A Statistical Method for Lung Tumor Segmentation Uncertainty in PET Images Based on User Inference

Chaojie Zheng, Xiuying Wang, Dagan Feng

***Abstract*— PET has been widely accepted as an effective imaging modality for lung tumor diagnosis and treatment. However, standard criteria for delineating tumor boundary from PET are yet to develop largely due to relatively low quality of PET images, uncertain tumor boundary definition, and variety of tumor characteristics. In this paper, we propose a statistical solution to segmentation uncertainty on the basis of user inference. We firstly define the uncertainty segmentation band on the basis of segmentation probability map constructed from Random Walks (RW) algorithm; and then based on the extracted features of the user inference, we use Principle Component Analysis (PCA) to formulate the statistical model for labeling the uncertainty band. We validated our method on 10 lung PET-CT phantom studies from the public RIDER collections [1] and 16 clinical PET studies where tumors were manually delineated by two experienced radiologists. The methods were validated using Dice similarity coefficient (DSC) to measure the spatial volume overlap. Our method achieved an average DSC of 0.878±0.078 on phantom studies and 0.835±0.039 on clinical studies.**

I. Introduction

Positron emission tomography (PET) provides the molecular information of cellular activities and has been widely used for primary cancer detection, diagnosis, staging and follow-up therapy in oncology applications [2]. For instance, in order to assess the treatment effectiveness before, during and after treatment, physicians and researchers use PET images to measure and analyze the functional morphology such as the area, volume and the activities of the targeted tissues [3]. The segmentation and separation of tumor from neighboring tissues also helps to minimize the harm and effects to healthy tissues during radiation therapy. Computer-aided segmentation methods provide an efficient way to assist researchers in practice.

The lung tumor segmentation methods from PET images typically utilize the standard uptake value (SUV), which is a semi-quantitative expression of tumoral ^{18}F-Fluorodeoxyglucose (FDG) uptake. The tumor region has higher SUV values and appear as "hot spots" on PET images because lung tumor has increased FDG uptake than surrounding normal tissues [4]. Thresholding-based tumor delineation methods include the fixed threshold of 40% to 50% of maximum SUV (SUV_{max}) [5], the adaptive threshold which is estimated by the source-to-background ratio [6] and iterative thresholding [7]. And other segmentation methods such as fuzzy c-means and gradient based methods for instance watershed [8] were also applied in the tumor delineations. However, these methods may result in over-segmentation when the tumor has heterogeneous FDG uptake or fail to separate the tumor from adjacent similar FDG uptake regions [9]. A tumor-customized downhill region growing method [9] was proposed by taking into account the gradient-based stopping criterion and SUV distribution to better assist the tumor separation [10].

However, the contrast between different nearby objects is relatively low in PET images and there are also "high smoothing" issues which are caused by the cellular activation in nearby tissues [11]. And these will all result in the unclear tumor boundary definition and the "uncertainty" in boundary delineation.

Recently, the random walks (RW) algorithm, which firstly used for computer vision application [12] and later used for image segmentation [13], has been proved of the capability of robustness against noise and weak boundaries [13-15]. However, as pointed out in some literatures [16] [17] [18], because the RW algorithms focused on the local affinity changes, it sensitive to the locations and the number of the input seeds. The segmentation result by RW is a crisp labeling based on the probability map which indicates the probability of unlabeled pixels being reached first from labeled seeds by the random walker.

Due to the low resolution and high smoothing in PET images, it is common that neighboring pixels hold slightly different probability values, while labeled into different categories. Without further analysis on these pixels, the correctness of their labeling is uncertain. Such pixels with uncertain labeling form an uncertainty band along the boundary of target object, e.g. lung tumor. Therefore, the introduction of user inference is important and could guide the analysis on uncertainty band to solve the segmentation uncertainty problem. In our paper, we address the segmentation uncertainty problem by a statistical method analyzing user inference with Principal component analysis (PCA).

