Chitsaz M, Woo CS. Software agent with reinforcement learning approach for medical image segmentation. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 26(2): 247–255 Mar. 2011. DOI 10.1007/s11390-011-9431-8

Software Agent with Reinforcement Learning Approach for Medical Image Segmentation

Mahsa Chitsaz and Chaw Seng Woo, *Member*, *IEEE*

Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia

E-mail: mchitsaz@siswa.um.edu.my; cswoo@um.edu.my

Received December 3, 2009; revised December 4, 2010.

Abstract Many image segmentation solutions are problem-based. Medical images have very similar grey level and texture among the interested objects. Therefore, medical image segmentation requires improvements although there have been researches done since the last few decades. We design a self-learning framework to extract several objects of interest simultaneously from Computed Tomography (CT) images. Our segmentation method has a learning phase that is based on reinforcement learning (RL) system. Each RL agent works on a particular sub-image of an input image to find a suitable value for each object in it. The RL system is define by state, action and reward. We defined some actions for each state in the sub-image. A reward function computes reward for each action of the RL agent. Finally, the valuable information, from discovering all states of the interest objects, will be stored in a Q-matrix and the final result can be applied in segmentation of similar images. The experimental results for cranial CT images demonstrated segmentation accuracy above 95%.

Keywords biomedical image segmentation, multi-agent systems, reinforcement learning system, CT images

1 Introduction

Image segmentation techniques have been an invaluable task in many domains, such as, quantification of tissue volumes, medical diagnosis, pathological localization, anatomical structure study, treatment planning, partial volume correction of functional imaging data, and computer integrated surgery[1]. Image segmentation separates an image into some disjoint partitions whereas the whole of partitions reconstruct the primary image. One of the well-known methods of image segmentation is region-growing that is based on the growth of a region whenever its interior is homogeneous according to certain features as intensity, color or texture. Besides, the K-means clustering algorithm clusters data by iteratively computing a mean intensity for each class and segmenting the image by classifying each pixel in the class with the closest mean. The fuzzy c-means algorithm generalizes K-means, allowing for soft segmentations based on fuzzy set theory. For more information regarding image segmentation methods refer to [1-3]. Image segmentation is still a debatable problem although there have been many researches in the last few decades[4]. First of all, every solution to image segmentation is problem-based. Secondly, medical image segmentation methods generally have restrictions because medical images have very similar gray level and texture among the interested objects. Therefore, significant segmentation error may occur. Another difficulty may arise due to the lack of sufficient training samples. For instance, some supervised segmentation methods require training samples that are prepared by field experts. Consequently, a more universal approach to the segmentation requires decreasing the level of user interaction and minimum training dataset.

Bearing in mind the above obstacles to medical image segmentation, we propose new algorithm based on reinforcement learning (RL). Agent can learn to perform segmentation over time by systematic trial and error[5]. The reinforcement learning agent is trained by obtaining rewards or punishment based on its action in an environment. Due to the dynamic nature of RL agent, it is suitable for segmenting images with high complexity[6]. The goal of the RL agent is to find out an optimal way to reach the best answer given some signals obtained after each action.

The state and action should be defined when using RL method in medical image segmentation; the number of regions in the sub-image identifies the states. Firstly, the agent takes an image and applies default threshold value for each region; this threshold value is calculated by dividing the gray-scale value by the number of regions. The input image is divided into several sub-images, and each RL agent works on it to find a

Short Paper

This research was funded by Peruntukan Penyelidikan Pascasiswazah (PPP) under Grant No. PS349/2008C.

suitable value for each object in the image. Each state in the environment is associated with some actions; and a reward function computes reward for each action of the RL agent. Therefore, the agent tries to learn which actions can gain the highest reward. Finally, the gained valuable information will be used to segment other similar images.

The main purpose of this work is to segment CT images simultaneously with some different regions of interest. This is a significant advantage compared to other approaches because it can segment many objects within an image concurrently. In addition, our method does not need a large training set or priori-knowledge.

In this section, we present a short description of RL system. Section 2 is a brief summary of recent image segmentation methods based on RL system. Section 3 gives the details of the approach and discusses algorithms used in our work. Section 4 analyses the experimental results. Section 5 discusses the experimental results. Finally, last section concludes our work.

