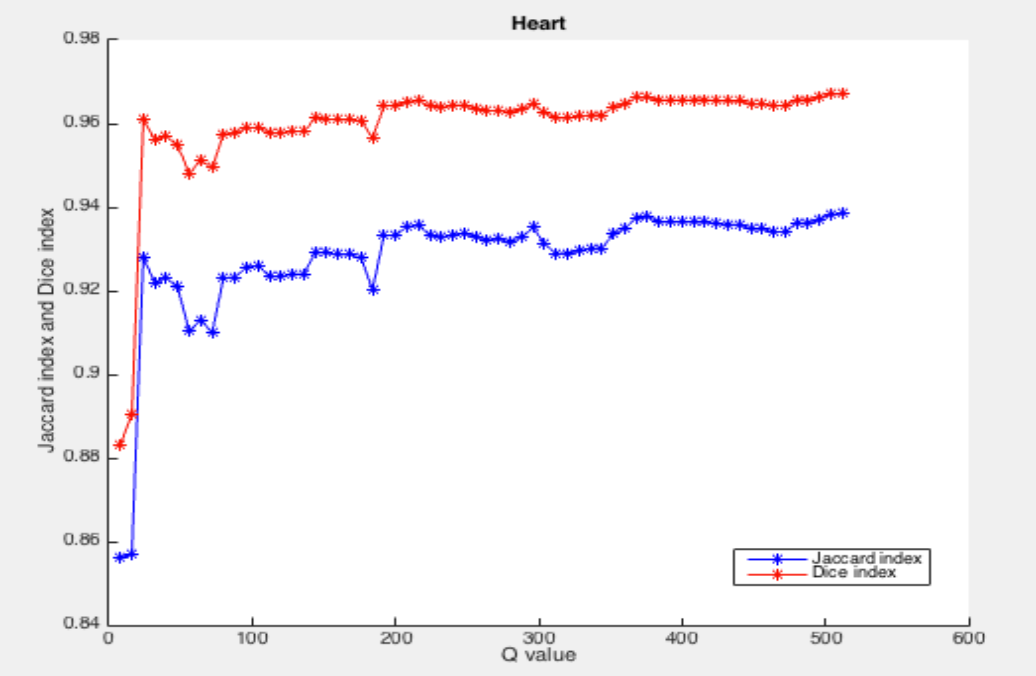

Figure 4: The mean Dice and Jaccard index values for 12 CT slices of the heart. SRM segmentation was performed for 64 Q values ranging from 8 to 512 in steps of eight.

Table 1 shows the mean Jaccard index for SRM segmentation with Q values of 8, 64, 128 and 512. This table shows the major differences between the Q values. Increasing the Q value from 8 to 64 results in a significant increase of 0.05 in the mean Jaccard index, thus indicating a higher initial rate of increase in the Jaccard index for smaller Q values and a lower rate of increase at high Q values.

Q values	**8**	**64**	**128**	**512**
Mean Jaccard index	0.85	0.90	0.92	0.93
Mean Dice index	0.88	0.95	0.95	0.96

Table 1: Evaluation of SRM for the heart using the mean Jaccard and Dice indices for different Q values.

Table 2 shows that the accuracy and the number of eligible components increase when the Q value is increased. For an accuracy of 80%, there are 23–35 eligible components across all 12 slices of the heart CT for Q values of 8 and 16. To increase the accuracy to 90%, the Q value is increased from Q = 16 (with 35 eligible components) to a Q value ranging between 24 and 512. In this range of Q values, the number of eligible components across all 12 slices are 52

(at Q=24) to 494 (at Q=512). As all Q values from 24 to 512 achieve an accuracy of 90%, the increase in accuracy reduces with each increase in Q value; it is thus advisable to choose a smaller Q in the suitable range (24 to 512) as it requires a smaller number of eligible components to be merged. For example, Q values of 24 and 512 have a very small difference in accuracy but the number of components differ by a factor of approximately 9 (=494/52). Therefore, Q = 512 uses approximately nine times more eligible components than Q = 24, without a significant increase in accuracy.

The accuracy of the segmentation	**80%**	**90%**
Q value	8–16	24–512
Number of eligible components used for all slices	23–35	52–494

Table 2: The accuracy of the segmentation and number of eligible components used across all slices of the heart.

Figure 5 shows a comparison of the result obtained by the SRM segmentation (on the right) and the expert-segmented image (on the left) for slice number 8 of the heart and a Q value of 128. The visual comparison shows that the two segmented organs are almost identical in shape, contour and size, indicating that the result of the SRM segmentation is highly accurate and consistent with the manual results by a human expert. Of the 12 slices for the heart, the middle slices are nearly 98% correctly segmented; however, the earlier slices are less accurate, as Figure 6 shows.

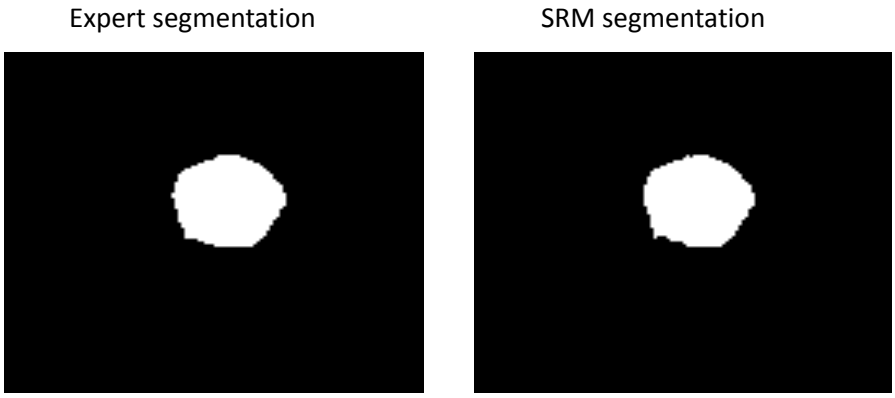

Figure 5: An example of the binary image of the ground truth overlaid with eligible components for slice no. 8 of the heart with a Q value of 128.

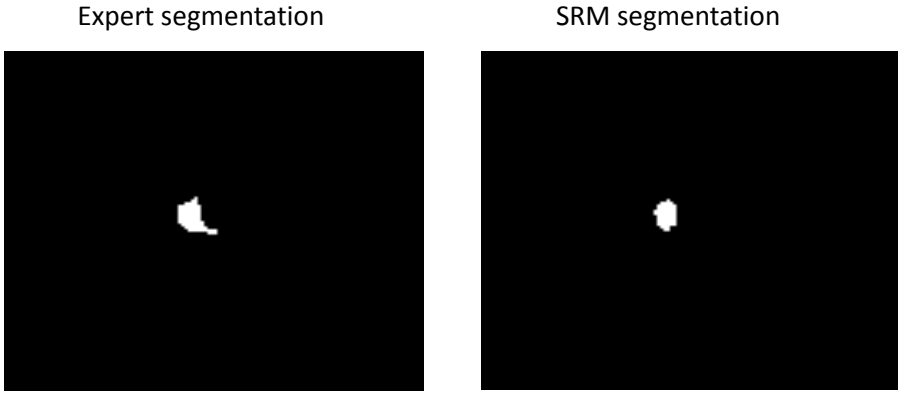

Figure 6: An example of the binary image of the ground truth overlaid with eligible components for slice no. 1 of the heart with a Q value of 128, which is less accurate.

To improve the display and understanding, the eligible components that merge to form the organ in any slice are shown in different colours in Figure 7. Notably, several components are not visible because one component is a subset of another component and is masked by the super set component. Typically, 20 or more components are required to create the heart organ in any slice.

Figure 7: An example of merging the eligible components in different colours.

As the Q value is increased, the number of eligible components per slice increases. For example, the maximum number of eligible components in the 12 slices for a Q value of 8 is four components in slice number 5, while the maximum number of eligible components for a Q value of 512 is 61 in slice number 7. Overall, the total number of components used in all 12 slices for a Q value of 8 is 23, which increases to 494 components for the combined 12 slices with a Q value of 512. Figure 8 shows the combined number of components (in all 12 slices) and the breakdown of the components per slice.

