Computers in Biology and Medicine 41 (2011) 56–65

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Tumor segmentation from computed tomography image data using a probabilistic pixel selection approach

Jung Leng Foo [a,*], Go Miyano [b], Thom Lobe [c], Eliot Winer [d]

[a] 1620 Howe Hall, Virtual Reality Applications Center, Iowa State University, Ames, IA 50011, USA
[b] Pediatric General and Urogenital Surgery Department, Juntendo University School of Medicine, Tokyo, Japan
[c] Pediatric Surgeon, Pediatric Surgery Department, Blank's Children Hospital, Des Moines, IA, USA
[d] Mechanical Engineering and Human–Computer Interaction Department, 1620 Howe Hall, Virtual Reality Applications Center, Iowa State University, Ames, IA 50011, USA

ARTICLE INFO

Article history:
Received 30 July 2009
Accepted 24 November 2010

Keywords:
Automatic seed relocation
Computed tomography
Image segmentation
Probabilistic selection
Tumor segmentation

ABSTRACT

Automatic segmentation of tumors is a complicated and difficult process as most tumors are rarely clearly delineated from healthy tissues. A new method for probabilistic segmentation to efficiently segment tumors within CT data and to improve the use of digital medical data in diagnosis has been developed. Image data are first enhanced by manually setting the appropriate window center and width, and if needed a sharpening or noise removal filter is applied. To initialize the segmentation process, a user places a seed point within the object of interest and defines a search region for segmentation. Based on the pixels' spatial and intensity properties, a probabilistic selection criterion is used to extract pixels with a high probability of belonging to the object. To facilitate the segmentation of multiple slices, an automatic seed selection algorithm was developed to keep the seeds in the object as its shape and/or location changes between consecutive slices. The seed selection algorithm performs a greedy search by searching for pixels with matching intensity close to the location of the original seed point. A total of ten CT datasets were used as test cases, each with varying difficulty in terms of automatic segmentation. Five test cases had mean false positive error rates less than 10%, and four test cases had mean false negative error rates less than 10% when compared to manual segmentation of those tumors by radiologists.

1. Introduction

The introduction of multi-slice computed tomography (CT) technology has established a method of acquiring and visualizing high resolution medical image data in a noninvasive manner. This has substantially improved the diagnosis of many ailments including tumor detection and removal. An important factor is the ability to properly segregate the tumor from the rest of the healthy organs in a patient. This can be a difficult task due to the complexity of the images and the irregular shapes and sizes of tumors. If objects within a dataset could be removed (i.e. segmented) and displayed individually, medical professionals would have powerful 2D and 3D visual representations to make crucial decisions in diagnosis and treatment. This capability would offer tremendous information not previously available that would allow errors in medicine to be reduced. Knowing the approximate size and shape of the tumor helps the surgeons for pre-operative planning and also offers doctors a way to track a patient's treatment progress by comparing segmented tumors from CT exams taken at different times.

This paper presents a novel approach to automate segmentation of tumors in CT image data using probabilities generated based on pixels' intensity and spatial properties. The segmentation procedure is performed two-dimensionally on a slice-to-slice basis, and an automatic procedure has been developed to facilitate the repositioning of the seed and search region. The next section of the paper presents a review of current related research in medical image segmentation, followed by a description of the algorithm. The results of 10 CT exams used as test cases are evaluated and discussed.

Statistical and probability based methods have been used with outstanding results as evident in the fields of heuristic optimization [1,2] and uncertainty decision modeling [3,4]. These implementations use probabilities that are generated in real-time as computations progress. The concept of using heuristic and probabilistic optimization algorithms has proven to be a success in various fields such as network traffic routing [5], water distribution design and planning [6], engineering product design [7], and protein designs [8]. An optimization algorithm is usually used to steer the design process towards an optimum design with the highest probability for success or towards a design with the lowest uncertainty probability. When used in engineering design, these probability based algorithms evaluate the design criteria and properties to generate a probability of success for each design. The success of heuristic algorithms demonstrated in various fields suggests a

---

* Corresponding author.
E-mail addresses: foo@iastate.edu (J.L. Foo), lobet2@ihs.org (T. Lobe), ewiner@iastate.edu (E. Winer).

0010-4825/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compbiomed.2010.11.006

strong applicability in medical image segmentation as well. Similar to when evaluating an engineering design or a protein design, pixels in a dataset can be evaluated based on pre-defined criteria such as intensity values, pixel location, shape of the object of interest, and other descriptive features within the image.

The complexity of medical image segmentation has encouraged many fields of research into improving automatic and adaptive segmentation algorithms. These methods perform by analyzing the pixels in an image and generating a statistical histogram or a probability map. Statistical histograms are generated based on the pixel intensities values and will incorporate statistical properties such as mean intensity, the distribution of intensity values, and/or the homogeneity of the pixel intensities. Similarly, probability maps are generated based on pre-defined criteria set by the user. In most cases, probabilities of a pixel being the object of interest are generated for each pixel based on its intensity and location [9–11].

