Empirical Evaluation of Segmentation Algorithms for Lung Modelling

S.L.A. Lee, A.Z. Kouzani, and E.J. Hu
School of Engineering and IT, Deakin University
Waurn Ponds, VIC 3217 AUSTRALIA
E-mail: slale,kouzani@deakin.edu.au

Abstract— **Lung modelling has emerged as a useful method for diagnosing lung diseases. Image segmentation is an important part of lung modelling systems. The ill-defined nature of image segmentation makes automated lung modelling difficult. Also, low resolution of lung images further increases the difficulty of the lung image segmentation. It is therefore important to identify a suitable segmentation algorithm that can enhance lung modelling accuracies. This paper investigates six image segmentation algorithms, used in medical imaging, and also their application to lung modelling. The algorithms are: normalised cuts, graph, region growing, watershed, Markov random field, and mean shift. The performance of the six segmentation algorithms is determined through a set of experiments on realistic 2D CT lung images. An experimental procedure is devised to measure the performance of the tested algorithms. The measured segmentation accuracies as well as execution times of the six algorithms are then compared and discussed.**

Keywords— **CT lung images, image segmentation**

I. INTRODUCTION

Lung diseases include conditions that inhibit the lungs function properly. They affect the patient's ability to carry out normal daily activities, and can be sometimes lethal. Lung diseases can be diagnosed in several ways. A common approach is through lung imaging techniques, e.g. computed tomography (CT) and magnetic resonance imaging (MRI), which can benefit from lung modelling.

The lung modelling refers to the process in which intelligent algorithms analyse 2D lung images, and form a realistic 3D model of the lung. The lung modelling process consists of several modules. Image segmentation is an important module of a typical lung modelling system.

Image segmentation is the process of dividing an image into distinct regions that altogether cover the whole image. The output of a segmentation algorithm is a label assigned to each image pixel. It facilitates analysing the image, and identifying objects within the image.

Every person has two lungs that are divided into lobes and associated fissure. Each lung contains an airway tree that according to some texts [1] consists of about 23 generation in the airways. Image segmentation algorithms are utilised in lung modelling to locate and extract out lung lobes as well as airways regions within CT or MRI lung images.

Lung modelling remains a challenging task as the current state-of-art lung imaging technology is still unable to detect the entire 23 generation of the lung airway tree [2, 3]. The resolution of each image and the spatial distance between two consecutive image slices are important factors dictating the accuracy and reliability of the formed model. Whilst the ill-defined characteristic of the image segmentation problem makes segmentation of lung images difficult, low resolution of lung images further increases the difficulty of the lung image segmentation. Although there exists a variety of image segmentation algorithms, some are not suitable for lung image segmentation, and do not perform well when applied to lung images.

The assessment of the performance of the image segmentation algorithms is not a straightforward task. The main challenge lies in the ill-defined nature of the segmentation problem [4], and the absence of an objective representation for segmentation results. This representation is called the ground truth which is a description of the result of an ideal segmentation process [5]. The ground truth is usually constructed through a semi-automated process in which a human expert often segments the images that are used in the assessment task.

This paper investigates six image segmentation algorithms, used in medical imaging, and their application to lung modelling. The algorithms are: normalised cuts, graph, region growing, watershed, Markov random field, and mean shift. The performance of the six segmentation algorithms is characterised through exploring the localisation as well as shape-accuracy of the boundary maps associated with the segmented regions within the input lung image and its associated ground truth. The measured segmentation accuracies as well as execution times of the six algorithms are then compared and discussed.

The paper is organised as follows. Section 2 describes the image segmentation algorithms whose performances are assessed in this work. Section 3 gives the details of the experiments that are carried out to evaluate the performance of the examined image segmentation algorithms. Section 4 discusses the measured performance of the six segmentation algorithms. Finally, the concluding remarks are given in Section 5.

II. IMAGE SEGMENTATION ALGORITHMS

Image segmentation plays a crucial role in lung modelling. The goal of image segmentation research is to increase the reliability, accuracy, precision, and to reduce the

computational cost of the algorithms. This paper investigates six image segmentation algorithms, used in medical imaging, and their application to lung modelling. The algorithms are: normalised cuts, graph, region growing, watershed, Markov random field, and mean shift. In the following, each of the stated algorithms is briefly described.

