14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

A ROBUST THRESHOLDING METHOD WITH APPLICATIONS TO BRAIN MR IMAGE SEGMENTATION

Ahmet Ekin and Radu Jasinschi

Video Processing Group, Philips Research, Prof. Holstlaan 4, Eindhoven, The Netherlands
{ahmet.ekin, radu.jasinschi}@philips.com

ABSTRACT

This paper brings forth two main novel aspects: 1) a generic thresholding method that is robust to degradation in the image contrast; hence, quality and 2) a new knowledge-based segmentation framework for brain MR images that first utilizes a clustering algorithm, and then the proposed thresholding method. The new thresholding method accurately computes a threshold value even for images with very low visual quality having very close class means. It also consistently outperforms known thresholding methods. The segmentation algorithm, on the other hand, generates almost constant segmentation performance in a wide range of scan parameter values. It utilizes first a clustering algorithm to identify the CSF (cerebrospinal fluid) region and then focuses on white matter (WM) – gray matter (GM) separation by using the novel thresholding method. We show the robustness of the proposed algorithms with a simulated dataset obtained with various parameter values and a real dataset of brain MR dual-echo sequences of patients with possible iron accumulation.

1. INTRODUCTION

Segmentation plays an essential role in many image-processing applications. In medical imaging, its importance will likely grow with the increasing activities in CAD (computer-aided detection) and CADx (computer-aided diagnosis) because segmentation is a bridge between low-level image processing and high-level interpretation. In addition to paving the way for the high-level analysis, segmentation also makes it possible for the medical experts to visualize the data in manners that improve the diagnosis.

Due to the importance of the topic, the segmentation literature is vast. There are many algorithms specializing in a *modality,* e.g., MR, CT, and PET, an *organ,* such as brain and lung, and *the type of the abnormality,* such as multiple sclerosis lesions and lung nodules. Most available methods start with certain assumptions about the image quality and the modality. The image quality is mostly related to the applied protocol; for example, repetition time (TR) and echo time (TE) affect the tissue contrasts in MR T1 and T2 images, respectively. The choice about the protocol and the quality resulting from it depend on many factors, such as cost, suspected pathology, the experience level of the radiologist, and the properties of the imaging device. Another factor that affects the image quality is pathological abnormalities. For example, the accumulation of iron in basal ganglia is common to patients with neurodegenerative diseases, such as Alzheimer's. Due to the magnetic properties of iron, iron accumulation in basal ganglia (a gray matter (GM) tissue) increases the intensity variance of the GM class. This invalidates the assumptions of many segmentation algorithms, such as unimodal distribution for each class, and has direct consequences in the quality of MR T2 contrast.

Some of the segmentation algorithms also assume the availability of certain MR contrasts. This may not be always applicable. Because of the suspected disease, the available scan time, and the associated cost, radiologists may skip some of the MR contrasts at the expense of some others.

The main aim of this paper is to propose a set of methods that is robust to the variations in the image quality and the type of the input. Our main application is the segmentation of brain MR images; however, one of the novel aspects of the paper is a generic thresholding method that is robust to variations in image contrast and is not limited to this application or to medical imaging. The proposed method finds a threshold value by minimizing the total segmentation error for the data at the class boundaries and it outperforms known methods. We also introduce a novel two-step segmentation framework for brain MR images based on this thresholding method.

In the next section, we present the proposed generic thresholding method. In Section 3, we explain the segmentation framework. Section 3.1 details the first stage of the brain image segmentation algorithm that identifies the CSF region by a clustering method and Section 3.2 explains the use of the proposed thresholding method to classify white and gray matter regions. Section 4 provides the experimental results on both synthetic and real data. Finally, Section 5 concludes the paper.

2. A NEW THRESHOLDING METHOD BASED ON VOXELS AT THE CLASS BOUNDARIES

In this section, we propose a new thresholding method that first selects data points that are close to the positions of class transitions, and then computes the threshold value to minimize the number of misclassified labels for those selected points. This algorithm is robust to image quality degradations as well as partial volume effect.