Appendix 1 gives the tabular data for the number of components used to construct the heart organ in each slice for all Q values.

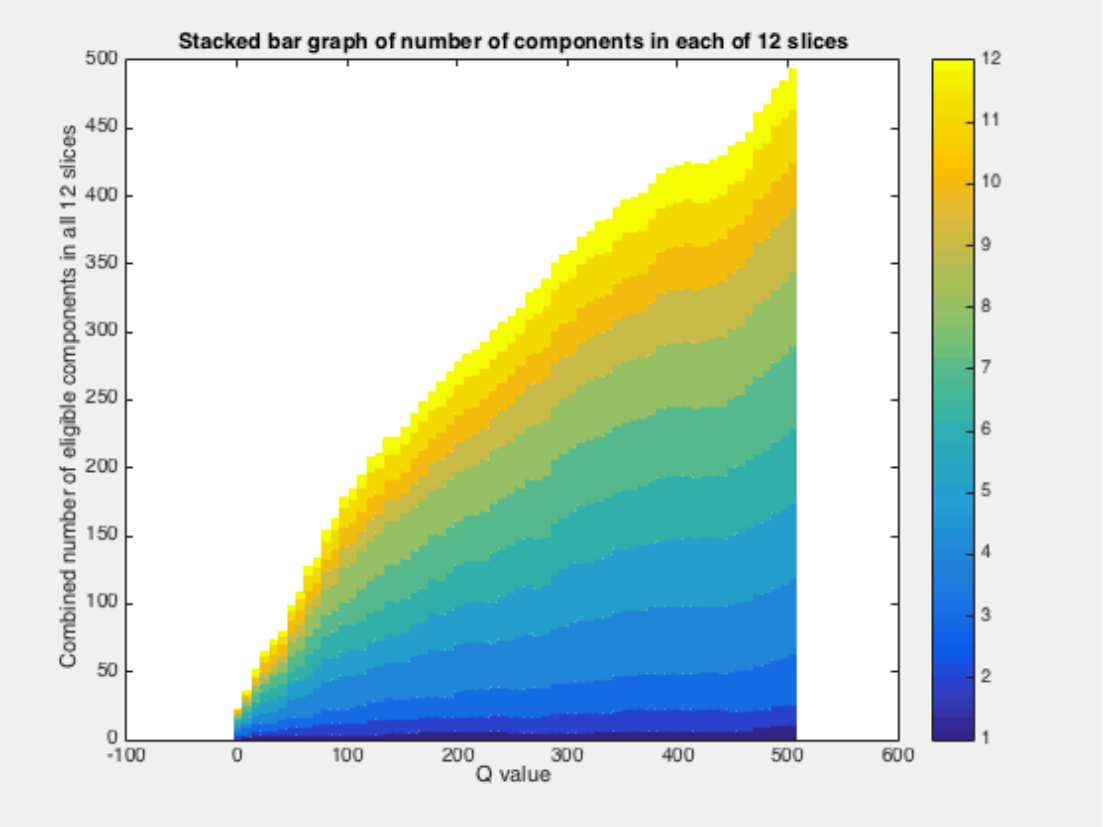

Figure 8: Number of eligible components used in all 12 slices with different Q values.

The Dice index values across 11 slices of the heart are above 0.80, and the Jaccard index values across 11 slices are above 0.70, when the Q value is 8. However, for slice 12, the Dice index is above 0.80 when the Q value is above 64, and the Jaccard index is above 0.70 when the Q value is above 64. As mentioned before, the heart organ is present in 12 CT slices. In some slices, the Jaccard index continuously increases as the value of Q is increased, as shown in Figure 9. Nevertheless, in heart slice number 8, shown in Figure 10, the Jaccard index remains almost constant at around 0.97 for all Q values of 8 to 512.

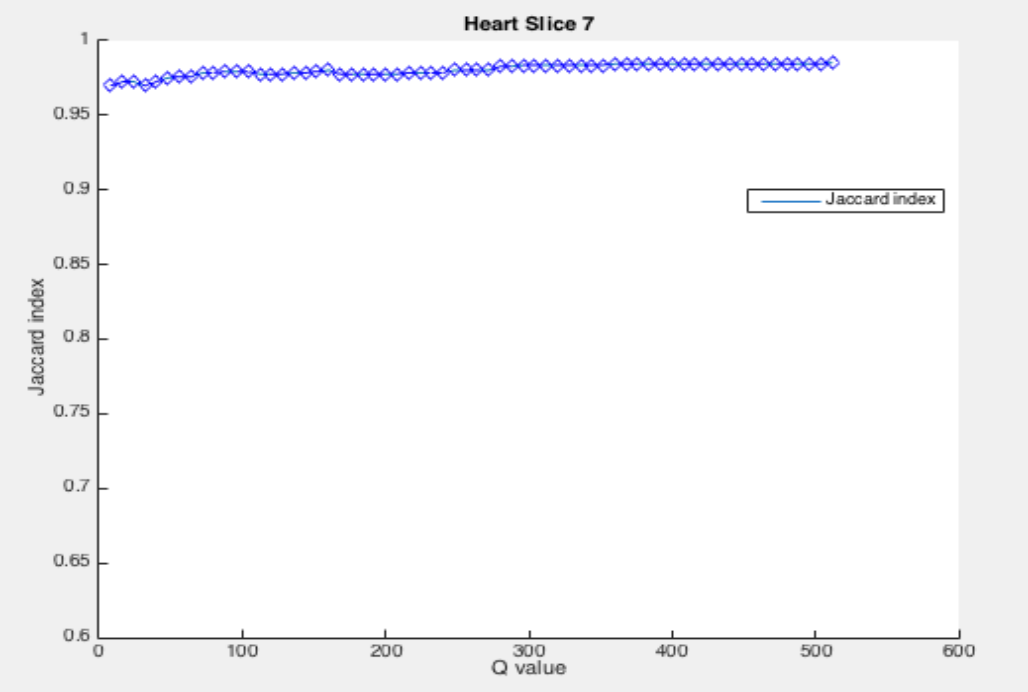

Figure 9: An example of the Jaccard index values for heart slice no. 7 across all Q values.

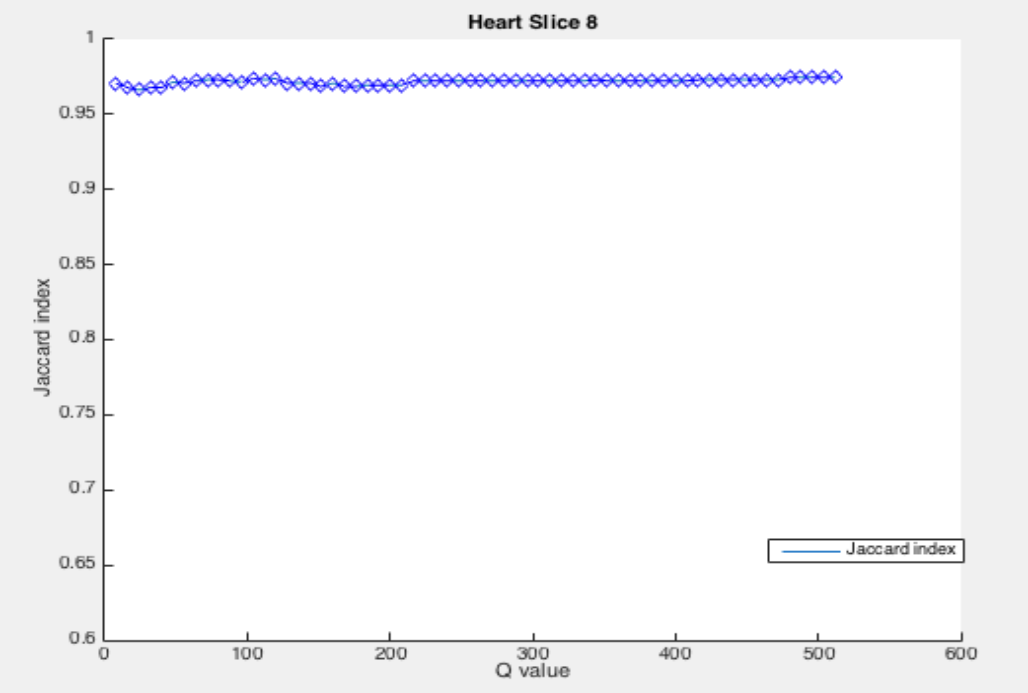

Figure 10: The Jaccard index remains at almost a constant value of 0.97 across all Q values from 8 to 512 for heart slice no. 8.