A. Normalised Cuts

Normalised cuts can be classified as a graph clustering approach [6]. Normalised cuts involves cutting the graph into two subsequent connected components where the cost of each cut is a division of the likeliness within each region. Normalised cuts is initiated to overcome the conventional graph-based-bias short-boundaries cut which results in small regions [6]. The normalised cuts algorithm avoids biasing by introducing a cost function called disassociation. This function measures the normalised cuts, but introduces another bias known as ratio cut which normalises the cost function by the length of the cut that is comparable to the weight partition [7]. However, the algorithm is slow for both planar graphs and non-planar graphs.

B. Graph

Graph based segmentation algorithm attempt to divide the graph into connected components corresponding to the clusters with large internal weights [8]. A set of graph is represented by $G = (V, E)$ where V symbolises the vertex of the image pixels and E corresponds to the edge weight between the two matching pixels of the two corresponding vertices. The method obtains similar elements in the same region, and dissimilar element in different regions. Therefore, the edges with the similar regions are comparatively low weights compared to dissimilar regions.

C. Region Growing

Region growing is a process that groups pixels into regions based on the predefined factor for growth. This approach requires a starting point known as "seed pixel" where the region attempts to adjoin similar neighbouring pixels, based on a homogeneity criterion, until the pixel contrasts are too disparate to the region to be added. Often the homogeneity criterion is based on specified features information or edges of the particular image [9]. Traditional region growing approaches randomly select the direction to grow the region where modern approaches used the edge information to place the seed pixel in the most accurate position. In order to prevent placing a seed in regions which is unstable, the position between regions' boundaries must be shunned. This region growing approach tends to be sensitive to the noise causing extracted regions to be detached or having holes [10]. [11] utilised partial volume effects to counter the detached regions.

D. Watershed

Watershed algorithm is a popular approach calculating the boundaries of the catchment basins which are situated at the high gradient point forming a gradient image. Homogeneous regions are clustered based on the edge detection and arithmetical morphology. Often, the watershed transform experiences over-segmentation problem resulting in the segmentation of unwanted regions. To overcome this issue,

marker based watershed transform is used rather than conventional watershed algorithm [12]. The markers are applied to the gradient image to avoid over-segmentation where it decreases the regional minima connecting them with the region of interest.

E. Markov Random Field

A Markov random field is a statistical model which uses a clustering technique such as k-means algorithm under a Bayesian prior estimation model [13]. This segmentation algorithm aims to obtain the posterior probability of the image data pixels. Accurate selection of the controlling strength of spatial interaction parameter is one of the difficulties that this approach faces. If the parameter is too high, it experiences a loss of the structural details. However, this approach suits applications in which the intensity in-homogeneities arise in texture properties.

F. Mean Shift

Mean shift algorithm conducts a preliminary pre-processing using filtering followed by clustering of the filtered image data points. The importance of filtering in the mean shift approach is to approximate the modes of the probability density function using kernel density estimation. The clustering algorithm groups each data point in the n-dimensional dataset through relating the peak of each point with the predefined probability density of the dataset. It includes a post-processing step that groups the modes and also their basin of attraction. In general, the mean shift approach is sensitive to the changes in the granularity and colour bandwidth features which results in a large change in the relative output image.

III. EXPERIMENTAL RESULTS

The six described image segmentation algorithms were employed in an empirical study to measure the performance of their application to 2D CT lung images for lung modelling. In this study, we have employed a subset of 2D CT lung images from the Early Lung Cancer Action Program (ELCAP) Public Lung Image Database [14] developed by Cornell University. This database contains CT lung images of 50 subjects. The number of image slices for different subjects varies. The average number of image slices per subject is around 260. The image slice thickness is 1.25 mm, and image size is 512×512 pixels. The images are in grayscale mode, stored in DICOM format.

