NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

A Graph-Theoretic Approach for Segmentation of PET Images

Ulaş Bağci[1,2], **Jianhua Yao**[2], **Jesus Caban**[3], **Evrim Turkbey**[2], **Omer Aras**[2,4], and **Daniel J. Mollura**[1,2]

[1]Center for Infectious Disease Imaging, MD, USA

[2]Department of Radiology and Imaging Sciences, MD, USA

[3]National Library of Medicine of National Institutes of Health (NIH), MD, USA

[4]Department of Radiology/Nuclear Medicine, University of Maryland Medical System, MD, USA

Abstract

Segmentation of positron emission tomography (PET) images is an important objective because accurate measurement of signal from radio-tracer activity in a region of interest is critical for disease treatment and diagnosis. In this study, we present the use of a graph based method for providing robust, accurate, and reliable segmentation of functional volumes on PET images from standardized uptake values (SUVs). We validated the success of the segmentation method on different PET phantoms including ground truth CT simulation, and compared it to two well-known threshold based segmentation methods. Furthermore, we assessed intra-and inter-observer variation in delineation accuracy as well as reproducibility of delineations using real clinical data. Experimental results indicate that the presented segmentation method is superior to the commonly used threshold based methods in terms of accuracy, robustness, repeatability, and computational efficiency.

I. INTRODUCTION

^{18}F-FDG PET functional imaging is a widely used modality for diagnosis, stating, and assessing response to treatment. The visual and quantitative measurements of radio-tracer activity in a given ROI is a critical step for assessing the presence and severity of disease. Overlap or close juxtaposition of abnormal signal with surrounding normal structures and background radio-tracer activity can limit the accuracy of these measurements, therefore necessitating the development of improved segmentation methods. Accurate activity concentration recovery, shape, and volume determination are crucial for this diagnostic process. PET segmentation can be challenging in comparison to CT because of the lower resolution which can obscure the margins of organs and disease foci. Moreover, image processing and smoothing filters commonly utilized in PET images to decrease noise can further decrease resolution.

Most of the studies regarding delineation of PET images are based on manual segmentation, fixed threshold, adaptive threshold, iterative threshold based methods, or region-based methods such as fuzzy c-means (FCM), region growing, or watershed segmentation [1], [2], [3], [4], [5]. Although these advanced image segmentation algorithms for PET images have been proposed and shown to be useful upto certain point in clinics, the physical accuracy, robustness, and reproducibility of delineations by those methods have not been fully studied and discovered. In this study, we present an interactive (i.e., semi-automated) image

ulas.bagci@nih.gov.

segmentation approach for PET images based on random walks on graphs [6], [7]. The presented algorithm segments PET images from standardized update values efficiently in pseudo-3D, and it is robust not only to noise, but also to patient and scanner dependent textural variability, and has consistent reproducibility. Although the presented method can be easily extended into a fully automated algorithm, interactive use of the method offers users the flexibility of choosing the slice and region of interest so that only the selected slices and region of interests in those slices are segmented. In the following section, we describe the basic theory behind the random walk image segmentation and its use in segmenting PET images using SUVs. Then, we give the experimental results on segmentation of PET images.