By contrast, Figure 11 shows unstable behaviour for slice 12. The Jaccard index reaches less than 0.7 for a Q value of 64, peaks at around 0.9 for a Q value of 80, remains around 0.9 up to a Q value of 176 and starts to drop slightly as Q increases further. Q values of 264 to 352 show a dip in the mean Jaccard index.

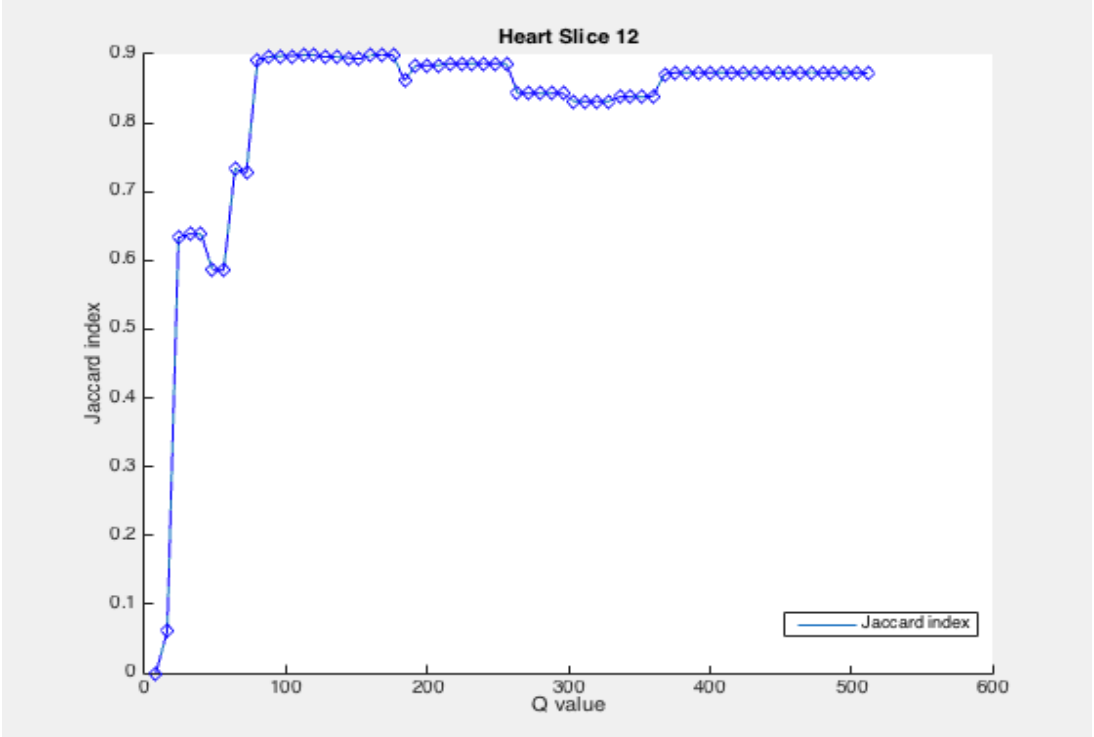

Figure 11: An example of unstable behaviour of the Jaccard values for Q values for heart slice no. 12.

Figure 12 shows a significant drop (from 0.84 to 0.73) in the mean Jaccard index for slice number 1 when the Q value changes from 56 to 64. An investigation of the problem found that the new smaller fragmented components were starting to fail the 50% overlap test, which resulted in a lower number of eligible components and the subsequent drop in the Jaccard index.

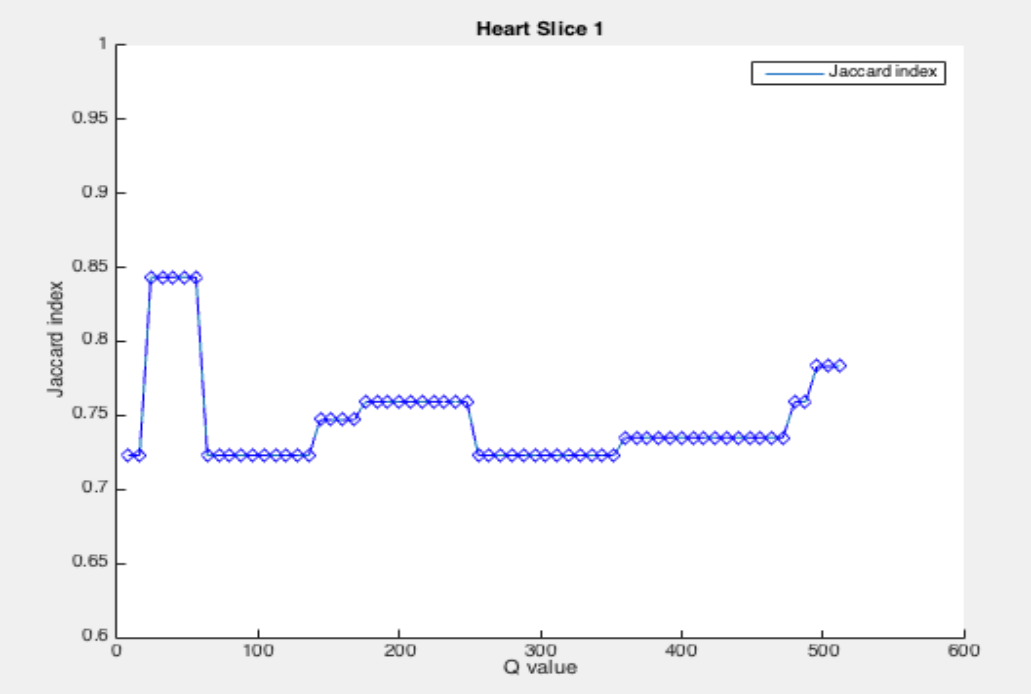

Figure 12: An example of unstable behaviour of the Jaccard values for the Q values for heart slice no. 1.

4.2 Spleen

Figure 13 shows the mean Dice index and the mean Jaccard index for the spleen. The SRM segmentation was performed on the spleen with Q values ranging from 8 to 512 in steps of eight. As the spleen organ appears in eight slices, the Dice and Jaccard indices display the mean Dice and Jaccard indices over all relate CT slices. The mean Jaccard index increases when the value of Q is increased. However, slight variations and occasional dips were evident at different Q values. Table 2 shows that the mean Jaccard index values for the spleen at Q values of 8 and 64 are very low (0.10 and 0.71, respectively). Similarly, the mean Jaccard index is 0.84 at a Q value of 128 and 0.90 at a Q value of 512.