The evaluation of the performance of the six segmentation algorithms was characterised using the Boundary-Displacement Evaluation (BDE) method [15, 16] that explores localisation as well as shape-precision of the segmented regions within a lung image. The boundary point set produced by the segmentation algorithm is named B, and the boundary point set associated with the ground truth is named G. Assume that a distance distribution signature from a boundary point set B_1 to another boundary point set B_2 is denoted by $D_{B_1}^{B_2}$ which is a function whose distribution represents discrepancy from B_1 to B_2. The distance from an arbitrary point x in B_1 to B_2 is the minimum absolute distance from x to all points in B_2, $d(x, B_2) = \min\{d_E(x, y)\}, \forall y \in B_2$, where d_E states the

Euclidean distance between points x and y. The discrepancy between B_1 to B_2 is given by the shape of the signature measured by its mean and standard deviation. Thus, $D_{B_1}^{B_2}$, with a near-zero mean and a small standard deviation, denotes a high quality segmentation. Two distance distribution signatures are employed, one from ground truth to the segmented, denoted by D_G^B, and the other from the segmented to the ground truth represented by D_B^G.

The procedure for conducting the experiments is as follows:

i. Five subjects were randomly selected from the ELCAP lung image database [14].

ii. Five original images were chosen from the set of available images for each subject (see Fig. 1). Since the number of image slices for different subjects varied, and also to ensure that images of different parts of the person's lung are included in the experiments, the following 5 images were selected. Assuming that n is the total number of images for the subject, the selected image slices were: $0.3 \times n$, $0.4 \times n$, $0.5 \times n$, $0.6 \times n$, $0.7 \times n$.

iii. Each original image was pre-processed in three steps (see Fig. 2). First, the image was histogram equalised. Next, the resulting image was median filtered. And finally, the filtered image was manually cleaned to retain only the regions containing the lung tissues. The resulting image was called "input image".

iv. The input images were presented to each of the six segmentation algorithms, and the segmented images were obtained. The segmentation parameters for each algorithm were varied through trial and error to obtain the parameters that minimised the BDE for the algorithm. In addition, the execution times were also recorded.

v. The boundary maps were produced for the region-based segmented images using an image gradient operator [17].

vi. The ground truths were constructed through a semi-automated process by the authors. Next, the boundary maps were also computed for all ground truth images (see Fig. 3).

vii. The boundary-based errors were calculated using each ground truth image as well as its six associated segmented images.

In our experiments with the normalised cuts, graph, region growing, Markov random field, and mean shift, we have employed publicly available source codes or executable binaries by Cour et. al.'s multi-scale normalised cuts [18], Felzenszwalb and Huttenlocher's efficient graph-based image segmentation [19], Deng and Manjunath's JSEG [20], Gradwohl and Kato's Markov random field [21], and Georgescu and Christoudias's EDISON [22], respectively. In our experiments with the watershed algorithm, the codes developed by the authors were used.
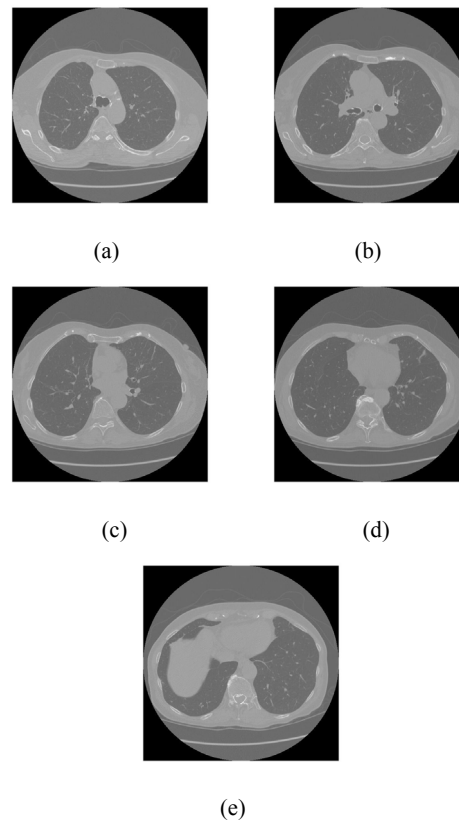

(a)   (b)

(c)   (d)

(e)