1.1 Reinforcement-Learning System

Learning to act in ways that are rewarded is a sign of intelligence. For example, it is natural to train an elephant in circus by rewarding it when it correctly acts in reaction to a command, or punishing it when it does not correctly act. This has been studied in experimental psychology[7]. In the standard reinforcement learning model, an agent interacts with its environment via perception and action as depicted in Fig.1. In each step of interaction the agent receives input, i, the current state, s, of the environment; the agent then chooses an action, a, to generate an output. The action changes the state of the environment and the value of this state transition is then received by the agent through reinforcement signal (reward/punishment), r. The agent's behavior, B, should choose actions that tend to increase the overall sum of values of the rewards. The figure also includes an input function I, which determines how the agent views the environment state. R is a set of scalar reinforcement signals. Finally, T is a set of tasks that the agent will perform in the environment.

Agent can learn to do this over time by systematic trial and error[5]. The reinforcement learning agent is trained by obtaining rewards or punishment based on its action performed in the environment.

It is important that the agent gathers useful experience about the possible system states, actions, rewards and punishment actively to act optimally.

Q-learning[7] is a recent form of RL algorithm. The values of state-action pairs will be estimated by Q-learning algorithm, and stored in Q-matrix. The value $Q(s,a)$ is the expected sum of future payoffs, r,

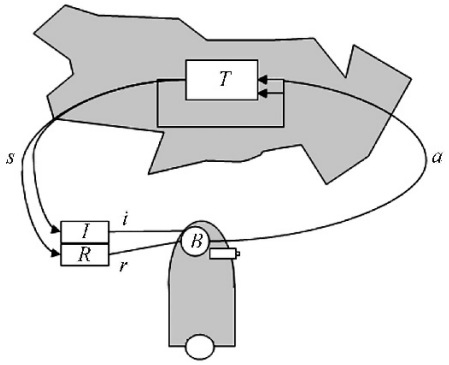

Fig.1. Standard RL model. The agent receives input i and current state s of environment, then based on its behaviour B, it does an action a that receives a reward, r[5].

obtained by taking action, a, from state, s. Once these values have been learned, the optimal action from any state is the one with the highest Q-value. At the beginning Q-matrix is initialized to arbitrary numbers, Q-values are estimated on the basis of experience as follows.

1) From the current state s, select an action a. This will cause a receipt of an immediate payoff, r, and coming at a next state, s'.

2) Update $Q(s,a)$ based on this experience as follows:

$$Q(s,a) = (1-\alpha)Q(s,a) + \alpha[r + \gamma \max_{a'} Q(s',a')] \quad (1)$$

where α is the learning rate and $0 < \gamma < 1$ is the discount factor.

3) Go to step 1).

The parameters α and γ should be between 0 and 1 as mentioned in (1). The learning rate (α) determines to what extent the newly acquired information will override the old information. A factor of 0 will make the agent not learn anything, while a factor of 1 would make the agent consider only the most recent information. The discount factor (γ) determines the importance of future rewards. A factor of 0 will make the agent opportunistic by only considering current rewards, while a factor approaching 1 will make it strive for a long-term high reward. If the discount factor meets or exceeds 1, the Q values will diverge[8]. Therefore, we chose 0.1 for α and 0.9 for γ. In ε-Greedy, the best lever is selected for a proportion $1-\varepsilon$ of the trials, and another lever is randomly selected (with uniform probability) for a proportion ε. The parameter value can vary widely depending on circumstances and predilections, here we assign 0.7 to ε. This algorithm is guaranteed to converge to the correct Q-values with the probability one if the environment is stationary and depends on the current state with the action taken in

it. A lookup table (Q-matrix) is used to store the Q-values, every state-action pair continues to be visited, and the learning rate is decreased appropriately over time. This exploration strategy does not specify which action to select at each step. In practice, a method of choosing action is usually to ensure sufficient exploration will be done to reach a steady state that chosen actions are optimal.

2 Literature Review

Many new methods can overcome the drawbacks of existing image segmentation methods. However each method has been further developed to produce better results[9]. A few researchers such as Peng, Shokri and Sahba used the RL agent in segmenting images.

Peng and Bhanu[10-12] proposed a framework for object recognition using RL approach. Some preprocessing steps are needed to achieve successful object recognition, like segmentation and feature extraction. The algorithm used for segmentation is Phoenix Segmentation Algorithm. This algorithm is based on a recursive region splitting. It uses information from histogram of red, green and blue image components to split the region on the basis of a peak/valley analysis of each histogram. The evaluation framework for object recognition is proposed by RL. As a result of the proposed method, the system is capable of exploring a significant portion of the search space, resulting in the discovery of good solutions due to the stochastic nature of RL. In general, this result cannot be achieved by any deterministic or simple supervised learning methods.