Thresholding aims at finding a boundary between two classes. In [1], the existing approaches for threshold value computation are classified based on the information they exploit: 1) histogram shape information, 2) clustering of measurements, 3) entropy, 4) image attributes, 5) spatial information, and 6) local characteristics. Most of these methods assume little overlap between class probability density functions. One of the most popular thresholding schemes is Otsu's method [2] and it generates better results when the two classes possess comparable number of

members [1]. Except for the last two types, the others do not use spatial information and work solely in a feature space. Furthermore, these methods assume that each data point in the image belongs to one class.

In the domain of medical imaging, partial volume effect (PVE) violates the crisp class membership assumption. PVE also has a smoothing effect to the probability density function. Furthermore, the scanning protocol may also reduce the contrast. We observed that when the contrast reduces significantly, some of the promising approaches in [1] become unpredictable and show large performance degradations. Finally, because of the anatomical variations and the pathological factors, it is usually impossible to speculate on the number of voxels in each tissue class.

Considering the above observations, we propose a new thresholding method. The proposed method first detects the spatial positions of class transitions, defined as edges, and constructs two histograms by using the data points on either side of the edge. Then, cumulative distribution function of each class is computed from the histograms, which are assumed to approximate the probability distribution function. Finally, a threshold value is computed to minimize the number of misclassified labels for the selected representative set. The advantages of our approach are that 1) it is robust to PVE, 2) it does not make any assumptions about the number of class members, 3) it can compute the threshold value even when the classes are significantly merged, and finally, 4) it makes use of spatial information.

We first compute the contrast in the image, or a region, as defined in Equation 1, where $\vec{x}$ and $\vec{y}$ denote *N*-dimensional spatial coordinates (3-D for medical data), and *R* is a region that may be the whole image or a region defined with a mask. According to Equation 2, the computed contrast value determines the edge threshold value, T_{edge}, which is automatically adapted to the image content. Afterwards, class borders are detected by comparing the gradient magnitude, computed as in Equation 3, with the edge threshold (as shown in Equation 4). The parameter Δ in Equation 4 determines the step size and the edge direction. When there is no PVE, the magnitude of Δ can be equal to one. Otherwise, a value greater than one should be used so that PVE has less effect in the compared pixel locations.

$$C = \max\left(I(\vec{x}); \vec{x} \in R\right) - \min\left(I(\vec{y}); \vec{y} \in R\right) \quad (1)$$

$$T_{edge} = \max\left(k * C, 1\right) \quad (2)$$

$$G(\vec{x}) = \left|I(\vec{x} + \Delta) - I(\vec{x} - \Delta)\right| \quad (3)$$

$$E(\vec{x}) = \begin{cases} 1 & if\ G(\vec{x}) \geq T_{edge} \\ 0 & otherwise \end{cases} \quad (4)$$

For each detected border point, two data points that are used in Equation 3 are selected. These points are assigned to two separate histograms, H_L and H_U, whereby the smaller intensity value is assigned to the former and the larger one to the latter. In this way, we select equal number of points from each class.

In the final step, we compute the cumulative distribution functions (C) of the two classes. Assuming normalized histograms approximate the probability density, they can be computed as in Equation 5. Then, as shown in Equation 6, the threshold value, *T,* is computed to maximize the sum of the correct decisions for the representative samples that are close to the boundary.

$$C_L(u) = \sum_{i=0}^{u} H_L(i) \quad (5)$$

$$T = \arg\max_{I} \left(C_L(I) - C_H(I)\right) \quad (6)$$

3. PROPOSED SEGMENTATION FRAMEWORK

We aim to segment brain MR images into three tissue types: white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF). Please note that the abnormalities or lesions are going to be assigned to one of these three classes. Their detection can be done with further analysis by using the segmentation result as well as other cues; but the detection of abnormalities is not the topic of this paper. For the three-class segmentation, we propose an integrated clustering and thresholding approach. In the first step, we aim to extract the CSF region and consider WM and GM as one class, and then in the second step, we focus on separating merged WM-GM class with a finer analysis. The reason for this is that WM and GM tissues may not be easily separable because of non-optimal settings of the scanning parameters and abnormalities, such as iron deposition and lesions. As a result, regular clustering approach may not distinguish one from the other. In the following, we first explain the CSF extraction step that is based on K-harmonic means (KHM) clustering. After that, we describe how to use the proposed thresholding method in Section 2 to separate WM and GM, which are merged in the output of the first step.