Figure 1: Original CT lung image of subject w0003 from ELCAP [14], slice: (a) 75, (b) 100, (c) 125, (d) 149, and (e) 174.
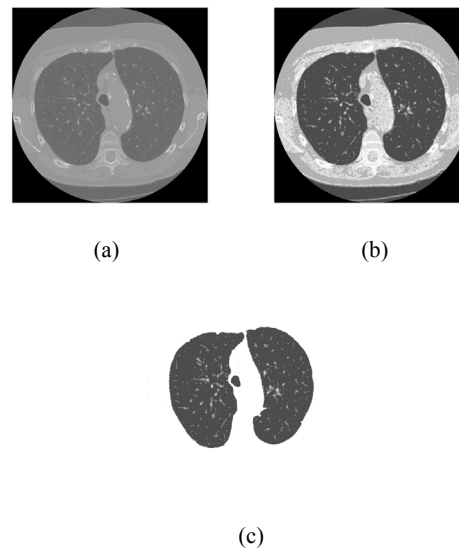

(a)   (b)

(c)

Figure 2: Pre-processing of the slice 73 of subject w0004: (a) original, (b) histogram-equalised-median-filtered, and (c) lung tissue extracted images.

Table I presents the evaluation results of the six examined segmentation methods in the form of the boundary displacement errors and also execution times. In the table, BDE represents the boundary displacement evaluation, and the execution times are in seconds. Figure 4 represents a comparison of the recorded results for various image slices.

(a)
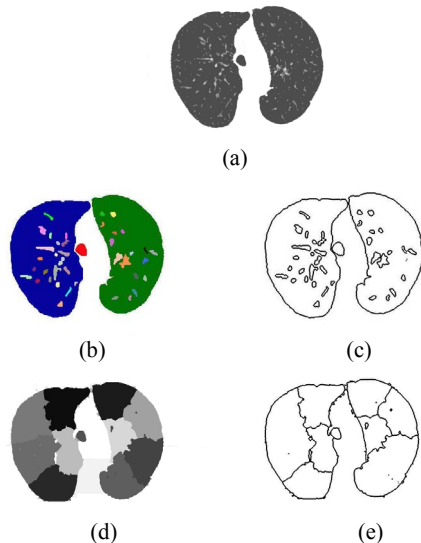
(b)   (c)

(d)   (e)

Figure 3: Sample images used in segmentation evaluation: (a) input, (b) ground truth, (c) boundary extracted ground truth, (d) normalised cuts segmented, (e) boundary extracted segmented images.

IV. DISCUSSIONS

Fig. 5(a) provides a comparison of the calculated average BDEs for the six segmentation algorithms. As can be seen from the figure, the mean shift algorithm was the best performer with an average BDE of only 0.47. The second best performer was the Markov random field algorithm with an average BDE of 1.26, followed by the watershed, region growing, and graph algorithms with a similar average BDE of approximately twice as that of the Markov random field algorithm. On the other hand, the normalised cuts was found to be the worst performer with an average BDE of almost 10 times higher than that of the best performer. It could hardly capture the lung airway regions, and also produced multiple regions for a single lung lobe. The described performance for the six segmentation algorithms was found to be consistent amongst all five subjects and their associated tested lung images.

The experiments were conducted on a desktop computer containing an Intel Core 2 CPU 6300 running at 1.86 GHz with 2GB of RAM. Four of the examined algorithms were executable binaries. However, normalised cuts as well as watershed algorithms were Matlab codes. The execution times were recorded during the course of the experiments (see Table I). Fig. 5(b) provides a comparison of the calculated average execution times for the six segmentation algorithms in log-linear form. As can be seen from the figure, the graph algorithm was the fastest with an average execution time of only 1.3 seconds. On the other hand, the normalised cuts

algorithm was the slowest with an average execution time of approximately 470 seconds. The mean shift algorithm that produced the least BDEs was only slightly slower than the graph, with an average execution time of 2.9 seconds. It should be however noted that the normalised cuts as well as watershed algorithms were tested in Matlab. They could be compiled to produce binary codes reducing their average computational time.

Considering the results of the evaluation of the six segmentation algorithms, it can be concluded that the mean shift method can be the best candidate for utilisation as the segmentation component in 2D and 3D lung modelling systems. It produces the lowest BDEs, and requires minimal CPU times.