Shokri and Tizhoosh[13-14] used the concept of the RL system for finding the best thresholding of image. The model has states, actions and a matrix that saved the reward or the punishment. The RL agent starts with a constant threshold and applies it to the image. The gray level could be a random number in the range 0 to 255. One may select the initial threshold using existing thresholding techniques. The state is based on the ratio of black pixels/total pixels, and the number of objects in the image. The action is finding an optimal thresholding range. There are two models of the reward/punishment; subjective and objective. Subjective case means an experienced user will assign a reward/punishment to the outcome image. And objective case is defined based on the black pixel ratio, the area of object, the tolerance for area deviation, and the number of the objects. The proposed method achieves accuracy of 87% for subjective method, and 60% for objective method. This method needs considerable user interaction to achieve a better performance.

Sahba *et al.*[15-16] proposed an RL model to segment an ultrasound image of the prostate. First, the image is divided into some sub-images. Then, agents find the optimal threshold of all sub-images. After completing the segmentation of all sub-images, the reward/punishment is assigned to the action of every agent via objective model or subjective model. After training, the agent finds the best threshold for the image and can segment another image of similar type. Furthermore, researchers have used a deformable model for extracting prostate from ultrasound image.

Table 1 summarizes the achievements of abovementioned methods. Outdoor (indoor) image is an image that is taken in open (close) environment. Synthetic images are combination of two or more parts which are constructed by human. Ultrasound image is a type of 3D imaging which is noisier than the other imaging techniques. Table 1 shows no research has been done to segment CT images although some researchers applied their method to different image modalities such as outdoor/indoor, synthetic and ultrasound images. Moreover, they used a narrow range of images such as the prostate. On the other hand, we test many images covering the upper half of a human body. Finally, the best accuracy of previous methods is limited to 91%.

Table 1. Achievement of RL-Based Segmentation Method (Achievement shows that all researches have promising results that illustrate the possibility of RL-based method in image segmentation.)

Researchers	Image Size and Modality	Achievement
Peng and Bhanu	Outdoor and indoor images	The use of RL algorithm as part of the evaluation function for image segmentation gives rise to significant improvement of the system performance.
Shokri and Tizhoosh	Synthetic images	Achieved performance of 87% for subjective method, and 60% for objective method.
Sahba *et al.*	Ultrasound image of prostate	The mean quality percentage is equal to 90.65%.

Therefore, we propose a new method using the RL agent to achieve better result in comparison with the mentioned researches.

3 Methodology

Although RL agents have been used in some image processing tasks, according to the work in [11-14, 16], the application of Q-learning in image segmentation has not been explored until recent years. We show that RL agent is suitable for medical image segmentation where several regions are processed simultaneously. This method is specifically useful for medical images where there are several images of a patient that have very similar characteristics. In such a case, some of the

images can be used as training image. Then, the appropriate parameter can be found for segmenting the other similar images. In the proposed method, as shown in Fig.2, an image is divided into several sub-images. First of all, the number of interest regions should be entered by a user.

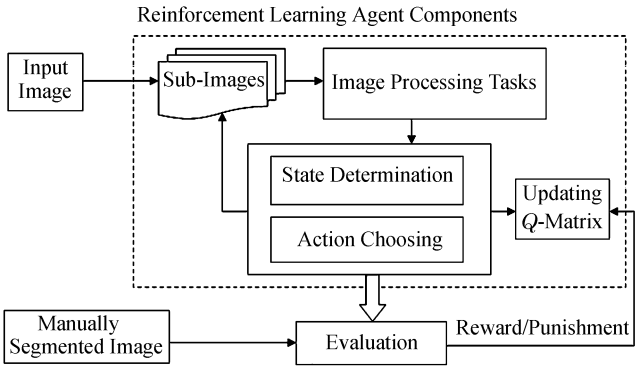

Fig.2. Global view of our proposed method. RL agent processes sub-images, after the state is determined, agent chooses an action to do changes in sub-images and updates (s, a) pair in Q-matrix. There is another component called evaluation that compares the result of segmented image from RL agent to manually segmented image to evaluate RL agent work and gives it a reward.

The RL agent determines the local thresholding value for each individual sub-image via dividing the maximum gray-scale value of the input image by the given number of objects within the image. The image processing task in Fig.2 refers to a set of actions as follows: apply thresholding to each sub-image, then process the sub-images, and then provide the state and action for each of them. The RL agent needs three components to learn from its environment, i.e., states, actions and reward. The Q-matrix is then constructed with regard to states and actions.