II. METHODOLOGY

Our algorithm consists of two stages: uncertain band definition and statistical model for uncertainty band labeling. The workflow of our algorithm is illustrated in Figure 1.

A. Uncertainty Band Definition

In the graph-based segmentation, an image is represented as a weighted undirected graph $G=(V,\mathrm{E})$ with vertices $v_i \in V$ corresponding to an image pixel x_i and an edge

*Research supported by ARC Grants.

C. Zheng, X. Wang, and D. Feng are Biomedical and Multimedia Information Technology research group, School of Information Technologies, University of Sydney, Sydney, Australia (e-mail: xiu.wang@sydney.edu.au). D. Feng is also with Med-X Research Institute, Shanghai Jiao Tong University, China (e-mail: Dagan.feng@sydney.edu.au).

$e_{ij} \in E \subseteq V \times V$ connecting two neighboring nodes. The weight $w_{ij} \in W$ of an edge e_{ij} is to reflect the intensity changes in the input image and defined by un-normalized Gaussian weighting function $w_{ij} = \exp\left(-\beta\left(g_i - g_j\right)^2\right)$, where g_i represents the intensity value of x_i and β is the weighting parameter.

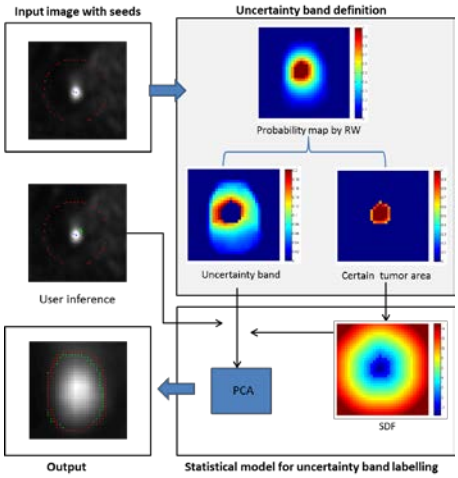

Figure 1. The workflow of our proposed algorithm.

Given initially user specified labels $l_k \in \{l_1, l_2\}$ where l_1 indicates the foreground and l_2 for background and a set of labeled nodes $V_M \subseteq V$, RW calculates the probabilities $Pr = [Pr_{l_1}, Pr_{l_2}]$ of reaching labels l_1 and l_2 from unlabeled nodes $V_U = V \setminus V_M$. And the segmentation is achieved by assigning each unlabeled node a label with larger probability.

The probability Pr is obtained by using function (1)

$$L_U Pr = -B^T Y \tag{0}$$

where Y is the binary matrix of 1 if $v_i \in V_M$ and 0 otherwise; L_U and B^T are the submatrices of the Laplacian matrix $L = D - W$ where $D = diag(\sum_j w_{ij})$ and

$$L = \begin{bmatrix} L_M & B \\ B^T & L_U \end{bmatrix} \tag{0}$$

To solve the segmentation uncertainty problem, we should first define the uncertainty band which is spread along the boundary of target object. In our algorithm, we extract the uncertainty band on the basis of gradient magnitude of the probability map generated from RW as below:

$$I_{uncertain}\left(x_i\right) = \left|\nabla \Pr\left(x_i\right)\right| \tag{1}$$

where Pr is the probability map from RW. In our experiments, we define pixels with probability over 90% as certain tumor areas, and the rest pixels with gradient magnitude of the probability map over 1% as uncertainty band.

Since there is no sufficient information determining the labels of the pixels in uncertain band, additional user inference is required. A small section of the expected segmentation boundary is delineated by the user, and this information is utilized by our algorithm to estimate the rest of the boundary.