V. CONCLUSION

This paper investigated six popular image segmentation algorithms and their application to lung modelling. The performance of the six segmentation algorithms was characterised through a set of experiments. The measured segmentation accuracies of the six algorithms are compared and discussed. An experimental procedure was devised to measure the performance of the tested algorithms. A subset of 2D CT lung images from the ELCAP database was used. The segmentation performances were characterised using the boundary displacement evaluation method. The execution times were all recorded. The mean shift algorithm was the best performer. It produced the lowest average BDE of 0.47 and required a small average CPU time of 2.99 seconds. It can be a good candidate for use in lung modelling systems

REFERENCES

[1] D. Karthikeyan, *High-Resolution Computed Tomography of the Lungs - A Pattern Approach*, 1st ed. London: Hodder Arnold, 2005.

[2] I. C. Sluimer, "Automated Image Analysis of Pathology Lung in CT," in *Image Sciences Institute*, vol. Masters Degree. Utrecht: University Medical Centre Utrecht, Sept. 2005.

[3] I. C. Sluimer, A. Schilham, M. Prokop, and B. Ginneken, "Computer Analysis of Computed Tomography Scans of the Lung: A Survey," *IEEE Transactions on Medical Imaging*, vol. 25, pp. 385-405, 2006.

[4] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "Toward Objective Evaluation of Image Segmentation Algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 929-944, June 2007.

[5] O. Okun and M. Pietikainen, "Fast and Accurate Ground Truth Generation for Skew-Tolerance Evaluation of Page Segmentation Algorithms," *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. 1-10, 2006.

[6] Z. Wu and R. Leahy, "An optimal graph theoretic approach to data clustering: theory and its application to image segmentation," *Transaction on Pattern Analysis and Machine Intelligence*, vol. 15, pp. 1101-1113, 1993.

[7] S. Wang and J. Siskind, "Image segmentation with ratio cut," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 675–690., 2003.

[8] D. Forsyth and J. Ponce, *Computer Vision - A Modern Approach*. London: Prentice Hall, 2003.

[9] R. M. Haralick and L. G. Shapiro, "Image segmentation techniques," presented at Proc. of Computer Vision Graph Image 1985.

[10] D. L. Pham, C. Xu, and J. L. Prince, "Current Methods in Medical Image Segmentation," *Annual Review of Biomedical Engineering*, vol. 2, pp. 315-337, 2000.

[11] M. Á. G. Ballester, A. P. Zisserman, and M. Brady, "Estimation of the partial volume effect in MRI," *Medical Image Analysis* vol. 6, 2002.

[12] R. Shojaii, J. Alirezai, and P. Babyn, "Automatic Lung Segmentation in CT images using Watershed Transform," in *Proc. ICIP*, vol. 2, Sept. 2005, pp. 1270-1273.

[13] Z. Kato, M. Berthod, and J. Zerubia, "Multiscale Markov random field models for parallel image classification," presented at Proc. of 4th International Conference on Computer Vision, 1993.

[14] "ELCAP Public Lung Image Database," Vision & Image Analysis Group (VIA) and International Early Lung Cancer Action Program (I-ELCAP) Labs, Cornell University, 2007.

[15] J. Freixenet, X. Muñoz, D. Raba, J. Martí, and X. Cufí, "Yet Another Survey on Image Segmentation: Region and Boundary Information Integration," in *Computer Vision - ECCV 2002: 7th European Conference on Computer Vision, Copenhagen, Denmark, May 28-31, 2002. Proceedings, Part III*. Berlin: Springer Berlin / Heidelberg, May 2002, pp. 21-25.

[16] Q. Huang and B. Dom, "Quantitative methods of evaluating image segmentation," presented at Proc., International Conference on Image Processing, Washington, DC, USA, 1995.

[17] A. Y. Yang, J. Wright, M. Yi, and S. Sastry, "Unsupervised Segmentation of Natural Images via Lossy Data Compression," University of California, Berkerley EECS-2006-195, 28 December 2006 December, 2006.

[18] T. Cour, F. Benezit, and J. Shi, "Multiscale Normalised Cuts Segmentation Toolbox for MATLAB," 1.0 ed. Philadelphia: Computer Science Department, University of Pennsylvania, 2006.

[19] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient Graph-Based Image Segmentation," 1 ed: Massachusetts Institute of Technology and Cornell University, 2004.

[20] Y. Deng and B. S. Manjunath, "JSEG." Santa Barbara: The Regents of the University of California, Samsung Electronics Corporation., 1999, pp. Unsupervised Segmentation of Color-Texture Regions in Images and Video.