The RL agent starts its work using an image and its corresponding desired results. Figs. 3(a) and 3(b) show a cranial CT image and its manually segmented version. They are used as an input for the RL agent to obtain segmentation knowledge. The agent starts to find the

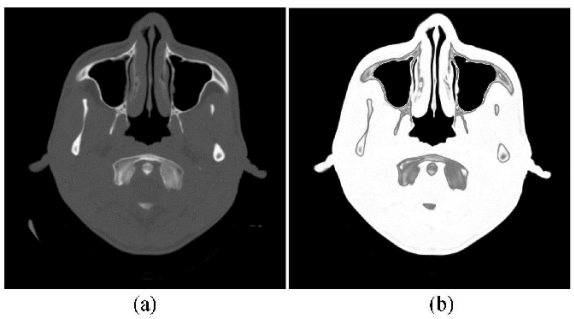

Fig.3. (a) Original image. (b) Manually segmented image.

appropriate state of the image, after that it chooses one defined action.

During this time the RL agent changes the local thresholding values for each sub-image individually. By taking each action the agent receives corresponding reward for that state-action pair and updates the corresponding value in the Q-matrix. The RL agent explores many actions in this cycle and tries to exploit the most rewarding actions.

Fig.4 shows the flowchart of RL agent behavior. At the beginning, the RL agent discovers all pixels in the sub-image, and marks it based on some fixed thresholding range. This thresholding information is acquired from dividing the maximum gray-scale, 256, by the number of acquired regions within an image. As an example, if the image has three objects for segmenting, the first thresholding range will be [0, 85] for the first region, [85, 170] for the second one, and [170, 256] for the third region. The size of sub-image is randomly chosen using trial-and-error during system implementation. This size can be any small number but we found out 7×7 is appropriate to limit total number of agents for each sub-image. Moreover, the size should be odd because the agent is placed in the center of each sub-image. Besides, the size of sub-image is more important in computation time. If this size is too small, there will be too many agents in the sub-image for processing,

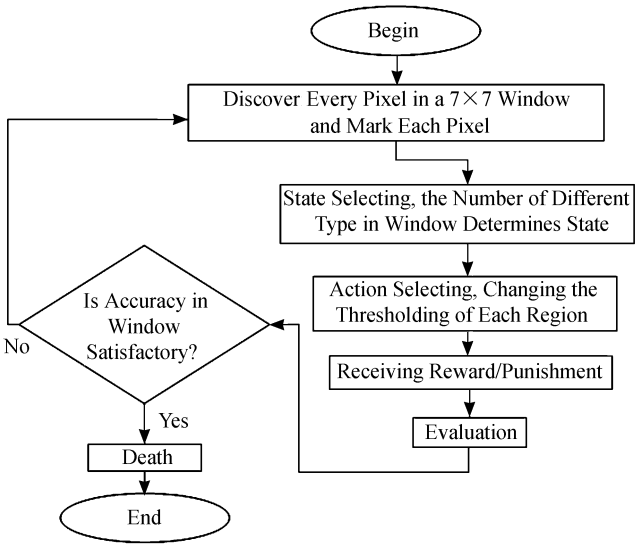

Fig.4. RL agent's behavior. RL agent discovers its window and starts to mark any pixel in it. After that, state of window will be determined based on the state an action will be chosen. This state-action pair results in a reward. After the whole window has been marked by RL agent, a globally evaluation will be done to decide how satisfactory the accuracy of current window is. If RL agent has achieved the satisfactory accuracy, its lifetime will be terminated.

and after a while they will be vanished, therefore too much computation will be required to produce agents and kill agents. On the contrary, if this size is too big, then the collaboration between agents would not be satisfied because in most of the pixels, there is one agent to decide without any neighbors.

After marking all pixels in the window, the RL agent finds its state. Subsequently, there are some actions for each state; it should select one of them. ε-Greedy is a method which helps the RL agent to choose better action, where ε is a probability to choose action with most Q-value. If ε is less than a predefined parameter, then the RL agent selects an action with the most Q-value from Q-matrix which is used to store the Q-values, otherwise it selects action randomly. After choosing the action, the RL agent alters the primary thresholding value of each region by means of maximum and minimum thresholding in the current sub-image. A reward function calculates the number of true segmented pixels. After that Q-matrix saves the information of this state-action pair using (1). Following that, the RL agent work is evaluated, if the result is satisfactory, it means the accuracy of segmented window is more than a fixed number, for example 95%, then the RL agent life time is finished because its duty is done. The accuracy is a fixed number, because sometimes the RL agent could not reach the goal of 100% accuracy.

3.1 Definition of States

The number of regions in the sub-image identifies the states. In each sub-image, there are pixels each of which should be marked as a specific region. For example, in Fig.3, there are three different regions, i.e., bone, skin, and air. Therefore, the state is the number of identified regions in a sub-image. Consequently, the numbers of states are seven described as follows:

1) the sub-image that includes the pixels of the first region type;

2) the sub-image that includes the pixels of the second region type;

3) the sub-image that includes the pixels of the third region type;

4) the sub-image that includes the pixels of both first and second region types;

5) the sub-image that includes the pixels of both first and third region types;

6) the sub-image that includes the pixels of both second and third region types;

7) the sub-image that includes the pixels of all region types.