II. Graph-based Methods for Image Segmentation

The "Graph-based" approaches [9], as alternatives to the boundary based methods, offer manual recognition, in which foreground and background or objects are specified through user-interactions. User-placed seed-points offer a good recognition accuracy especially in the 2D case. Graph-cut (GC) has been shown to be a very useful tool to locate object boundaries in images optimally. It provides a convenient way to encode simple local segmentation cues, and a set of powerful computational mechanisms to extract global segmentation from these simple local (pairwise) pixel similarity [9]. Using just a few simple grouping cues, called *seed points* and serve as segmentation hard constraints, one can produce globally optimal segmentations with respect to pre-defined optimization criterion. A major converging point behind this development is the use of graph based technique. In other words, GC represents space elements (*spels* for short) of an image as a graph with its nodes as spels and edges defining spel adjacency with cost values assigned to edges, and to partition the nodes in the graph into two disjoint subsets representing the object and background. This is done by trying to find the minimum cost/energy among all possible cut scenarios in the graph where GC optimizes discrete energies combining boundary regularization with regularization of regional properties of segments [9]. A common problem in GC segmentation is the "small cut" behaviours, happening in noisy images, or images including weak edges. This behaviour may produce unexpected segmentation results along the weak boundaries. Since PET images are poor in terms of resolution, and since weak boundaries often exist in PET images, we propose to use random walk (RW) segmentation method for segmentation of PET images to provide globally optimum delineations. By using RW segmentation method, we also often avoid "small cut" behaviours because weak object boundaries can be found by RW segmentation method as long as they are part of a consistent boundary. Due to the properties of providing globally optimum delineations and being less susceptible to the "small cut" behaviours, we propose to use RW segmentation method in delineating PET images.

A. Random Walks for Image Segmentation

Among graph-based image segmentation methods, RW segmentation has been shown very useful in interactive image delineation. It appeared in computer vision applications in [8], and then extended for image segmentation in [6], [7]. In this section, we describe the basic theory of RW image segmentation and its use in segmentation of PET images from SUVs.

Suppose G = (V,E) is a weighted undirected graph with vertices $v \in$ V and edges $e \in$ E⊆V×V. Let an edge spanning two vertices, v_i and v_j, be denoted e_{ij}, and weight of an edge is defined as w_{ij}. As common to graph-based approaches, edge weights are defined as a function, which maps a change in image intensity to edge weights. In particular, we use un-normalized Gaussian weighting function to define edge weights as: $w_{ij} = exp(-(g_i - g_j)^2)$, where g_i represents the SUV of pixel i. Assuming that the image corresponds to a lattice where SUV of each pixel is mapped to edge weights in the lattice such that some of the

nodes of the lattice are known (i.e., fixed, labelled), V_M, by user input (i.e., seeds, marks), and some are not known, V_U, such that $V_M \cap V_U = V$ and $V_M \cup V_U = \emptyset$. The segmentation problem in this case is basically to find the labels of unseeded (not fixed) nodes. A combinatorial formulation of this situation is nothing but the Dirichlet integral, as stated previously in [6],

$$D[x] = \frac{1}{2}(Ax)^T C (Ax) = \frac{1}{2}x^T Lx = \frac{1}{2}\sum_{e_{ij} \in E} w_{ij}\left(x_i - x_j\right)^2, \quad (1)$$

where C is the diagonal matrix with the weights of each edge along the diagonal, and A and $L(= A^T CA)$ are incidence and Laplacian matrices indicating combinatorial gradients in some sense, and defined as follows:

$$A_{e_{ij}}v_k = \begin{cases} 1 & \text{if } i=\text{k} \\ -1 & \text{if } j=\text{k} \\ 0 & \text{otherwise} \end{cases} \quad . \quad (2)$$

Solution of the combinatorial Dirichlet problem may be determined by finding the critical points of the system. Differentiating $D[x]$ with respect to x, and solving the system of linear equations with $|V_U|$ unknowns yield a set of labels for unseeded nodes if every connected component of the graph contains a seed (See [6] for non-singularity criteria). In other words, RW efficiently and quickly determines the highest probabilities for assigning labels to the pixels by measuring the "betweenness" through starting pixel of the RW (labeled pixel) to the un-labeled pixel, reached first by the random walker.

III. Evaluations and Results

We performed a retrospective study involving 20 patients with PET-CT scans. PET scans' resolution is limited to $4\text{mm} \times 4\text{mm} \times 4\text{mm}$. The patients are having diffusive lung parenchymal disease abnormality patterns including ground glass opacities, consolidations, nodules, tree-in-bud patterns, lung tumours and non-specific lung lesions. PET scans are having more than 200 slices per patient with in-plane resolution varying from 144×144 to 150×150 pixels with 4mm pixel size. The average of the time for delineating one slice by the RW segmentation method was around 0.1 seconds. We qualitatively and quantitatively demonstrated the success of the presented delineation method both on phantom and real clinical data. We also compared the RW segmentation to the threshold based methods: fuzzy locally adaptive bayesian (FLAB) and FCM clustering [1]. Comparative results are explained in the following subsections.