Our algorithm extracts the pixel value and distance to certain tumor areas as a feature vector for each training pixel. Signed distance function (SDF) is used to represent the distance between training pixels to certain tumor areas. The boundaries of certain tumor areas, denoted by $C \subset \Omega$, are represented as the zero level set of Lipschitz function $\varphi : \Omega \to R$, such that

$$\begin{cases} C = \partial\omega = \left\{(x, y) \in \Omega : \varphi(x, y) = 0\right\} \\ inside(C) = \omega = \left\{(x, y) \in \Omega : \varphi(x, y) > 0\right\} \\ outside(C) = \Omega \setminus \overline{\omega} = \left\{(x, y) \in \Omega : \varphi(x, y) < 0\right\} \end{cases} \tag{2}$$

$$SDF\left(x\right) = \begin{cases} d\left(x, \partial\omega\right) & if\ x \in \Omega \setminus \overline{\omega} \\ 0 & if\ x \in \partial\omega \\ -d\left(x, \partial\omega\right) & if\ x \in \omega \end{cases} \tag{3}$$

where $d\left(x, \partial\omega\right) = \inf_{y \in \partial\omega} d\left(x, y\right)$. Thus, for each training pixel, we could extract the distance feature from SDF values, and comprise the feature vector with intensity value.

B. *Statistical Model for Uncertainty Band Labeling*

The statistical model is constructed with the set of feature vectors representing the training pixels. For each training pixel, we form a column feature vector for those intensity and distance values. Thus, from *n* training pixels, each training pixel is represented by one dimensional column vector, denoted by $t_i = \begin{bmatrix} g_i \\ SDF\left(x_i\right) \end{bmatrix}$, $\mathrm{x_i}$ where $i \in 1 \ldots n$. Principal component analysis (PCA) is then used during the model construction to capture the variability of the features. Through PCA, we get the mean feature vector $\overline{t}$ representing the mean of training pixels. In addition, the covariance matrix S is also calculated, using

$$S = \frac{1}{n}\sum_{t=1}^{n} dt_i dt_i^T \tag{4}$$

where $dt_i = t - \overline{t}$

Through calculating the eigenvectors and eigenvalues of S, shapes can be derived from the model using

$$t = \overline{t} + Pb \tag{5}$$

where x is the resulted feature vector, $\overline{t}$ is the mean feature vector, $P = \left(P_1, P_2 \ldots P_k\right)$ is the matrix representing the first k variation mode, i.e. first k eigenvectors, and $b = \left(b_1, b_2 \ldots b_k\right)^T$ is a column vector of weights. In order to keep the feature consistency, b is limited to $-3\sqrt{\lambda} < b < 3\sqrt{\lambda}$, where λ is the eigenvalue of the corresponding eigenvector. The derived t represents the features satisfied the statistical model.

The statistical model could therefore be used to classify pixels in uncertain band. The statistical model would check the satisfaction of each uncertain pixels in the uncertain band by calculating:

$$b = P^T \left(t - \bar{t} \right) \tag{6}$$

If $-3\sqrt{\lambda} < b < 3\sqrt{\lambda}$, the pixel should be labeled inside segmentation.

III. Experiments and Discussion

A. Validation Datasets

We validated our algorithm on both phantom and clinical studies. The ten phantom studies from the public RIDER collections [1] were acquired from the Cancer Imaging Archive. The target/background ratio was 4:1. The diameters of the 6 spheres were 10 mm, 13 mm, 17 mm, 22 mm, 28 mm and 37 mm. The PET data were reconstructed using a matrix of 128 × 128 with voxel size 2.73 × 2.73 × 3.27mm.

Sixteen clinical PET studies from patients with NSCLC for further experimental validation and comparison. The scans were carried out on a GEMINI TF Big Bore PET/CT scanner (Philips Medical Systems). PET data were reconstructed into 144 × 144 matrices with pixel size of 4 mm × 4 mm. The slice thickness in PET-CT was 4mm. Each study was delineated by two experienced radiologist and both of the manual segmentations served as "ground truth" for validating the accuracy of segmentation methods.

B. Comparison Methods

We evaluated RW method and our algorithm on phantom studies and applied four segmentation methods, Fuzzy c-means clustering (FCM) [19], tumor-customized downhill (TCD) method [20], RW, and our algorithm to clinical datasets. We optimized the parameters used in each method to reduce bias caused by the parameter settings. In our experiments, pixels with SUV value over 90% maximum SUV are set as foreground seeds for RW and our method. To reduce the bias caused by the introduction of user inference, we also set user inference stroke as additional foreground seeds for RW.