3.2 Definition of Actions

Actions change the local thresholding value of each sub-image. There are some actions for each state. An action employs the maximum and minimum gray-scale value in the sub-image. This distance can be divided into several intervals with a predefined parameter. Each action is defined as one of the intervals of this distance to threshold the sub-image. The number of actions depends on the predefined parameter, and also the state. For example, if the predefined parameter is 2, and maximum and minimum gray-scale values of sub-image are 25, and 27 respectively, the actions for the fourth, fifth or sixth state of previous subsection are to choose one of these sets {[25, 25], [25, 27]}, {[25, 26], [26, 27]}, or {[25, 27], [27, 27]}, thus the total number of actions is three.

If the minimum and maximum values are similar for the action interval of n, it means pixels in the sub-image do not include the region number of n. For example, the first interval in the first set of the previous example is [25, 25]. The minimum and maximum values are similar. It means there are no pixels which can be marked by first region type in the sub-image, so all the pixels are from the second region type. For another example, in this set {[25, 27], [27, 27]}, there is an interval with the same value at the first and the last interval, so it means there are no pixels with the second region type in the sub-image. In other words, if the minimum and maximum values are similar for the action interval of n, it means there is no pixels can be marked as region type of n. This feature helps the RL agent to change its state to another state when the current state is not correct.

3.3 Definition of Reward

The reward should show how well the image is segmented by the RL agent. As a result, an appropriate segmented image is needed for evaluation in lieu of true delineation. The reward function is defined based on the number of pixels which are segmented correctly.

4 Experimental Result

The experimental results of the proposed method have been considered qualitatively and quantitatively through image display and experiment measurements respectively. There are two different CT image sets from human body. In the first experiment, head CT images[17] are acquired on a CT scanner with an image size of 512×512, and a pixel size of $0.5\,\text{mm} \times 0.5\,\text{mm}$. Upper human body CT images for the second experiment[18] are acquired on the same machine. The used imaging protocol has image size of 512×512, and a pixel size of $0.55\,\text{mm} \times 0.55\,\text{mm}$.

We used ε-Greedy method for choosing an action, where ε was placed at 0.7. Also, α and γ, parameters