3.1 CSF extraction by K-Harmonic Means clustering

Our aim in this step is to classify brain tissue into two classes: CSF and WM-GM. The input is a 3D skull-stripped brain MR image and we do not make any assumptions about the available contrasts, such as T1 and T2. The skull stripping can be done by applying one of the publicly available tools, such as [3]. We define the segmentation problem as an unsupervised clustering problem where the number of classes is two. For clustering, we use K-harmonic means (KHM) algorithm [4] because it is reported to be more robust to initialization [5][6]. Even though KHM is robust to initialization, the accurate identification of the modes is important because it speeds the segmentation up by reducing the convergence time. To this effect, we use the mean shift algorithm to accurately identify the modes of the distribution. The use of both mean-shift based initialization and KHM-based clustering provides more robustness to the overall segmentation algorithm.

For the initialization of the seed points, we follow a similar approach to the work in [7]. Because we know the number of clusters and our application is medical imaging, there are some differences. First, we build the intensity histogram of the brain tissue. The number of contrasts, such as T2 and PD, determines the dimension of the histogram. The proposed algorithm is not dependent on this value. Given the brain tissue mask (the result of

the skull stripping software, e.g., [3]), the steps of the algorithm are as follows:

1. Compute the histogram of the input brain image (this can be done for a representative slice)
2. Compute the Chamfer distance transform [8] of the brain tissue mask image (the peripheral region is assigned to the lower distance values whereas the image center will have the highest distance value) to form mask distance image I_D
3. Select N spatial locations for each distinct distant value in I_D (the reason for making the selections as a function of the distance value is that CSF and GM mainly occupy the peripheral regions and the random selection of feature points is not the most efficient method)
4. Starting with the feature (intensity) vector of the selected location, apply the mean shift algorithm to find the closest mode
5. Pick the most salient and sufficiently apart two modes and assign them as seed/center points for the clustering

After the initialization, clustering is performed as usual by first assigning class membership values to the voxels, and then updating class means by class membership and feature values, and continuing these iterations until convergence. As mentioned before, for the clustering, we use KHM instead of common choices of K-means or fuzzy C-means. KHM uses harmonic mean of the distance from the feature vectors to the mean [4]. It differs from both K-means and fuzzy C-means in that unlike K-means, class membership values are soft but at the same time, voxel weights are variable, which differs from fuzzy C-means [5].

3.2 White matter (WM) – Gray matter (GM) separation

Once CSF is extracted, we focus on the separation of WM and GM. The reasons for doing this are that: 1) the contrast between WM and GM is usually less than the contrasts between each of WM and GM, and CSF in many MR contrasts. In an MR scan with non-optimal parameter settings, WM and GM may end up as one cluster in a straightforward clustering algorithm. 2) Abnormalities may affect the magnetic properties of WM and GM. This makes finer analysis of WM and GM regions necessary. For example, large amounts of iron accumulation change the T2 characteristics of basal ganglia and result in hypo-intensities in those regions.

We use the thresholding method proposed in Section 2 to label WM and GM. After applying the thresholding method, the decision about labeling the two regions as WM and GM uses the knowledge about their expected relative intensity values. For example, the region with the higher intensity value is assigned as WM if T1-weighted images are used. The relative locations of WM and GM tissues can also be used for this purpose, e.g., GM has a peripheral region surrounding the CSF.