C. Similarity Measure

For quantitative analysis, we used the Dice Similarity Coefficient (DSC) as the similarity measure, which is defined as:

$$DSC\left(V_g, V_s\right) = \frac{2\left|V_g \cap V_s\right|}{\left|V_g\right| + \left|V_s\right|} \times 100\% \tag{7}$$

where V_s is the volume of the segmentation result of each segmentation methods and V_g is the volume of the ground truth.

D. Quantitative Comparison of Segmentation Methods

Figure 2 shows the DSC results of 6 spheres from ten PET phantom studies using RW, RW with additional seeds, and our algorithm. The spheres are listed in ascending order on their diameters. It shows that our algorithm outperformed the RW methods. Figure 3 shows a sample segmentation results from these methods on a phantom study. With user inference as additional seeds, the improvement on the accuracy of RW could be expected. However, the improvement mainly appears in the areas between foreground seeds and additional seeds from user inference. Information from user inference is not fully utilized to affect the decision made on labeling pixels which are away from the foreground seeds. Besides, leakage appears due to the introduction of addition seeds.

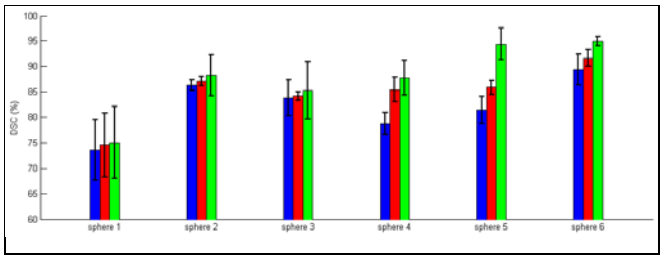

Figure 2. DSC results of applying RW (blue), RW with additional u (red) and our algorithm (green).

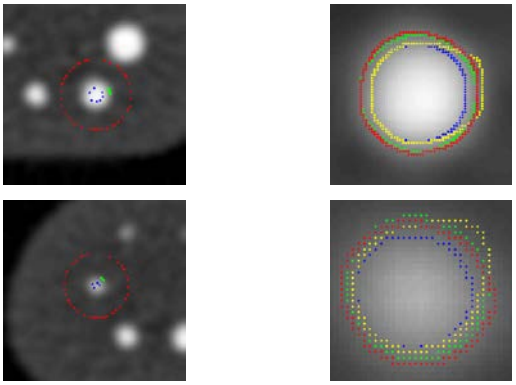

Figure 3. Sample segmentation results on phantom study. Ground truth is delineated in green color. The first row shows seeds location (blue dots for foreground seeds and red dots for background seeds) and segmentation result by RW delineated in blue line. The second row shows the user inference (green strokeline in left column), and the RW segmentation result (yellow line) with these additional seeds. The third row shows the segmentation result of our algorithm.

Figure 4 shows the DSC results of the five segmentation methods on 16 clinical PET studies, and our method could consistently provide better segmentation results than others. Figure 5 shows sample segmentation results of the five segmentation methods on four clinical studies.

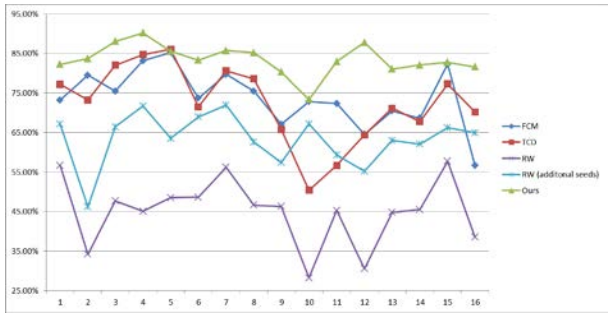

Figure 4. DSC results of 5 segmentation methods on 16 clinical PET studies.