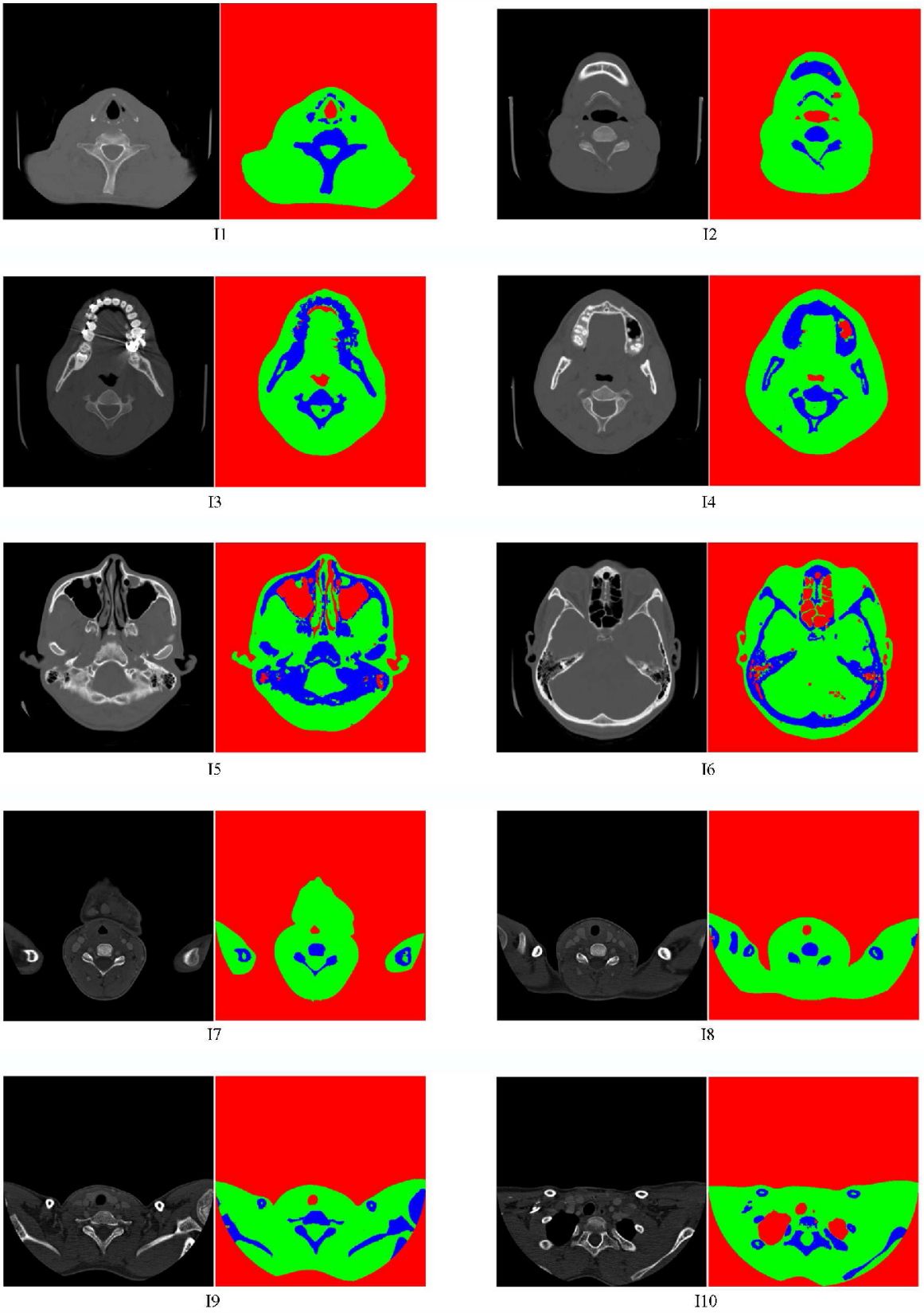

Fig.5. The left and the right images in each sub-figure show the original image and segmented image by our propose method, respectively. The accuracy for each image (I1∼I10) is illustrated in Table 2.

Table 2. TPVF and FPVF for Each Image Sample I1∼I10 That Is Shown in Fig.5

Dataset	TPVF (%)			FPVF (%)		
	BG	Skin	Bone	BG	Skin	Bone
I1	99.99	97.44	98.34	0.36	0.10	0.56
I2	99.99	95.76	94.74	0.45	0.37	0.92
I3	99.99	97.55	97.66	0.55	0.18	0.47
I4	99.99	98.83	97.87	0.13	0.22	0.34
I5	99.86	96.85	96.43	0.41	0.63	0.91
I6	99.86	97.79	97.11	0.37	0.46	0.75
I7	99.99	99.74	96.52	0.11	0.09	0.03
I8	99.99	99.69	96.80	0.18	0.09	0.04
I9	99.99	99.60	97.19	0.15	0.21	0.07
I10	99.97	99.46	98.29	0.27	0.12	0.08

in (1), are set to 0.1 and 0.9 respectively. Another fixed parameter employed in this framework is the number of iteration to examine every action of a specified state, it was set to 200. The dataset tested consists of a series of medical images with image size 512×512 pixels.

A subjective inspection discovered that in all experiments and in all data, the results are very close to the manually segmented images. Some examples are displayed in Fig.5; meanwhile, initialization was done in identical manner for all experimental images to evaluate the results of input images.

Quantitative segmentation evaluation has been used to assess segmentation methods[19]. The accuracy of a segmentation technique refers to how far an actually segmented image is from the manually segmented image. As a result, an appropriate segmented truth is needed for evaluation. In all experiments, all datasets have been manually segmented in the domain. For any image $A = (C, f)$, where C is a 2D (or higher-dimensional) rectangular pixels array, and $f(c)$ denotes the intensity of any pixel c in C, let C_d^M be the segmentation result obtained from C, and C_{td} is the true delineation. U_d is a binary image representation of a reference superset of pixels that is used to express the two measures as a fraction. We used true positive volume fraction (TPVF) and false positive volume fraction (FPVF)[20].

These equations are sufficient to describe the accuracy of the proposed method:

$$TPVF_d^M = \frac{|C_d^M \cap C_{td}|}{|C_{td}|} \times 100 \tag{2}$$

$$FPVF_d^M = \frac{|C_d^M - C_{td}|}{|U_d - C_{td}|} \times 100. \tag{3}$$

Table 2 lists TPVF and FPVF achieved in our experiment for each image in Fig.5. The TPVFs of all datasets are above 95%, and their FPVFs do not exceed 0.9%.

The efficiency of segmentation method provides information on the sensible use of the proposed algorithm. Table 3 contains the mean computation time for each image in Fig.5. For every image, the program ran 15 times. The proposed method was implemented on 2.00 GHz Intel Core 2 Duo and 2.00 GB RAM. The reward function needs the manually segmented image of a current image, therefore, it takes in average 10 minutes to segment an image by means of imaging software.