For example, Vincken et al. [12] and Kupinski and Giger [13] developed methods of using probability as a means to merge regions. Vincken's method divides a three-dimensional image data representation created from slices into a voxel data tree structure, and links adjacent voxels based on a parent–child voxel probability linking scheme. Although this approach produced promising results, the computational resources are very expensive and the processing can be time consuming. Kupinski's method performs a statistical analysis on the pixel intensity histogram using a small-variance Gaussian function on previously defined partitions of the image data, to generate a probability distribution for these partitions. Based on these probability distributions, the partitions are evaluated and the partition with the maximum probability is selected as the region of interest. Another method by Anbeek et al. implemented a probabilistic algorithm for image segmentation using *k*-nearest neighbors (KNN) means to calculate the probability of a voxel being either: white matter, gray matter, cerebro-spinal fluid, ventricles, and white matter lesions in cranial MR images [14,15]. Algorithms can also analyze the dataset prior to segmentation to build a predictive model before employing maximum-likelihood probability estimation for segmentation of brain tissues [16,17]. Melas and Wilson [18] used Markov Random Fields (MRF) as a model estimation tool to generate likelihood values based on texture information within the image, and later segmented using a Bayesian approach. Zhang et al. [19] used hidden MRF models with model fitting performed using the Expectation–Maximization (EM) algorithm, to extract features from brain MR images [20,21]. However, these model estimation methods are highly dependent on the initial model formed and the segmentation results can be wrong if the initial model is not well defined.

There are also methods that will incorporate probability or statistical analysis with other segmentation approaches. Learning algorithms like fuzzy logic can be integrated to represent spatial relations within the image when performing segmentation using deformable models [22] or incorporating neural networks with probability and statistical pixel labeling for segmentation of brain tissues [23]. Toledo et al. [24] introduced the idea of a deformable model called eigen-snakes to segment vessels in angiographs based on a probabilistic framework that performed with encouraging results. The snake deforms to maximize the probability of defining the vessels based on the trained features.

Even though current probabilistic segmentation methods have shown success in segmenting the required object of interest, these current implementations usually require large amounts of computing and processing resources or a significant amount of *a priori* knowledge of the data. These methods perform a first-pass inspection of the entire medical study in order to generate statistical histograms or probability maps. Loading the entire image study in 3D for this processing requires immense computational resources. Thus, very few methods are able to generate and calculate probabilities for segmentation in real-time.

Based on the review of the methods discussed, it has shown that a probability based segmentation algorithm has been a prevalent concept in image segmentation. It is based on the success of these past methods that a probabilistic algorithm is implemented in the proposed segmentation method for generating pixel selection probabilities. The formulation of the probability calculations using a pixel's spatial and intensity relations in this proposed method is a unique approach based on the literature review performed.

All the methods discussed represent this rich and diverse research area. While each of these methods successfully segment the objects of interest in test cases, specific requirements limit their widespread use and adoption as outlined in the previous sections. Using these concepts, a segmentation method that performs in two-dimensional space with the implementation of a probabilistic pixel selection process was created. Since the method functions in two-dimensional space, computational needs are substantially reduced, allowing for efficient performance on a typical personal computer or laptop and also making this method highly robust to work on both small and large datasets. In addition, since the method processes each two-dimensional slice image data independently using the already available spatial and pixel intensity properties, the need for *a priori* knowledge of the data is virtually eliminated, apart from the need of a user input in the first slice. As the results from distinct test cases will show, the proposed method performs efficiently and with very high accuracy.

Prior to the segmentation process, the image data is first pre-processed to reduce acquisition noise and increase the contrast of the object of interest by adjusting the window center and width settings. The method then implements probabilistic concepts for image segmentation by making use of intensity and spatial properties of the image data. A probability of a pixel being selected as part of the region of interest is generated based on that pixel's intensity value and location. The proposed method is highly customizable to suit different needs of a user and requires information from only a single two-dimensional image.

2. Methods

The method begins with a pre-processing step that includes noise removal and performing a window center/width operation to increase the contrast between the object of interest and the background. To initialize the segmentation process, a user selects seed point(s) within the first slice in the study and defines a search region. The results presented in this paper are based on only a single seed point for all test cases. The placement of the seed point to initialize the segmentation process is a critical step since this is the only time where the user would provide any form of information to the method and future steps of the method is dependent on this first seed point. For each remaining slice in the set, an automated algorithm is performed to automatically place the seed and redefine the search region for segmentation. Lastly, probabilities are generated for every pixel in the search region, and pixels with high probabilities are segmented as the object of interest. The process is then repeated for every image slice in the dataset. The flowchart in Fig. 1 shows the steps of the segmentation process.

2.1. Image pre-processing

Before any segmentation is performed, the image data is pre-processed. The first step in this pre-processing is to obtain the image data in terms of pixel intensities via windowing. The window center and width of the Hounsfield Units are set to improve the contrast between the object of interest and the background,