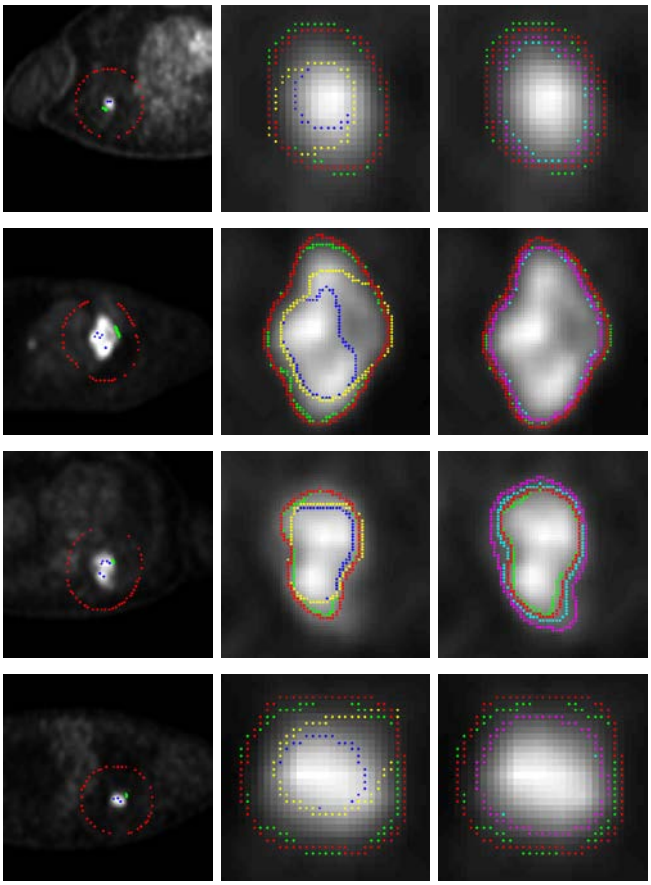

Figure 5. Sample segmentation results on clinical studies. The first column shows seeds locations (foreground in blue, background in red, user inference in green). The second and third column shows segmentation results by RW (blue), RW with user inference (yellow), FCM (cyan), TCD (magenta), our algorithm (red) and ground truth (green).

E. User Inference Conformation

To validate the degree of user inference conformation, inter-observer agreement of the manual delineations from the two experienced radiologist was evaluated with mean DSC 0.820 ± 0.068 . User inference conformation could be reflected from the deviation of the DSC of segmentation compared with both of the ground truth. The lower the mean of this deviation, the more stably the method conformed to user inference. Table 1 shows the evaluation results of user inference conformation for 4 methods on clinical PET studies our method could better conform to user expectation.

TABLE I. USER INFERENCE CONFORMATION

	FCM	TCD	RW	Ours
DSC deviation	0.043±0.042	0.056±0.047	0.038±0.026	0.031±0.023

IV. CONCLUSION

We proposed a statistical method for segmentation uncertainty problem in PET images by analyzing the features of user inference to label the uncertainty band. Our experimental validation on phantom and clinical PET studies demonstrated that our algorithm provided more accurate results and better conformed to user expectation.

REFERENCES

[1] *RIDER Phantom PET-CT: UNIVERSITY OF WASHINGTON*. Available: https://wiki.cancerimagingarchive.net/display/Public/RIDER+Collections

[2] B. Foster, U. Bagci, A. Mansoor, Z. Xu, and D. J. Mollura, "A review on segmentation of positron emission tomography images," *Computers in Biology and Medicine*, vol. 50, pp. 76-96, 7/1/ 2014.

[3] P. G. C. Philipp A. Kaufmann, "Myocardial Blood Flow Measurement by PET: Technical Aspects and Clinical Applications," *Journal of Nuclear Medicine*, vol. 46, pp. 75-88, 2005.