Table 3. Efficiency of Each Image Sample I1∼I10 That Is Shown in Fig.5

Dataset	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10
Computation Time (s)	10	12	12	12	14	15	8	6	7	7

5 Discussion

Quantitative comparison of the proposed method is not intended; however, having the manually segmented image side by side gives an opportunity to consider the advantages of the proposed method. Furthermore, the qualitative comparison shows accurate result. The proposed method is almost automatic; it has just required the manually segmented image for the reward function. The most significant advantage of proposed method is segmenting image to more than two regions in parallel way.

It means the regions of interest can be more than one and with different characteristics. For example, the CT image of head consists of three different regions, such as air, bone, and skin, the proposed method segments the image to three different objects simultaneously. Also the number of training dataset decreased in comparison to the neural networks, and the other learning method. The efficiency illustrates this method is too fast in comparison to the other method. In Table 4, the efficiency of some segmentation methods is listed.

The proposed method focuses on a simple yet efficient approach in segmentation by omitting other image features such as texture and color. Although only gray-scale value is used in segmenting CT images, the results of the proposed method are promising. The point is that with this feature we can obtain results quickly and accurately. However, the qualitative result shows accuracy of proposed method; but the method does not work for all images in the dataset because of some reasons. First of all, at the beginning of algorithm, some predetermined conditions, such as the number of iteration to fill the Q-matrix, ε, α, and γ, had been set fixedly, therefore, these conditions could not change in the duration of program running.

Table 4. Efficiency Comparison Among Different Image Segmentation Methods

Researcher	Method	Dataset	PC Specification	Efficiency
T. Liang, J. J. Rodriguez[20]	Fuzzy C-mean	MR images of a patient's head	Sun SPARCstation 10/50	Pixel-based: 9.7 min Region-based: 0.73 min
Z. Pan[21]	Region-growing	CT image of skull and liver	Win2k and VC++ 6.0 platforms	Skull: more than 10 s Liver: less than 5 s
H. Lu[22]	Extended image force model of snakes	Heart and Lung gradient image	N/A	Heart (160×169): 8.3 s Lung (225×211): 30.2 s
Chitsaz and Woo	Reinforcement learning	2 different datasets of the CT image (512×512)	2.00 GHz Intel Core 2 Duo and Java platform	First dataset: 13 s Second dataset: 7 s

Consequently, maybe these predetermined conditions should be changed for some specific images in duration of program running. For example, the number of iteration have been set to 200 times in our implementation, for images with such a narrow histogram, this number is not sufficient to fill all the cells in Q-matrix. Moreover, states are defined based on gray-scale value; this should be improved to cover more image features, like texture, shape, or special information, in future.

Besides, the reward function can be changed to a totally autonomous method. Finally, the number of actions for each state is rigid. For each window of image which covers the more distance of gray-scale, this amount of actions is not satisfactory, because the distance is too long and finding the appropriate gray-scale for each region in window is not possible. Meanwhile, the chosen method of action is ε-Greedy in our framework; it can be changed to a more comprehensive method.

6 Conclusion

The proposed method utilizes standard RL model to achieve segmentation. State, action, and reward function are defined where the RL agents use them to learn from the image. Therefore, every state of the environment and image has associated actions. The RL agent in each situation decides to choose one action. As a result, the image is marked by the RL agent; it means each pixel in sub-images is labeled as a specific region of image such as bone or skin. Finally, a reward function evaluates the accuracy of the segmented image, and gives a reward signal to the RL agent.

Our proposed method is almost automatic; it works without user interaction in segmenting the image. The most significant advantage of this method is segmenting image into more than two regions in a parallel way, it means the detected regions can have different characteristics when the image is segmenting. Also, the number of training dataset decreased in comparison to the artificial neural network-based methods. The efficiency of our method illustrated in Table 3 is significantly high compared with other methods that listed in Table 4. We have shown that the method can be used to segment different anatomic structure in medical images. The main results of this method are summarized below.

- We attained significant result in segmentation accuracy; the accuracy is more than 95% for each region in the images.
- We achieved satisfactory result in computation time; the mean computation time of all datasets is less than 13 seconds.
- The number of training dataset is minimal.
- We have the ability to segment simultaneously an image into some distinct regions and thus saving processing time.

Some improvements can be done in the future. States are defined based on gray-scale value; this should be improved to cover more image features such as texture and shape. Besides, the reward function can be changed in order to achieve total autonomy. Finally, the number of actions for each state is rigid; it can be changed for future work.