Figure 1 shows the performance of the presented RW segmentation method qualitatively. Each row of Figure 1 shows steps of the RW segmentation on PET images for different anatomical location such that: (a) PET images for different anatomical locations, (b) seed locations (blue for object, red for background), (c) segmented object regions.

A. Reproducibility

Repeatability, or reproducibility in other words, is vital for many computational platform including segmentation of images. Particularly in seed based segmentation, it is of interest to know the effects of locating the user defined seeds on images for segmentation. Figure 2 shows an example segmentation of PET images examining the sensitivity of the RW segmentation to the user defined seeds. Not only the number of background and foreground seeds were changed in each case, but also their locations were changed while the consistency between separation of object and foreground was still kept. Apart from this qualitative evaluations, we also computed quantitative scores of the correlation between

segmentation results via dice similarity scores. For each segmented volume $\mathscr{V}_s$ and a reference segmentation $\mathscr{V}_r$, the dice similarity score is calculated as

$$\mathscr{D}\,(\mathscr{V}_r,\mathscr{V}_s) = \frac{2|\mathscr{V}_r \cap \mathscr{V}_s|}{|\mathscr{V}_r|+|\mathscr{V}_s|}100\%. \quad (3)$$

From each of 20 patients, we selected slices with high uptake regions (i.e., regions are either small or large depending on the abnormality: nodule, non-specific lesion, consolidation, TIB, and etc.), and for each slice we repeated the segmentation experiments 10 times by putting the seeds randomly over the image regions while keeping the seeds belonging to object and foreground. Resultant dice similarity scores for each segmentation experiment was compared after taking the mean over all pairs for the same patients, and eventually for all patients. The mean, standard deviation (std), max, and min scores are reported in Table I. Note that the presented segmentation method has a variation in segmented accuracy <1%. (See [11] for a full sensitivity analysis of segmentation to the number of seeds).

B. Validation by Ground Truth

We validated the presented segmentation algorithm via IEC image quality phantom [10], containing six different spherical lesions of 10, 13, 17, 22, 28 and 37 mm in diameter (See Figure 3). Two different signal to background ratios (i.e., 4:1 and 8:1), and two different voxel sizes (i.e., $2\times2\times2$ and $4\times4\times4$ mm^3) were considered [1] as reconstruction parameters. Figure 3 (a–d) show phantoms with different reconstruction parameters, and Figure 3(e) shows ground truth, simulated from CT. Segmented objects from the phantoms are shown in the last row of Figure 3. Resultant segmentation variations due to the use of different reconstruction parameters (i.e., signal to background ratios and voxel size) are reported in Figure 4. As expected, low resolution and signal to background ratios degrade the delineation performance. Note also that the variations were computed with respect to the ground truth simulated CT scan.

C. Intra-and Inter-observer Variations

Manual segmentations were performed by three expert radiologists, and the experiments were repeated three times in order to assess intra- and inter observer variations over segmentation accuracies. The inter-observer and intra-observer variations are reported as $22.27\pm6.49\%$ and $10.14\pm4.23\%$, respectively. Intra-observer variations (i.e., delineation by the same expert in different time points) on a sample scan is illustrated in Figure 5. Even though the window level is fixed for each segmentation experiment for the same anatomical location of the same patient, an average of 10% variation was inevitable on average.