[21] C. Gradwohl and Z. Kato, "Markov random field (MRF) based image segmentation algorithm ": Université de Nice-Sophia Antipolis, France, 2005.

[22] B. Georgescu and C. M. Christoudias, "Edge Detection and Image SegmentatiON (EDISON) System ", 1.1 ed. Piscataway: Robust Image Understanding Laboratory, Rutgers University, New Jersey, 2002.

BDE

(a)
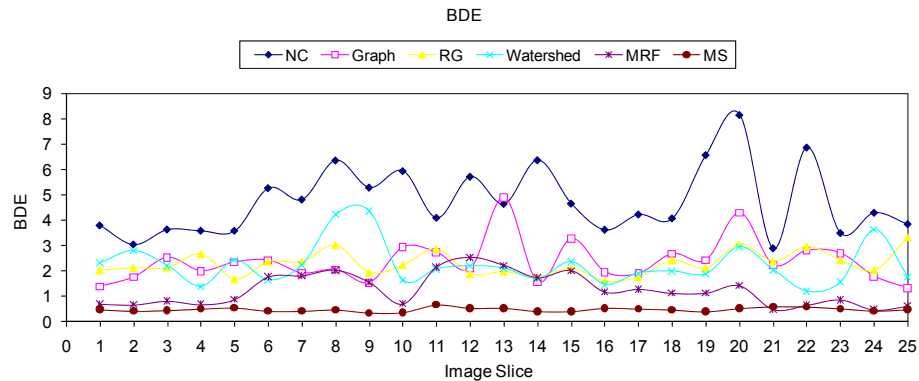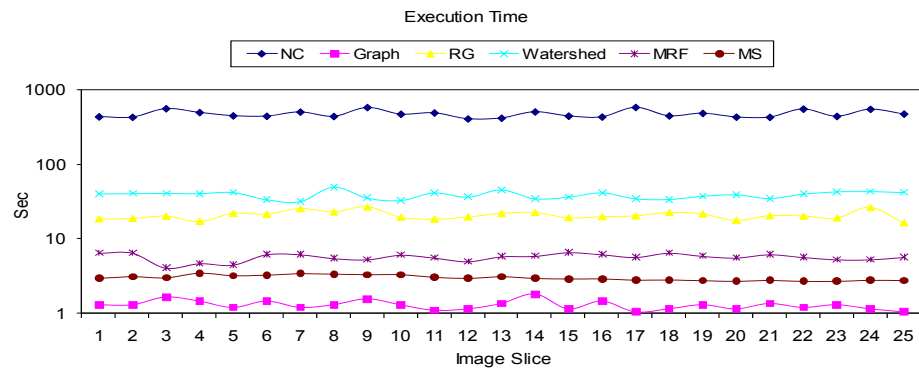
Execution Time

(b)
Figure 4: Experimental results: (a) BDEs, and (b) execution times in log-linear.

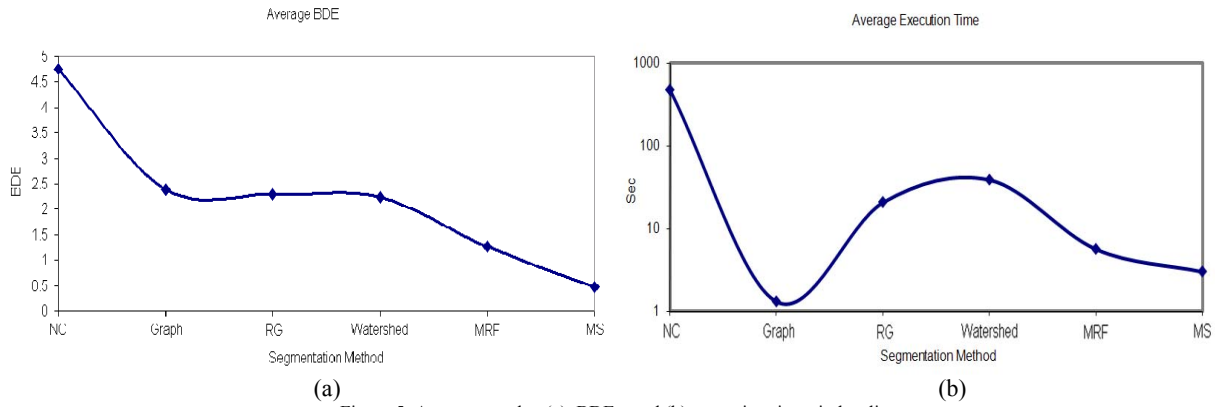(a)   (b)
Figure 5: Average results: (a) BDEs, and (b) execution times in log-linear.