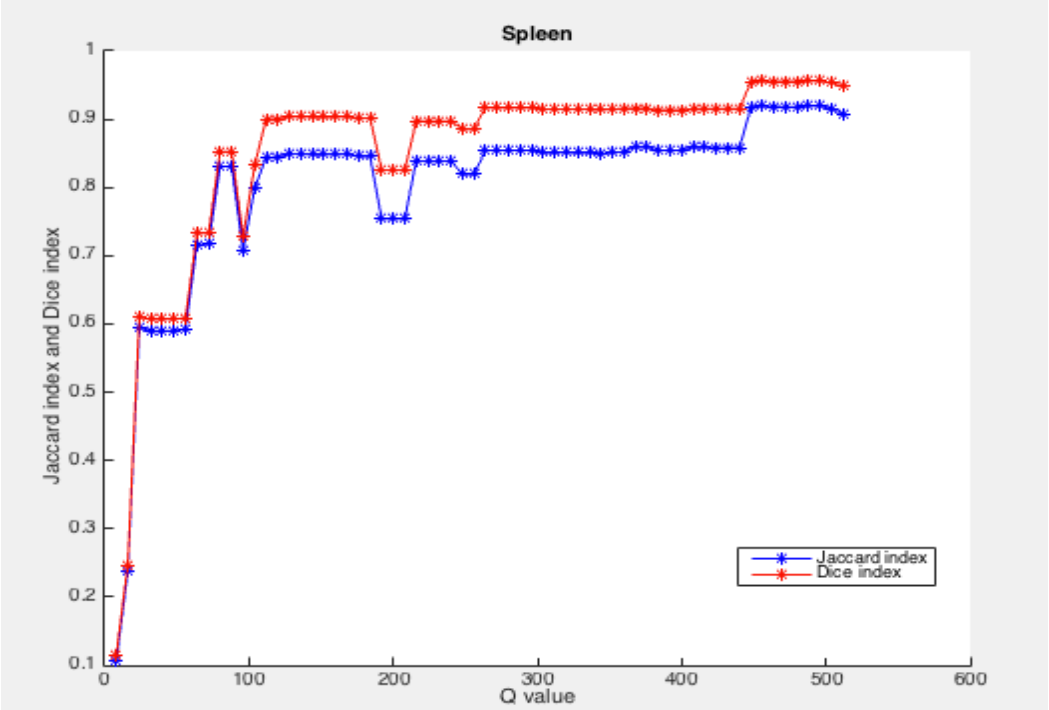

Figure 13: The mean Dice and Jaccard index values for eight CT slices for the spleen. SRM segmentation was performed for 64 Q values ranging from 8 to 512 in steps of eight.

Q value	8	64	128	512
Mean Jaccard index	0.10	0.71	0.84	0.90
Mean Dice index	0.11	0.73	0.90	0.95

Table 3: Evaluation of SRM for the spleen using the mean Jaccard and Dice indices for different Q values.

From Table 4, it can be seen that the accuracy and subsequently the number of eligible components increases when the Q value is increased. For an accuracy of 80%, there are 20 to 90 eligible components across all eight slices of the spleen CT for Q values of 80 and 440. To increase the accuracy to 90%, the Q value is increased from Q = 440 (with 90 eligible components) to a Q value ranging between 446 and 512. Therefore, for 80% accuracy, the number of eligible components across all eight slices are 20 (at Q=80) to 90 (at Q=440), and for 90% accuracy, the number of eligible components across all eight slices are 95 (at Q=448) to 107 (at Q=512). All values of Q between 80 and 440 achieve an accuracy of 80% and all values of Q between 448 and 512 achieve an accuracy of 90%. It is thus advisable that for a

required accuracy (of, e.g., 80%), a smaller Q in the suitable range (80 to 440) should be chosen, as it requires a smaller number of eligible components to be merged. To achieve 90% accuracy, Q values of 448 and 512 have a very small difference in accuracy; thus, in the case of the spleen, the difference in minimum and maximum eligible components is very high. For example, the number of eligible components in all eight slices is 20 at Q values of 80 and 90 at a Q value of 440.

The accuracy of the segmentation	**80%**	**90%**
Q value	80–440	448–512
Number of eligible components used for all slices	20–90	95–107

Table 4: The accuracy of the segmentation and the number of eligible components used across all slices for the spleen.

Figure 14 shows a comparison of the result obtained by the SRM segmentation (on the right) and the expert-segmented image (on the left) for slice number 1 of the spleen for a Q value of 512. The visual comparison shows that the two segmented organs are almost identical in shape, contour and size, indicating that the result of the SRM technique is highly accurate and consistent with the manual results by a human expert. There are eight slices for the spleen. At a Q value of 512, the organ contour in all the slices are correctly reconstructed by SRM. The worst case is slice number 6; however, this is still very close to the expert-segmentation shown in Figure 15. The outcomes for all the other slices are almost identical.

Expert segmentation SRM segmentation

Figure 14: An example of the binary image of the ground truth overlaid with eligible components for slice no. 6 of the spleen with a Q value of 512.

Expert segmentation SRM segmentation

Figure 15: An example of the binary image of the ground truth overlaid with eligible components for slice no. 8 of the spleen with a Q value of 512.

Figure 16: An example of merging the eligible components in different colours.

As the Q value is increased, the number of eligible components per slice increases. For example, the number of eligible components in all eight slices for a Q value of 96 has a maximum of five eligible components slice number 2. The maximum number of eligible components for a Q value of 512 is 29 for slice number 3. Overall, the total number of components used in all eight slices for a Q value of 96 is 21, which increases to 107 components for all eight slices for a Q value of 512. Slice 8 has only one component for all Q value of 104 or higher. There is no component for slice number 8 in the lower Q values. Slices 1 to 7 also exhibit no matched components for the smaller Q values. Figure 17 shows the combined number of components (in all eight slices) and the breakdown of the components per slice.

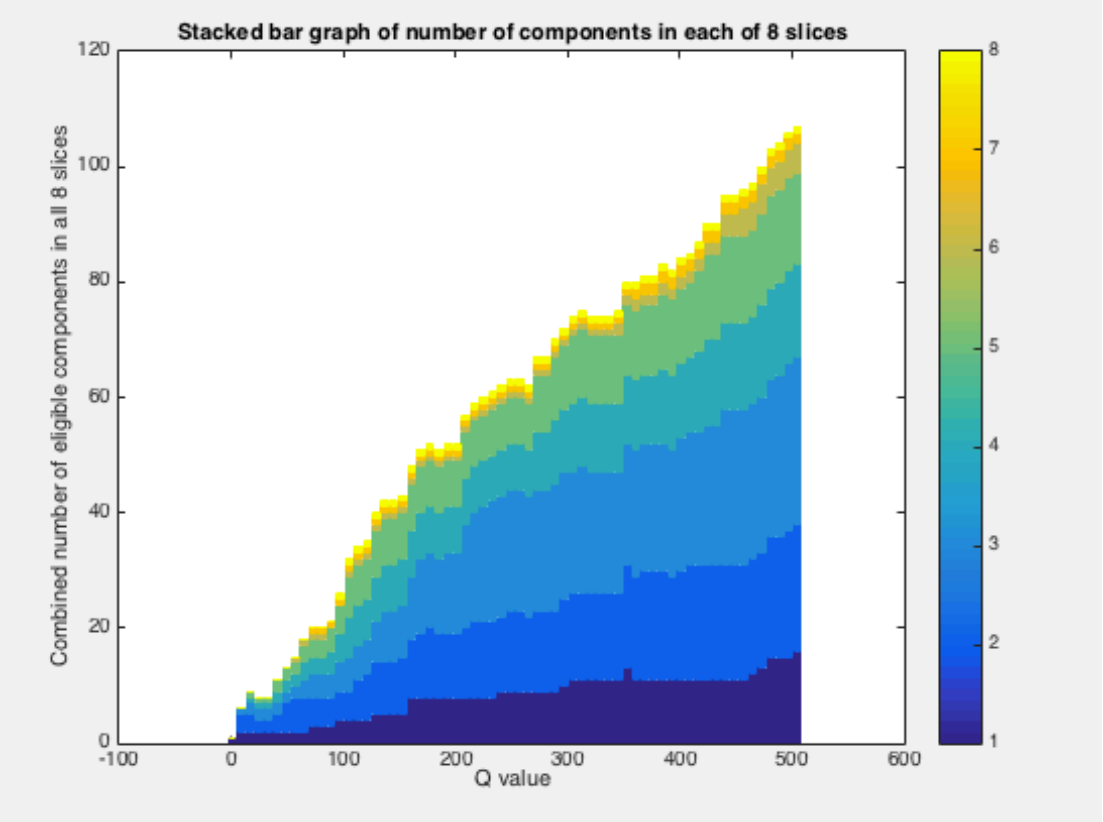

Figure 17: Number of eligible components in all eight slices with different Q values.

Appendix 2 gives the tabular data for the number of components used to construct the spleen organ in each slice for all Q values.

Some problems were encountered for different slices. For example, Figure 18 shows that no eligible components occurred in slice 8 up to a Q value of 96. After that, there is a sudden improvement in the merging, and the Jaccard index is approximately 0.75 for the remaining tested Q values. Slice number 8 is the most extreme case; however, there are several other slices where the overlap does not occur up to the threshold value of Q (which is less than Q=96 for slice 8).