4. RESULTS

We evaluate the robustness of the proposed thresholding and segmentation algorithms on simulated and real datasets. The simulated dataset has been obtained from BrainWeb site [9]. The key purpose of using this simulated data is to evaluate the performance of the proposed algorithms under different image quality values. In this context, the primary factor affecting the image quality is the contrast between WM and GM. The simulated dataset consists of a T1-weighted set, which is obtained by varying the TR value (in the range 17ms - 100ms) with a fixed TE of 15 ms and a flip angle of 20 degrees, and a T2-weighted set, which resulted from changing the TE value (in the range 40ms - 180ms) with a fixed TR of 3000ms. The slice thickness for the T1w dataset is 1 mm whereas that for T2w dataset is 3mm. The choice for the latter is that we wanted to have equal slice thickness to our real T2w dataset, which is explained below, so that the performance values could be comparable. The ground truth is available on the BrainWeb site.

The real dataset of brain MR images was obtained from Leiden university medical center (LUMC). These are dual-echo MR images (T2-weighted and proton density (PD)) acquired with Philips 1.5T Intera MR System with TR 3300 ms, and TE of 27ms -120ms. The slice thickness is 3 mm. This dataset consists of patients with possible iron accumulation in basal ganglia organs. As a result, T2-weighted images are not reliable in the WM-GM separation. We use T2w and PD in the first stage of the segmentation, and then use only PD in the second stage. Because of the PVE resulting from 3mm slice thickness, low contrast due to scan parameters, and the pathological factors, this set is very challenging. We will present segmentation results as images due to the lack of the ground truth for this set. However, we also obtained simulated data from BrainWeb with the same settings to be able to report a performance value for comparison.

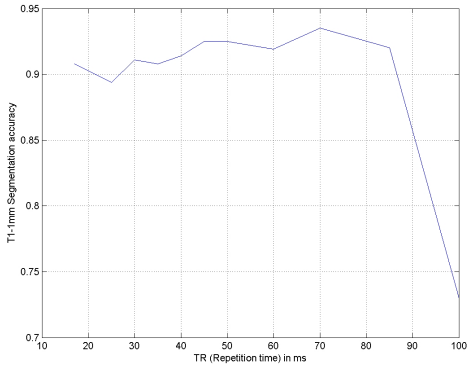

Figure 1: The segmentation performance in T1 images as a function of repetition time

Figure 1 is a plot of the segmentation accuracy, defined as the ratio of the total number of accurately classified voxels to the total number of voxels, in the T1 set as a function of TR (repetition time). The highest segmentation accuracy is 92.5% and it stays within 3% of this value in the interval between 17 ms and 85 ms of TR. The contrast between the WM and GM tissues decreases significantly for TR values larger than 70 ms. For example, Figure 2 shows an image obtained with TR = 85 ms. The borders between WM and GM tissues are below JND (just noticeable difference) levels. In spite of that, the proposed segmentation algorithm achieves 92% accuracy. This is significantly better than a method that is completely based on a clustering approach. As shown in Figure 3, we observe same type of improvement for the T2w set. Because of the larger slice thickness (3mm), the maximum accuracy is 79% and the performance value stays within the 3% of its maximum in a large interval of TE values.

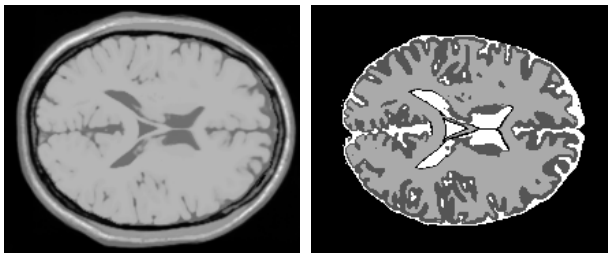

Figure 2: Segmentation result of a T1w image with TR = 85ms; the boundaries between WM and GM tissues are not easily noticeable to the eye, but the proposed algorithm can accurately segment the image (92% accuracy)