Acknowledgements The authors wish to thank the University of Malaya for postgraduate fellowship.

References

[1] Pham D L, Xu C, Prince J L. A survey of current methods in medical image segmentation. *Annual Review of Biomedical Engineering*, 2000, 2: 315-337.

[2] Jain A K. Fundamentals of Digital Image Processing. Printice Hall, 1989.

[3] Chen P, Pavlidis T. Image segmentation as an estimation problem. *Computer Graphics and Image Processing*, 1980, 12(20): 153-172.

[4] Liu J. Synergistic hybrid image segmentation: Combining model and image-based Sstartegies [Ph.D. Dissertation]. Univ. Pennsylvania, 2006.

[5] Kaelbling L P, Littman M L, Moore A W. Reinforcement learning: A survey. *Artificial Intelligence Research*, 1996, 4: 237-285.

[6] Chitsaz M, Woo C S. Medical image segmentation by using reinforcement learning agent. In *Proc. International Conference on Digital Image Processing* (*ICDIP 2009*), Bangkok, Thailand, Mar. 7-10, 2009, pp.216-219.

[7] Watkins C J C H. Learning from delayed rewards [Ph.D. Dissertation]. Cambridge, 1989.

[8] Watkins C, Dayan P. *Q*-learning. *Machine Learning*, 1992, 8(3): 279-292.

[9] Chitsaz M, Woo C S. The rise of multi-agent and R.L. segmentation methods for biomedical images. In *Proc. The 4th Malaysian Software Engineering Conference* (*MySEC 2008*), Kuala Terengganu, Malaysia, 2008, p.5.

[10] Bhanu B, Peng J. Adaptive integrated image segmentation and object recognition. *IEEE Transaction on Systems, Man, and Cybernetics*, 2000, 30(4): 427-441.

[11] Peng J, Bhanu B. Delayed reinforcement learning for adaptive image segmentation and feature extraction. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 1998, 28(3): 482-488.

[12] Peng J, Bhanu B. Closed-loop object recognition using reinforcement learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, 20(2): 139-154.

[13] Shokri M, Tizhoosh H R. Using reinforcement learning for image thresholding. In *Proc. IEEE Canadian Conference on Electrical and Computer Engineering*, May 4-7, 2003, pp.1231-1234.

[14] Shokri M, Tizhoosh H R. A reinforcement agent for threshold fusion. *Applied Soft Computing*, 2008, 8(1): 174-181.

[15] Sahba F, Tizhoosh H R, Salama M M A. A reinforcement learning framework for medical image segmentation. In *Proc. International Joint Conference on Neural Networks*, Vancouver, Canada, Jul. 16-21, 2006, pp.511-517.

[16] Sahba F, Tizhoosh H R, Salama M M A. A reinforcement agent for object segmentation in ultrasound images. *Expert Systems with Applications*, 2008, 35(3): 772-780.

[17] Obaidellah U H B. A finite element approach for the planning and simulation of 3D mandibular osteotomy for orthoganathic surgery [Master Thesis]. Faculty of Computer Science and Information Technology, University of Malya, 2006.

[18] DICOMsample: DICOM Files. July 2008, http://pubimage.hcuge.ch:8080/.

[19] Zhang Y J. A review of recent evaluation methods for image segmentation. In *Proc. The Sixth International Symposium on Signal Processing and Its Applications*, Kuala Lumpur, Malaysia, Aug. 13-16, 2001, pp.148-151.

[20] Te-shen Liang, Rodriguez J J. MR cranial image segmentation — A morphological and clustering approach. In *Proc. IEEE Southwest Symp. Image Analysis and Interpretation*, San Antonio, USA, Apr. 8-19, 1996, pp.184-189.

[21] Pan Z, Lu J. A Bayes-based region-growing algorithm for medical image segmentation. *Computing in Science & Engineering*, 2007, 9(4): 32-38.

[22] Lu H, Bao S. An extended image force model of snakes for medical image segmentation and smoothing. In *Proc. The 8th International Conference on Signal Processing*, Beijing, China, Nov. 16-20, 2006.

Mahsa Chitsaz received her B.S. degree in compt. eng. from Shiraz University in 2006. She then obtained her M.Sc. degree from University of Malaya in 2010 under the guidance of Chaw Seng Woo in the area of reinforcement learning in medical image segmentation. Chitsaz's research interest is in artificial intelligence in real-time systems, mainly in the areas of machine learning and dynamic processes. She also works at the intersection of learning and topics as varied as medical image segmentation, telemedicine, and head-mounted display.

Chaw Seng Woo is a senior lecturer at the Faculty of Computer Science and Information Technology, University of Malaya. His research interests include image processing and mobile applications.