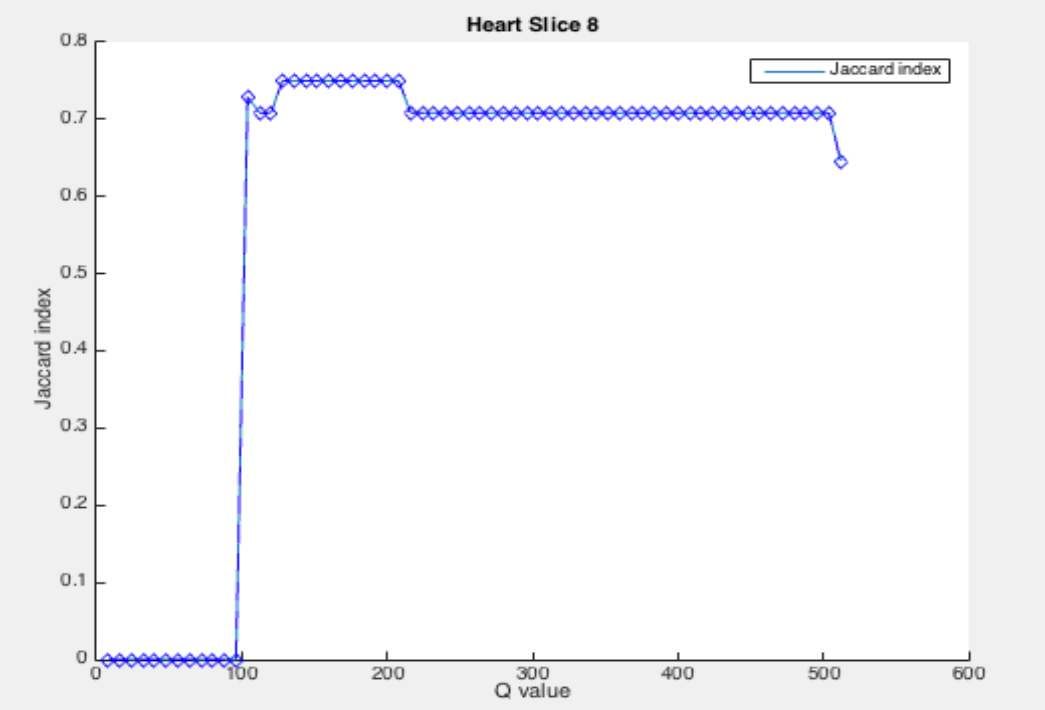

Figure 18: An example of unstable behaviour of the Jaccard values for Q values for spleen slice no. 8.

Some very sharp dips are also evident in the Jaccard index values when Q increases. For example, the Jaccard index for slice 6 of the spleen shows that changing the Q value from 88 to 96 reduces the Jaccard index from 0.98 to 0. Similarly, when the Q value changes from 360 to 368, the Jaccard index dips from 0.65 to 0.5. When the Q value increases, the number of eligible components increases because the previous eligible components (for a smaller value of Q) are now fragmented into a larger number of smaller components. In some cases, the new smaller components fail the 50% threshold test used to confirm the eligibility of the component. This reduces the Jaccard index for some values of Q as seen in Figure 19.

The opposite occurs when the Q value is changed from 104 to 112. In this case, even though the fragments are still getting smaller, the new smaller components pass the threshold test and are used in the statistical merging. Therefore, the reconstruction improves and the mean Jaccard index increases from 0 to 0.39. The situation remains constant until the Q value changes to 256, when another jump occurs due to the inclusion of further eligible components, as shown in Figure 19.

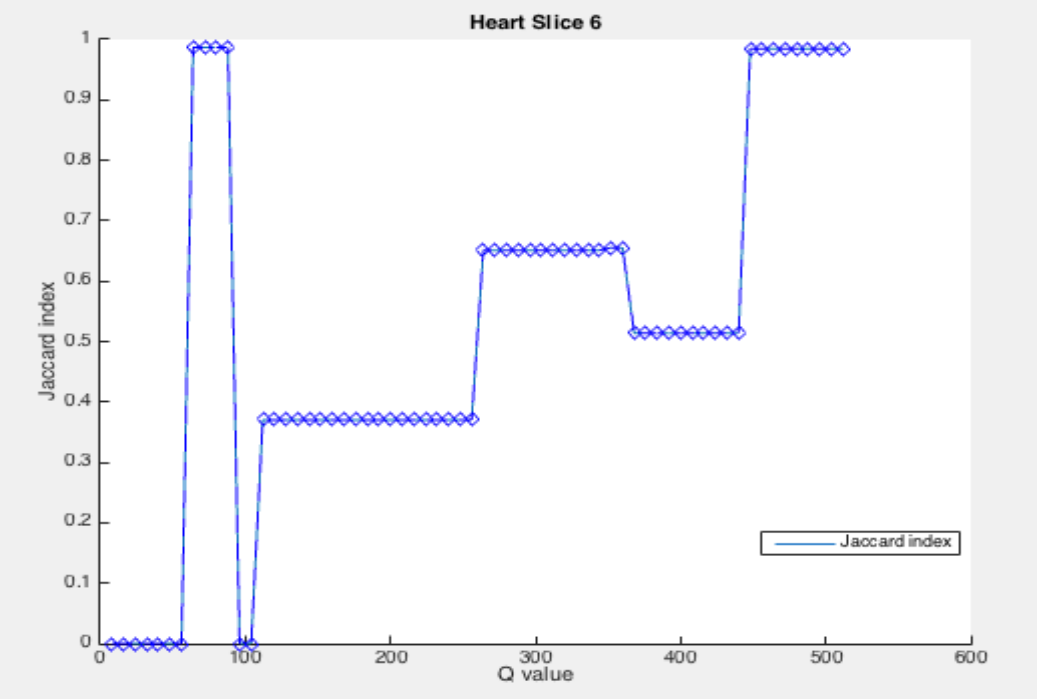

Figure 19: An example of unstable behaviour of the Jaccard values for Q values for spleen slice no. 6.

CHAPTER 5: DISCUSSION AND CONCLUSION

5.1 Discussion

The results found very low mean Jaccard index values for the spleen at Q values of 8 and 64 (0.10 and 0.71, respectively) compared to the mean Jaccard index values for the heart, which were 0.85 and 0.90, respectively. Similarly, the mean Jaccard index at a Q value of 128 was 0.84 compared to 0.92 for the heart. Table 2 (in the Results section) shows that an accuracy of 80%–89.9% for the heart slices can be achieved with Q values of 8 or 16. For a 90% accuracy in heart reconstruction, the Q value needs to be in the range 24 to 512, and for the spleen (Table 4 in the Results section), an 80% accuracy requires a Q value in the range of 80 to 440, while the value of Q should be between 448 and 512 for an accuracy of at least 90%. The spleen therefore requires a much higher Q value to achieve comparable results with the heart. The primary reason for the discrepancy in the results is that the heart is relatively easy to segment out from the larger organs, such as the left and right lungs, whereas the spleen is surrounded by other organs such as the stomach and is thus harder to segment out from other organs. The CT scan slices for the spleen are 50, 51, 52, 53, 58, 59, 60 and 61. The spleen does not appear in slices 54, 55, 56 and 57 because other organs occlude it. The heart is available in all sequential slices from 38 to 49.

Tables 2 and 4 (in the Results section) show that the number of eligible components in the heart slices is much higher than that of the spleen. One reason for this is that the heart appears in 12 slices and the spleen only appears in eight slices. Even considering the number of slices, the heart still has a much higher number of eligible components. For example, the number of eligible components varies from 24 (at Q=8) to 494 (at Q=512) for the heart, while the number of eligible components varies from 51 (at Q=192) to 107 (at Q=512) for the spleen. Because at least one or more slices have missing spleen components for Q values less than 192, these results are skewed and have not been included.