D. Comparison to Threshold-based methods

Figure 6 shows the delineations of the same anatomical regions with FCM, FLAB, and RW segmentation algorithms. Note that in FCM and FLAB, there is a need to define region of interest to exclude false positives as a further refinement of the segmentation. On average, dice similarity index between RW and FCM and between RW and FLAB are reported as: $\mathscr{D}\,(\text{RW}, \text{FCM}) = 87.31 \pm 5.16\%$ and $\mathscr{D}\,(\text{RW}, \text{FLAB}) = 92.1688 \pm 4.16\%$, respectively. In addition, segmentation results by the methods FLAB and FCM on the phantom sets show that (1) FLAB performs better than FCM as agreed with [1], (2) reproducibility of the segmentation algorithm is not comparable to the RW segmentation algorithm: FLAB has reproducibility variation of <4% whereas RW has variation of <1%, (3) accuracy of the delineations in RW segmentation are superior to FLAB and FCM: in the most favourable phantom in terms of image quality, FCM and FLAB have average delineation errors of 5–15% and 15–20%, respectively, for objects larger than >2cm, whereas RW algorithm has an

average delineation error of 3.3–16.4% without any restriction in object size, and less than 10% if only objects >2cm are considered. As reported in [1], FCM and FLAB have problems in segmentation of small regions with high uptake, and often failed to segment lesions <2cm, and produce false positives. However, on the other hand, our presented framework successfully segment all regions with great accuracy, and manual removal of any false positive regions are avoided by the user defined object-foreground separation prior to beginning of delineation.

IV. Conclusion

In this study, we presented a fast, robust, and accurate graph based segmentation method for PET images using SUVs. We qualitatively and quantitatively evaluated the RW image segmentation method both in phantom and real clinical data. We also compared the presented RW segmentation algorithm with two well-known and commonly used PET segmentation methods: FLAB and FCM. We conclude from the experimental results that interactive RW segmentation method is a useful tool for segmentation of PET images with high reproducibility.

Acknowledgments

REFERENCES

[1]. Hatt M, et al. A Fuzzy Locally Adaptive Bayesian Segmentation Approach for Volume Determination in PET. IEEE TMI. 2009; 28(6):881–893.

[2]. Montgomery, et al. Fully automated segmentation of oncological PET volumes using a combined multiscale and statistical model. Med.Phys. 2007; 34(2):722–736. [PubMed: 17388190]

[3]. Erdi YE, et al. Segmentation of lung lesion volume by adaptive positron emission tomography image thresholding. Cancer. 1997; 80(12 Suppl):2505–2509. [PubMed: 9406703]

[4]. Day E, et al. A region growing method for tumor volume segmentation on PET images for rectal and anal cancer patients. Med Phys. 2009; 36(10):4349–4358. [PubMed: 19928065]

[5]. Jentzen W, et al. Segmentation of PET Volumes by Iterative Image Thresholding. J Nucl Med. 2007; 48:108–114. [PubMed: 17204706]

[6]. Grady L. Random Walks for Image Segmentation. IEEE TPAMI. 2006; 28(11):1768–1783.

[7]. Andrews S, et al. Fast Random Walker with Priors Using Pre-computation for Interactive Medical Image Segmentation. MICCAI. 2010; 3:9–16. [PubMed: 20879377]

[8]. Wechsler H. A random walk procedure for texture discrimination. IEEE TPAMI. 1979; 3:272–280.

[9]. Boykov Y, et al. Fast Approximate Energy Minimization via Graph Cuts. IEEE TPAMI. 2001; 23(11):1222–1239.

[10]. Jordan K. IEC emission phantom Appendix Performance evaluation of positron emission tomographs. Medical and Public Health Research Programme of the European Community. 1990