TABLE I
EVALUATION RESULTS OF THE EXAMINED SEGMENTATION METHODS: BDE-BOUNDARY DISPLACEMENT EVALUATION; TIME- IN SECONDS.

Subject	CT Slice	Normalised cuts		Graph		Region growing		Watershed		Markov random field		Mean shift	
		BDE	Time	BDE	Time	BDE	Time	BDE	Time	BDE	Time	BDE	Time
W0003	75	3.80	432.10	1.40	1.30	2.02	18.44	2.32	40.00	0.69	6.40	0.47	2.95
	100	3.05	426.70	1.77	1.30	2.12	18.70	2.80	40.40	0.66	6.40	0.41	3.10
	125	3.64	549.70	2.54	1.64	2.13	20.10	2.20	40.50	0.81	4.00	0.44	3.00
	149	3.59	490.60	2.01	1.45	2.66	17.01	1.38	40.30	0.67	4.60	0.50	3.45
	174	3.59	444.90	2.36	1.20	1.67	21.86	2.42	41.60	0.88	4.40	0.54	3.20
W0004	73	5.27	439.40	2.43	1.45	2.36	21.33	1.63	33.50	1.77	6.10	0.41	3.25
	97	4.82	497.60	1.93	1.20	2.36	25.35	2.25	31.50	1.82	6.10	0.41	3.40
	122	6.36	436.70	2.06	1.30	3.02	22.93	4.23	49.20	2.04	5.40	0.46	3.35
	146	5.29	569.00	1.55	1.55	1.93	26.83	4.37	35.20	1.55	5.20	0.34	3.30
	170	5.94	466.80	2.95	1.30	2.23	19.46	1.63	32.60	0.70	6.00	0.36	3.30
W0005	82	4.10	484.40	2.74	1.10	2.86	18.26	2.08	41.20	2.17	5.50	0.66	3.05
	109	5.72	406.00	2.11	1.15	1.88	19.56	2.20	36.20	2.55	4.90	0.53	2.95
	137	4.65	415.10	4.92	1.35	1.97	21.86	2.10	45.10	2.23	5.70	0.53	3.10
	164	6.37	500.00	1.57	1.80	1.77	22.43	1.69	34.50	1.74	5.80	0.40	2.95
	191	4.66	441.60	3.28	1.15	2.10	19.09	2.36	36.10	2.02	6.50	0.40	2.90
W0006	77	3.63	429.70	1.97	1.45	1.60	19.75	1.46	41.30	1.17	6.10	0.53	2.90
	102	4.23	570.50	1.91	1.05	1.75	20.28	1.91	34.50	1.28	5.60	0.50	2.80
	128	4.08	443.90	2.69	1.15	2.42	22.46	1.99	33.70	1.12	6.40	0.46	2.81
	153	6.57	478.70	2.44	1.30	2.12	21.52	1.88	37.20	1.13	5.80	0.39	2.75
	179	8.15	428.50	4.31	1.15	3.05	17.64	2.96	38.90	1.42	5.50	0.52	2.70
W0010	86	2.90	426.50	2.25	1.35	2.38	20.36	2.04	34.80	0.48	6.10	0.59	2.77
	115	6.87	543.10	2.82	1.20	2.96	20.15	1.19	39.60	0.65	5.60	0.58	2.70
	144	3.50	438.00	2.71	1.30	2.42	18.98	1.57	42.60	0.86	5.20	0.51	2.70
	174	4.30	539.60	1.80	1.15	2.03	26.35	3.64	42.90	0.46	5.20	0.42	2.78
	202	3.85	469.10	1.33	1.05	3.33	16.28	1.76	41.70	0.61	5.60	0.47	2.76
Average		4.76	470.73	2.39	1.30	2.29	20.68	2.24	38.60	1.26	5.60	0.47	2.99