The mean Jaccard index values for the heart at Q values of 8, 64, 128 and 512 are 0.85, 0.90, 0.92 and 0.93, respectively. This means that increasing the Q value from 8 to 128 increases the Jaccard index by 0.07, while there is only an increase of 0.01 when increasing the Q value from 128 to 512. The additional number of eligible components does not contribute

significantly to the increase in accuracy. However, the mean Jaccard index values for the spleen at Q values of 8, 64, 128 and 512 are 0.10, 0.71, 0.84 and 0.90, respectively. The Q values of 8 and 64 in this case are unusable because of a very low mean Jaccard index. For the spleen, slices 6, 7 and 8 do not have any eligible components for Q values lower than 64. All eight slices start showing at least one eligible component for a Q value of 104. Although a Q value of 128 increases the mean Jaccard index to 0.84 and the value is sufficiently high, the individual slice number six at a Q value of 128 only has a Jaccard index of 0.37. Only the results for Q values of 192 and higher are used because all slices at these Q values have a mean Jaccard index greater than 0.7. The mean Jaccard index increases to a 0.90 at a Q value of 512. Although there are some Q values at which the Jaccard index for individual slices suddenly drops due to the smaller fragmented components failing the threshold, overall the mean Jaccard index increases as the Q values increase.

Accuracy is not the only criteria. The number of components should be kept as small as possible to ease the labelling task. Hence, for a required accuracy, the number of components can be minimised by choosing the right Q value, which uses the minimum number of eligible components to achieve that particular accuracy.

Tables 2 and 4 present the suitable ranges of Q values that achieve a particular/specified threshold accuracy (80% and 90%) for the heart and the spleen. Different applications require different accuracies, so the value of Q can be adjusted based on the accuracy required and the type of organ. For example, as applications such as CT scans for detecting cancer require a higher accuracy, the Q value should be chosen from the correct range. The results obtained in this study validate the results obtained in previous studies (Lee, Bajger and Caon 2012). This study provides a method of determining the Q value based on the required accuracy and relates the accuracy with the number of eligible components. As this study used different techniques of segmentation from previous studies, which used atlas or probability shape models, the results of this study are not comparable with the other studies.

5.2 Conclusion

The SRM segmentation technique was used to automatically segment two organs (heart and spleen) in this study. The results confirm that the technique has the potential to provide Q

high accuracy (mean Jaccard index of 0.9) for both organs. However, neither the heart nor the spleen can be segmented at this level of precision because an additional post-processing would be necessary to merge eligible regions into one piece corresponding to the whole organ. The SRM segmentation technique can potentially be used to closely match the accuracy obtained by a human expert, and the Q value should be chosen depending on the accuracy required. Tables 2 and 4 (see Section 4) present the ranges of Q required for accuracies of at least 80% and 90% for the heart and spleen, respectively. The strength of this technique is that it does not require any image pre-processing. Manual segmentation is a time-consuming process and requires a human expert with sufficient experience. Automatic segmentation using SRM is much faster, does not require a human expert and retains adequate the same level of accuracy.

5.3 Future work

This study only automatically segmented the heart and spleen; however, future studies can extend the method to other organs. Additionally, future work will incorporate the labelling task into the SRM segmentation technique.

REFERENCES

Artaechevarria,X, Munoz-Barrutia, A, and Ortiz-de Solorzano, C 2009, ' Combination strategies in multi-atlas image segmentation: Application to brain MR data', *IEEE Trans. Med. Imag.,* vol. 28, no. 8, pp. 1266–1277.

Bajger, M, Ma,F, and Bottema, MJ 2009, 'Automatic tuning of MST segmentation of mammograms for registration and mass detection algorithms', in *Digital Image Computing Techniques and Applications (DICTA2009)*, Melbourne, pp. 400–407.

Bajger, M, Ma, F, Williams, S & Bottema, M 2010, 'Mammographic Mass Detection with Statistical Region Merging, 2010 Digital Image Computing: Techniques and Applications2010 IEEE', *Computer Society.*

Bajger, M, Lee, G & Caon, M 2012, 'Full-body CT segmentation using 3d extension of two graph-based methods: a feasibility study', *Refereed proceedings of the IASTED International Conference Signal Processing, Pattern Recognition and Applications (SPPRA 2012),* Crete, Greece, pp. 43-50.

Ballard, D and Brown,C 1982, *Computer Vision*, Englewood Cliffs, NJ, Prentice-Hall.

Battiato, S, Bosco, C, Farinella, GM, and G. Impoco, G 2006, '3D CT segmentation for clinical evaluation of knee prosthesis operations' , *in Proceedings of the Eurographics Italian Chapter*'.

Calderero, F 2010, 'Region Merging Techniques Using Information Theory Statistical Measures', *IEEE transactions on image processing*, vol.19, no.6.

Caon, M, Bibbo,G, and Pattison, J 1999, 'An EGS4- ready tomographic computational model of a 14-yearold female torso for calculating organ doses from CT examinations' , *Physics in Medicine and Biology*, vol.44, no. 9, pp. 2213–2225.

Celebi, ME, Kingravi, HA, Lee, J, Stoecker, WV, Malters, JM, Iyatomi, H, Aslandogan, YA, Moss, R, and Marghoob, AA 2007, 'Fast and accurate border detection in dermoscopy images using Statistical RegionMerging', *in J. P. W. Pluim and J. M. Reinhardt, editors, Proceedings of SPIE: Medical Imaging,* vol. 6512, pp.1-10.

Celebi, ME, Kingravi, HA, Uddin, B, Iyatomi, H, Aslandogan, YA, Stoecker, WV, Moss, RH, Malters, JM, Grichnik, JM, Marghoob, AA, Robinovitz, HS, Menzies, SW 2008, 'Border detection in dermoscopy images using statistical region merging', *Skin Research and Technology,* vol. 14, pp. 347-353.

Commowick, O, and Malandain, G 2007, 'Efficient Selection of the Most Similar Image in a Database for Critical Structures Segmentation', in *MICCAI 2007*, Springer, Berlin, Germany, vol. 4792, Lecture Notes in Computer Science, pp. 203–210.

Cuadra, M, Pollo, C, Bardera, A, Cuisenaire, O, Villemure, JG, Thiran, JP 2004,' Atlas-based segmentation of pathological MR brain images using a model of lesion growth', *IEEE Trans. on Medical Imaging,* vol. 23, pp. 1301–1314.

Dawant, BM, Hartmann, SL, Thirion, JP, Maes, F, Vandermeulen, D, Demaerel, P 1999, 'Automatic 3-D segmentation of internal structures of head in MR images using a combination of similarity and free-form transformations: Part I, methodology and validation on normal subjects', *IEEE Trans. on Medical Imaging,* 18, 909–916.

Ding, F, Leow, WK & Wang, SC n.d '*Segmentation of 3D CT Volume Images Using a Single 2D Atlas*'.

Dvorak, P, Bartusek, K & Gescheidtova, E 2014, 'Automatic Segmentation of Multi-contrast MRI Using Statistical Region Merging, Progress in Electromagnetics Research Symposium', *Refereed proceedings,* pp. 1865-1869.

Felzenszwalb, PF, & Huttenlocher, DP 2004, 'Efficient graph-based image segmentation', *Int. J. Comput. Vision*, vol. 59, no. 2, pp. 167–181.

Fleckenstein, P, Jensen, JT, 1993, 'Anatomy in Diagnostic Imaging', *Munksgaard.*

Fukunaga, K 1990, *Introduction to statistical pattern recognition,* 2nd ed., Academic Press.

Gies, V and Bernard, T 2004, 'Statistical solution to watershed over-segmentation', in *Proc. ICIP*, Oct. 24–27, vol. 3, pp. 1863–1866.

Gonzalez RC, Woods RE 2004, *Digital image processing, 2nd ed.,* Pearson Education.

Han, X, Hoogeman, M, Levendag, P, Hibbard,L, Teguh, D, Voet, P, Cowen, A, and Wolf, T 2008, 'Atlas-Based Auto-Segmentation of Head and Neck CT Images', in *MICCAI ,* Springer Berlin, Germany, vol. 5242, Lecture Notes in Computer Science, pp. 434–441.