[11]. Sinop AK, Grady L. A Seeded Image Segmentation Framework Unifying Graph Cuts and Random Walker Which Yields A New Algorithm. Proc. of ICCV. 2007:1–8.
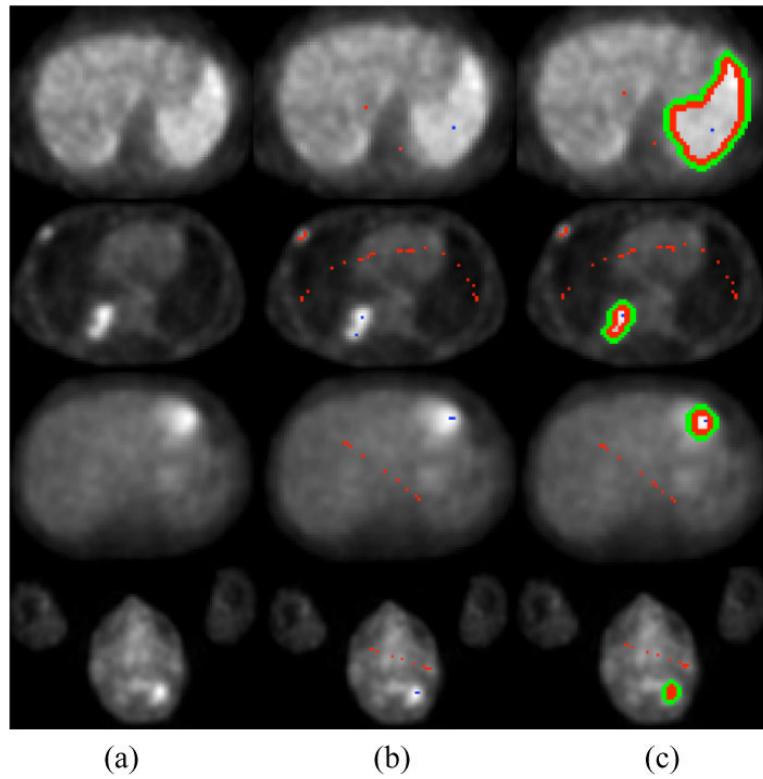
Fig. 1.
(a) different anatomical positions, (b) seeded object (blue) and background (red), (c) delineation results by RW segmentation method.
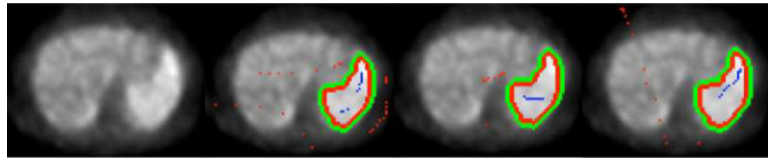
Fig. 2.
Study of RW segmentation reproducibilty by locating seeds in different regions of object and background.
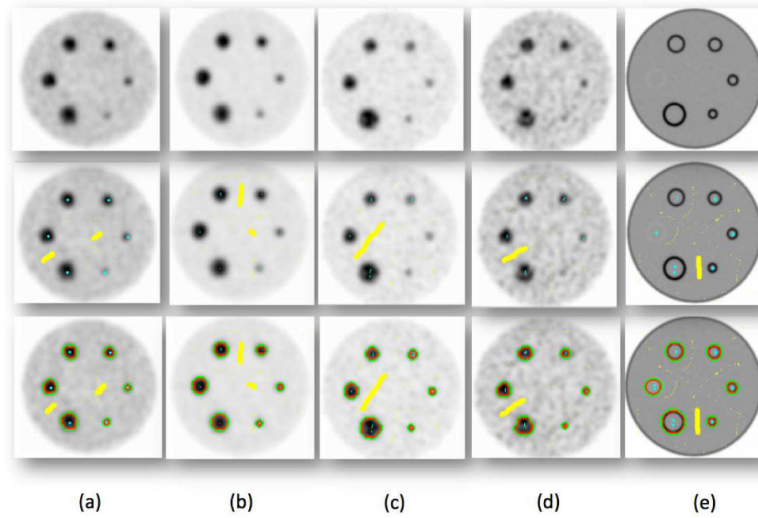
Fig. 3.
First row shows phantom images with following properties: (a) ratio 4:1, 64mm^3, (b) ratio 8:1, 64mm^3, (c)ratio 8:1, 8mm^3, (d) ratio 4:1, 8mm^3, (e) CT acquistion-ground truth. Second row shows user labeling, and the third row shows RW segmentation results.