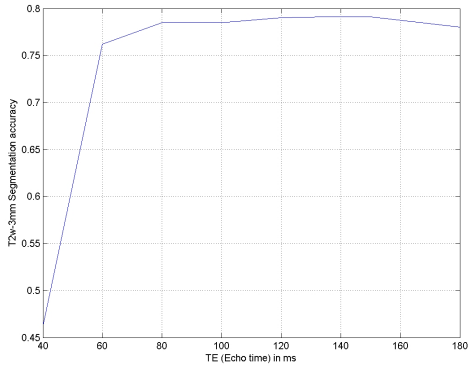

Figure 3: The segmentation performance in T2w images as a function of echo time

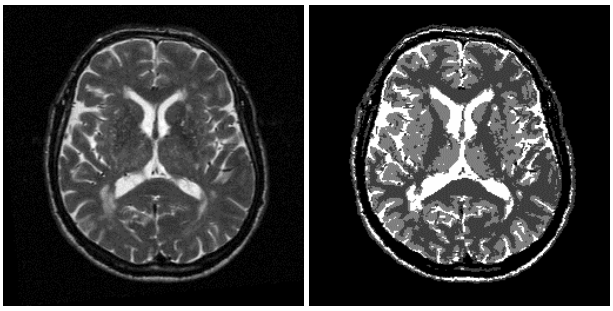

Figure 4: T2w image (left) with the resulting segmentation map; the effects of pathological factors and the low contrast between WM and GM were effectively handled by the proposed method (Courtesy of LUMC)

For the LUMC set, we used both T2w and PD images in the CSF extraction whereas WM and GM are separated from only PD because iron concentration degrades the T2w contrast more adversely. Figure 4 shows a segmentation result over a real dataset. Visual evaluation of the segmentation results for more than 30 scans was positive. The algorithm effectively handled both low contrast between WM and GM and the iron-related hypo-intensities in basal ganglia. The BrainWeb simulation having the same parameters was segmented at 79% accuracy. In the evaluation, we discretized PVE voxels as belonging to the class having the highest membership value. This reduces the reported performance value.

5. CONCLUSIONS

We introduced a generic thresholding method and a fast segmentation algorithm for brain MR images that are robust to visual quality degradations independently from the input contrast. The proposed algorithms have been shown to be consistent across a large range of parameter space as well as to be effective in the segmentation of pathological data. In the future, we plan to investigate the efficiency of the algorithm for other modalities, such as PET. Another interesting future direction could be the use of a brain atlas. This demands extensive care because of pathological abnormalities and anatomical variations. For example, most of the healthy patients in the LUMC set are healthy in the sense that they do not have large iron concentration in basal ganglia; however, most of them have other forms of white matter lesions.

6. REFERENCES

[1] B. Sankur and M. Sezgin, "A survey of image thresholding techniques and quantitative performance evaluation," *Journal of Electronic Imaging,* 13(1), 146-165, Jan. 2004.

[2] N. Otsu, "A threshold selection method from gray level histograms," *IEEE Trans. Systems, Man and Cybernetics,* vol.9, March 1979, pp. 62-66.

[3] S. Smith, BET2: Brain extraction tool, University of Oxford http://www.fmrib.ox.ac.uk/fsl/bet2/

[4] B. Zhang, "Generalized K-harmonic means - boosting in unsupervised learning," Technical Report HPL-2000-137, Hewlett-Packard Labs, 2000.

[5] G. Hammerly and C. Elkan, "Alternatives to the k-means algorithm that find better clusterings," In *Proc. ACM on information and knowledge management,* pp. 600-607, Nov. 2002.

[6] A. Ekin, S. Pankanti, and A. Hampapur, "Initialization-independent spectral clustering with applications to automatic video analysis," in *Proc. IEEE ICASSP,* May 2004, Montreal, Canada.

[7] D. Comaniciu and P. Meer, "Robust Analysis of Feature Spaces: Color Image Segmentation," in *Proc. IEEE CVPR*, San Juan, Puerto Rico, 750-755, 1997.

[8] G. Borgefors, "Hierarchical Chamfer Matching: A Parametric Edge Matching Algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.,* 10(6): 849-865, 1988.

[9] [Online] BrainWeb: simulated brain database, http://www.bic.mni.mcgill.ca/brainweb/