Hance, GA, Umbaugh, SE, Moss, RH, and Stoecker WV 1996,'Unsupervised Color Image Segmentation with Application to Skin Tumor Borders', *IEEE Engineering in Medicine and Biology Magazine*, vol. 15, no. 1, pp. 104-111.

Haralick RM, 1979, 'Statistical and structural approaches to texture', *Proc IEEE,* vol. 67, pp.786–804.

Hartmann, SL, Parks, MH, Martin, PR, Dawant, BM 1999, 'Automatic 3-D segmentation of internal structures of head in MR images using a combination of similarity and free-form transformations: Part II, validation on severely atrophied brains', *IEEE Trans. on Medical Imaging*, 18, pp. 917–926.

Isgum, I, Staring, M, Rutten, A, Prokop, M, Viergever, M, and van Ginneken, B 2009, 'Multi-atlas-based segmentation with local decision fusion— Application to cardiac and aortic segmentation in CT scans', *IEEE Trans. Med. Imag.*, vol. 28, no. 7, pp. 1000–1010.

Julesz, B 1981, 'Textons, the element of texture perception and their interactions', *Nature,* vol. 290, pp. 91–97.

Kass, M, Witkin, A, and Terzopoulos, D 1987, 'Snakes: active contourmodels', *Int. J. Comput. Vision*, vol. 1, no. 4, pp.321– 331.

Koss JE, Newman FD, Johnson TK, Krich DL 1999, 'Abdominal organ segmentation using texture transform and Hopfield neural network', *IEEE Trans Med Image*, vol. 18, pp.640–648.

Lau, HT & Al-Jumaily, A 2009, 'Automatically Early Detection of Skin Cancer: Study Based on Neural Network Classification', *Refereed proceedings of the IEEE computer society International Conference on Soft Computing and Pattern Recognition*, pp. 375-380.

LaValle, S and Hutchinson, SM 1993, 'Bayesian region merging probability for parametric image models', in *Proc. IEEE Computer Soc. Conf. Computer Vision Pattern Recognition*, June 15–17, pp. 778–779.

Lee, G, Bajger, M & Caon, M 2012*, '*Multi-organ segmentation of CT images using statistical region merging', *Refereed proceedings of the IASTED International Conference Biomedical Engineering (BioMed 2012),* pp.199-206.

Li, H, Gu, H, Han, Y & Yang, J 2009, 'An Efficient Multiscale SRMMHR (Statistical Region Merging and Minimum Heterogeneity Rule Segmentation Method for High-Resolution Remote Sensing Imagery', *IEEE journal of selected topics in applied earth observations and remote sensing*, vol.2, no.2, pp. 67-73.

Linguraru, MG, Sandberg, JK, Li, Z, Shah, F & Summers, RM 2010, 'Automated segmentation and quantification of liver and spleen from CT images using normalized probabilistic atlases and enhancement estimation', *Med. Phys*, vol.37, no.2, pp. 771-783.

Lorenzo-Vald´es, M, Sanchez-Ortiz, GI, Elkington, AG, Mohiaddin, RH, Rueckert, D 2004, 'Segmentation of 4D cardiac MR images using a probabilistic atlas and the EM algorithm' , *Medical Image Analysis,* vol*.* 8, pp. 255–265.

Ma, F, Bajger, M, Slavotinek,JP and Bottema, MJ 2007, 'Two graph theory based image segmentation methods for identifying pectoral muscle in mammograms', *Pattern Recognition*, vol. 40, pp. 2592–2602.

Ma,F, Bajger, M, and Bottema, MJ 2009, 'Automatic mass segmentation based on adaptive pyramid and sublevel set analysis,' in *Digital Image Computing Techniques and Applications (DICTA2009)*, pp. 236–241.

Nock, R & Nielsen, F 2004, 'Statistical region merging', *IEEE transactions on pattern analysis and machine intelligence,* vol.26, no.11, pp. 1452-1458.

Okada, T, Yoshida, Y, Hori, M, Summers, RM, Chen, Y-W, Tomiyama, N, and Sato, Y 2012, 'Abdominal Multi-organ Segmentation of CT Images Based on Hierarchical Spatial Modeling of Organ Interrelations,' in *Proceedings of the Third International Conference on Abdominal Imaging: Computational and Clinical Applications*, ser. MICCAI'11. Berlin, Germany, Springer-Verlag, pp. 173–180.

Pal NR, Pal SH 1993, 'A review on image segmentation techniques', *Pattern Recog.,* vol. 26, pp. 1277–94.

Park, H, Bland, P, and Meyer, C 2003, 'Construction of an abdominal probabilistic atlas and its application in segmentation', *IEEE Trans. Med. Imag.*, vol. 22, no. 4, pp. 483–492.

Pham, DL, Xu, C, and Prince JL 2000, 'Current methods in medical image segmentation', *Ann Rev Biomed Engg.,* vol. 2, pp.315–337.

Prasad M, Brown MS, Abtin F, Vasunilsahorn S, Goldin JG 2008, 'Comparison of scleroderma detection techniques in CT images using supervised and unsupervised methods', *Int J Biomed Engg Technol,* vol.1, no.4.

Pratt KW 2001, *Digital image processing*, 3rd ed., Willey, pp. 551–87.

Prince JL, Links JM 2006, *Medical imaging signals and system,* Pearson Education.

Ramus,L, Commowick, O, and Malandain, G 2010, ' Construction of Patient Specific Atlases from Locally Most Similar Anatomical Pieces, ser', MICCAI , Springer-Verlag, Berlin, pp. 155–162.

Rueckert, D, Sanchez-Ortiz, GI, Lorenzo-Vald´es, M, Chandrashekara, R, Mohiaddin, R 2002, 'Non-rigid registration of cardiac MR: Application to motion modelling and atlas-based segmentation'*, IEEE Int. Symp. on Biomedical Imaging*.

Sharma, N & Aggarwal, LM 2010, 'Automated medical image segmentation techniques', *J Med Phys,* vol.35, no.1, pp. 3-14.

Sonka M, Hlavac V, and Boyle R 1999, *Image processing, analysis and machine vision*, Singapore: Thomson Learning.

Srinivasan, GN, Shobha, G 2007, 'Segmentation Techniques for Target Recognition', *International Journal of Computers and Communications,* vol. 1, issue 3, pp. 75-81.

Tesar L, Shimizu A, Smutek D, Kobatake H, Nawano, S 2008, 'Medical image analysis of 3D CT images based on extension of Haralick texture features', *Comput Med Imaging Graphics,* vol.32, pp.513–520.

Thurfjell, L, Bohm, C, Greitz, T, Eriksson, L 1993, 'Transformations and algorithmsin a computerized brain atlas', *IEEE Tran. on Nuclear Science*, vol. 40, pp. 1187–1191.

van Rikxoort, EM, Isgum, I, Arzhaeva, Y, Staring, M, Klein, S, Viergever, MA, Pluim, JP, and van Ginneken, B 2010, 'Adaptive local multi-atlas segmentation: Application to the heart and the caudate nucleus', *Med. Image Anal.,* vol. 14, no. 1, pp. 39–49.

Wang, H, Stout, D, and Chatziioannou, A 2012, 'Estimation of mouse organ locations through registration of a statistical mouse atlas with micro-ct images', *IEEE Trans. Med. Imag.*, vol. 31, no. 1, pp. 88–102.

Wolz, R, Chu, C, Misawa, K, Fujiwara, M, Mori, K & Rueckert, D 2013, 'Automated Abdominal Multi-Organ Segmentation with Subject-Specific Atlas Generation', *IEEE transactions on medical imaging*, vol.32, no.9, pp. 1723-1730.

Xu, C, and Prince, JL 1998,'Snakes, shapes, and gradient vector flow', *IEEE Transactions on Image Processing*, vol.7, no.3, pp. 359–369.

Zortea, M, Skrovseth, SO, Schopf, TR, Kirchesch, HM, and Godtliebsen, F 2011, 'Automatic segmentation of dermoscopic images by iterative classification', *International Journal of Biomedical Imaging*.