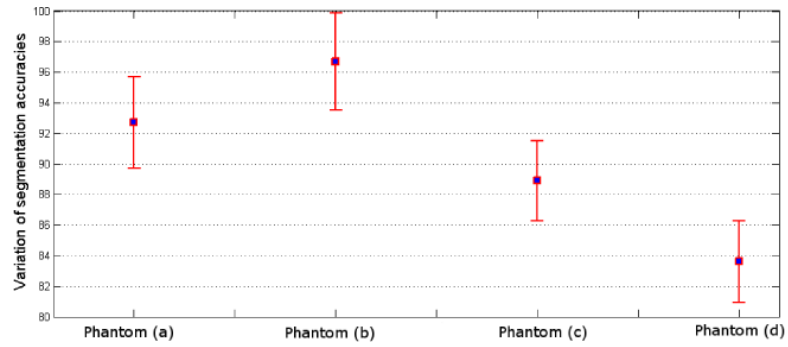
Fig. 4.
Segmentation accuracy of phantoms with respect to the ground truth (Figure 3.e) is shown. Dice scores are listed as 92.7±2.99%, 96.7±3.18%, 88.9±2.62%, and 83.6±2.68%, for phantoms given in Figure 3 (a), (b), (c), and (d), respectively.
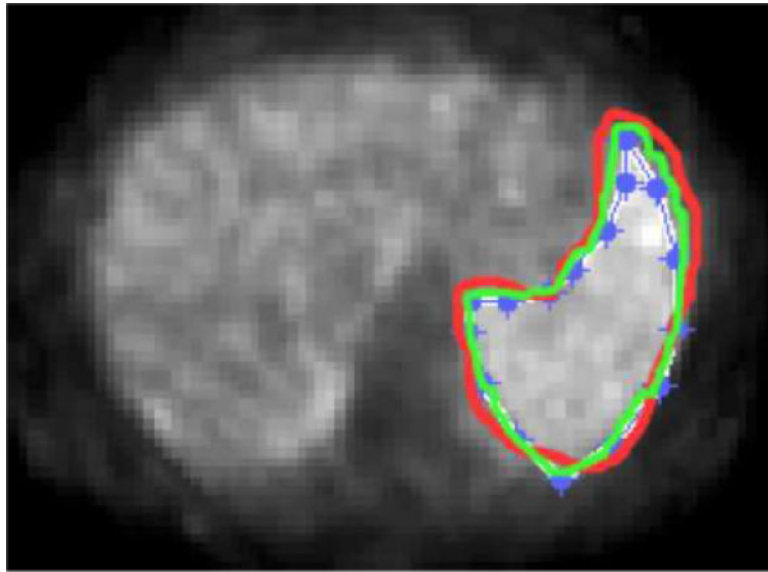
Fig. 5.
Intra-observer variation study: delineation of the high uptake region by the same expert in three different time points is reported.
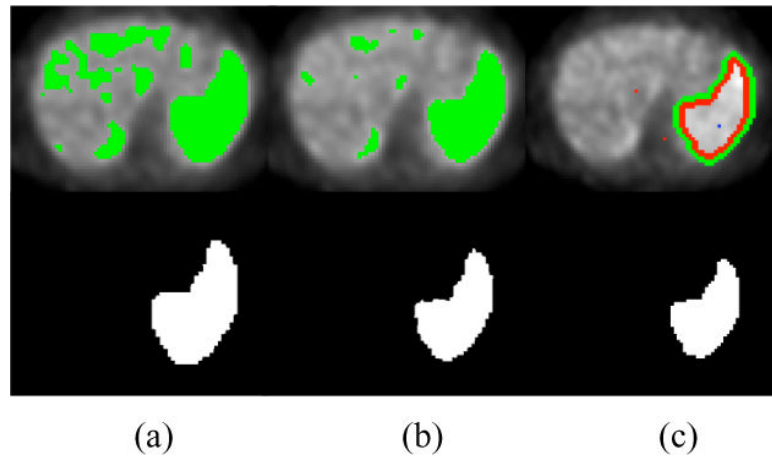
Fig. 6.
Delineation by FCM (a), FLAB (b), and RW (c).

TABLE I

Study of RW segmentation reproducibility via dice similarity scores.

	Mean	Std	Max	Min
Dice Scores	99.0077%	0.6865%	99.6346%	97.2189%