[4] ACS. (2014). *American Cancer Society How is non-small cell lung cancer diagnosed?*

[5] M. Vanderhoek, S. B. Perlman, and R. Jeraj, "Impact of the Definition of Peak Standardized Uptake Value on Quantification of Treatment Response," *Journal of Nuclear Medicine*, vol. 53, pp. 4-11, 2012.

[6] I. S. G. Quinten C. Black, Larry L. Kestin, Ching-Yee O. Wong, John W. Wong, Alvaro A. Martinez, Di Yan, "Defining a radiotherapy target with positron emission tomography," *International Journal of Radiation Oncology Biology Physics*, vol. 60.

[7] F. L. Jentzen W, Eising EG, Heinze M, Brandau W, Bockisch A., "Segmentation of PET volumes by iterative image thresholding," *Journal of Nuclear Medicine*, vol. 48, pp. 108-114, 2007.

[8] P. Tylski, G. Bonniaud, E. Decenciere, J. Stawiaski, J. Coulot, D. Lefkopoulos, *et al.*, "(18)F-FDG PET images segmentation using morphological watershed: a phantom study," presented at the 2006 IEEE Nuclear Science Symposium Conference Record, 2006.

[9] C. Ballangan, X. Wang, M. Fulham, S. Eberl, Y. Yin, and D. Feng, "Automated Delineation of Lung Tumors in PET Images Based on Monotonicity and a Tumor-Customized Criterion," *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, pp. 691-702, 2011.

[10] X. Wang, H. Cui, C. Ballangan, and D. D. Feng, "Lung tumor segmentation and separation from PET volumes based on Tumor-Customized Downhill," presented at the IEEE-EMBS International Conference on Biomedical and Health Informatics, 2012.

[11] U. Bagci, J. K. Udupa, N. Mendhiratta, B. Foster, Z. Xu, J. Yao, *et al.*, "Joint segmentation of anatomical and functional images: Applications in quantification of lesions from PET, PET-CT, MRI-PET, and MRI-PET-CT images," *Medical Image Analysis*, vol. 17, pp. 929-945, 2013.

[12] H. Wechsler and M. Kidode, "RANDOM-WALK PROCEDURE FOR TEXTURE DISCRIMINATION," *Ieee Transactions on Pattern Analysis and Machine Intelligence*, vol. 1, pp. 272-280, 1979.

[13] L. Grady, "Random walks for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 1768-1783, 2006.

[14] L. Grady, T. Schiwietz, S. Aharon, and M. Westermann, "Random walks for interactive organ segmentation in two and three dimensions: Implementation and validation," presented at the Medical Image Computing and Computer-Assisted Intervention, 2005.

[15] M. Chen, E. Helm, N. Joshi, S. M. Brady, and Ieee, "Random Walk-based Automated Segmentation For The Prognosis of Malignant Pleural Mesothelioma," presented at the IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2011.

[16] L. Grady, "Multilabel random walker image segmentation using prior models," presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005.

[17] X. W. Hui Cui, Jianlong Zhou, Stefan Eberl, Yong Yin, Dagan Feng, and Michael Fulham, "Topology polymorphism graph for lung tumor segmentation in PET-CT images," *Physics in Medicine and Biology*, 2015.

[18] W. Yang, J. Cai, J. Zheng, and J. Luo, "User-Friendly Interactive Image Segmentation Through Unified Combinatorial User Inputs," *Image Processing, IEEE Transactions on*, vol. 19, pp. 2470-2479, 2010.

[19] D. W. G. Montgomery, A. Amira, and H. Zaidi, "Fully automated segmentation of oncological PET volumes using a combined multiscale and statistical model," *Medical Physics*, vol. 34, pp. 722-736, Feb 2007.

[20] C. Ballangan, W. Xiuying, M. Fulham, S. Eberl, Y. Yong, and F. Dagan, "Automated Delineation of Lung Tumors in PET Images Based on Monotonicity and a Tumor-Customized Criterion," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 15, pp. 691-702, 2011.