APPENDIX

Number of eligible components across 12 slices for heart

Q value	slice 1	slice 2	slice 3	slice 4	slice 5	slice 6	slice 7	slice 8	slice 9	slice 10	slice 11	slice 12	total
8	1	2	3	3	4	2	2	2	1	2	1	0	23
16	2	2	5	4	4	4	5	3	2	2	1	1	35
24	3	3	6	6	6	5	7	4	2	4	2	4	52
32	3	3	6	8	9	6	7	4	4	7	5	3	65
40	3	3	7	8	9	8	9	4	5	7	7	3	73
48	4	4	8	8	9	8	9	7	5	8	7	3	80
56	4	5	8	13	12	8	12	10	8	8	7	3	98
64	3	6	8	15	12	10	12	11	9	9	9	4	108
72	3	6	8	17	14	15	14	15	9	9	12	5	127
80	4	7	8	20	14	15	15	15	9	9	11	6	133
88	4	8	10	21	17	16	17	19	12	10	13	7	154
96	4	9	10	21	17	18	17	18	13	13	15	8	163
104	4	9	10	24	18	20	17	25	15	15	14	8	179
112	4	9	11	24	18	19	19	26	15	15	17	8	185
120	4	9	12	24	18	21	23	28	15	15	17	8	194
128	4	11	14	24	19	22	24	29	16	15	19	11	208
136	4	11	14	25	19	22	26	31	15	15	18	11	211
144	5	11	15	27	20	25	26	32	16	14	19	12	222
152	5	11	14	27	20	24	28	32	16	15	19	12	223
160	5	11	14	28	22	24	29	34	15	16	20	12	230
168	5	11	15	31	23	24	30	33	17	18	20	13	240
176	6	11	15	32	23	25	32	33	17	19	22	13	248
184	6	11	15	34	28	25	32	34	18	19	21	13	256
192	6	11	15	34	27	25	33	35	20	20	20	17	263
200	6	11	15	35	29	27	33	36	20	21	20	17	270
208	6	11	18	35	29	30	32	36	22	21	21	17	278
216	6	11	20	35	30	30	33	36	22	21	23	17	284
224	6	11	20	35	30	30	34	36	23	21	23	18	287
232	6	11	20	35	30	30	36	38	24	21	23	18	292
248	6	12	20	35	32	36	37	39	25	21	24	20	307
256	5	12	20	36	35	36	38	38	27	22	23	19	311
264	5	12	20	40	36	35	38	39	29	22	23	18	317
272	5	12	21	40	36	36	42	39	31	22	26	19	329
280	5	12	21	38	36	36	44	40	31	22	29	18	332
288	5	13	21	39	36	36	45	41	31	23	30	19	339
296	5	14	22	42	39	38	46	40	31	24	30	19	350
304	5	14	22	43	42	38	46	41	31	25	31	18	356
312	5	14	24	44	43	39	46	40	28	25	32	19	359

Number of eligible components across 12 slices for heart

Q value	slice 1	slice 2	slice 3	slice 4	slice 5	slice 6	slice 7	slice 8	slice 9	slice 10	slice 11	slice 12	total
320	5	14	24	45	43	42	47	42	31	26	32	19	370
328	5	14	24	45	43	43	47	43	32	27	32	19	374
336	6	14	24	46	43	43	48	43	32	27	32	23	381
344	6	14	25	46	45	43	48	43	32	26	32	23	383
352	6	15	26	47	46	43	48	45	34	26	32	23	391
360	7	15	26	47	46	43	50	45	34	28	33	23	397
368	7	15	26	47	47	43	50	45	34	28	32	24	398
376	7	15	27	47	47	44	50	45	34	28	32	25	401
384	7	15	27	49	47	45	50	44	37	30	33	25	409
492	7	15	27	49	49	46	51	47	38	30	32	25	416
400	7	15	27	50	49	46	51	47	39	31	32	26	420
408	7	15	27	50	49	46	51	47	39	33	32	26	422
416	7	15	27	50	50	46	51	48	39	33	32	27	425
424	7	15	27	49	50	45	51	48	39	33	32	27	423
432	7	16	27	49	50	45	51	49	39	31	32	27	423
440	7	16	27	49	50	45	51	50	39	33	32	27	426
448	7	15	27	49	50	47	51	50	41	33	32	27	429
456	7	14	29	50	51	49	52	51	41	33	32	27	436
464	7	14	31	50	51	49	53	51	42	32	32	27	439
472	7	14	31	51	53	50	55	51	43	32	33	27	447
480	9	14	32	52	53	51	57	53	45	32	37	27	462
488	9	14	33	53	53	51	59	53	45	33	37	27	467
496	11	14	34	53	55	52	59	54	46	34	40	27	479
504	11	15	34	55	55	54	59	55	46	34	40	27	485
512	11	15	37	56	55	55	61	55	46	34	40	29	494

Number of eligible components across 8 slices for spleen

Q value	slice 1	slice 2	slice 3	slice 4	slice 5	slice 6	slice 7	slice 8	total
8	1	0	0	0	0	0	0	0	1
16	2	3	1	0	0	0	0	0	6
24	2	3	2	1	1	0	0	0	9
32	2	2	2	1	1	0	0	0	8
40	2	2	2	1	1	0	0	0	8
48	2	3	2	2	2	0	0	0	11
56	2	5	2	2	2	0	0	0	13
64	2	6	2	2	2	1	0	0	15
72	2	6	3	3	3	1	0	0	18
80	3	5	4	3	3	1	1	0	20
88	3	5	4	3	3	1	1	0	20
96	3	5	4	4	4	0	1	0	21
104	4	5	5	5	5	0	1	1	26
112	4	5	6	7	7	1	1	1	32
120	4	7	6	7	7	1	1	1	34
128	4	8	6	7	7	1	1	1	35
136	5	9	7	8	8	1	1	1	40
144	5	9	9	8	8	1	1	1	42
152	5	9	9	8	8	1	1	1	42
160	5	10	9	8	8	1	1	1	43
168	8	10	11	8	8	1	1	1	48
176	8	11	13	8	8	1	1	1	51
184	8	12	13	8	8	1	1	1	52
192	8	11	13	8	8	1	1	1	51
200	8	11	14	8	8	1	1	1	52
208	8	11	14	8	8	1	1	1	52
216	8	12	18	8	8	1	1	1	57
224	8	13	19	8	8	1	1	1	59
232	8	13	20	8	8	1	1	1	60
248	9	13	21	8	8	1	1	1	62
256	9	14	21	8	8	1	1	1	63
264	9	14	21	8	8	1	1	1	63
272	9	13	21	8	8	1	1	1	62
280	9	14	21	10	10	1	1	1	67
288	9	14	21	10	10	1	1	1	67
296	9	14	22	11	11	1	1	1	70
304	10	15	22	11	11	1	1	1	72
312	11	15	21	12	12	1	1	1	74

Number of eligible components across 8 slices for spleen

Q value	slice 1	slice 2	slice 3	slice 4	slice 5	slice 6	slice 7	slice 8	total
320	11	15	22	12	12	1	1	1	75
328	11	15	21	12	12	1	1	1	74
336	11	15	21	12	12	1	1	1	74
344	11	15	21	12	12	1	1	1	74
352	11	15	21	12	12	2	1	1	75
360	13	18	21	12	12	2	1	1	80
368	11	18	22	12	12	2	2	1	80
376	11	19	22	12	12	2	2	1	81
384	11	19	22	12	12	2	2	1	81
492	11	19	22	13	13	2	2	1	83
400	11	18	22	13	13	2	2	1	82
408	11	19	23	13	13	2	2	1	84
416	11	20	23	13	13	2	2	1	85
424	11	20	23	14	14	2	2	1	87
432	11	20	24	15	15	2	2	1	90
440	11	20	24	15	15	2	2	1	90
448	11	20	27	15	15	4	2	1	95
456	11	20	27	15	15	4	2	1	95
464	11	20	27	15	15	5	2	1	96
472	12	20	27	15	15	5	2	1	97
480	13	20	27	16	16	5	2	1	100
488	15	21	27	16	16	5	2	1	103
496	15	21	28	16	16	5	2	1	104
504	15	22	29	16	16	5	2	1	106
512	16	22	29	16	16	5	2	1	107