A Graph-Based Method for PET Image Segmentation in Radiotherapy Planning: A Pilot Study

Alessandro Stefano[1,2], Salvatore Vitabile[3], Giorgio Russo[2], Massimo Ippolito[4], Daniele Sardina[5], Maria G. Sabini[5], Francesca Gallivanone[6], Isabella Castiglioni[6], and Maria C. Gilardi[6,7]

[1] Dipartimento di Ingegneria Informatica, University of Palermo, Palermo, Italy
[2] IBFM CNR - LATO, Cefalù (PA), Italy
[3] Dipartimento di Biopatologia e Biotecnologie Mediche e Forensi, University of Palermo, Palermo, Italy
[4] Nuclear Medicine Department, Cannizzaro Hospital, Catania, Italy
[5] Medical Physics Unit, Cannizzaro Hospital, Catania, Italy
[6] IBFM CNR, Segrate (MI), Italy
[7] University of Milano-Bicocca, Milano, Italy

Abstract. Target volume delineation of Positron Emission Tomography (PET) images in radiation treatment planning is challenging because of the low spatial resolution and high noise level in PET data. The aim of this work is the development of an accurate and fast method for semi-automatic segmentation of metabolic regions on PET images. For this purpose, an algorithm for the biological tumor volume delineation based on random walks on graphs has been used. Validation was first performed on phantoms containing spheres and irregular inserts of different and known volumes, then tumors from a patient with head and neck cancer were segmented to discuss the clinical applicability of this algorithm. Experimental results show that the segmentation algorithm is accurate and fast and meets the physician requirements in a radiotherapy environment.

Keywords: Segmentation, Graph, PET, Head and Neck cancer, Radiotherapy.

1 Introduction

Radiotherapy is the mainstay of treatment for head and neck cancer (HNC) and the computerized tomography (CT) is considered to be the gold standard for staging, target tumor delineation, dose calculation, monitoring and evaluation of treatment response. However, results in HNC radiotherapy are still disappointing due to the radio resistance of the tumor and/or inadequate dose to target for geographic miss. CT imaging is based on variation of tissue density and provides anatomical information with a high resolution. However, it has several limitations due to the insufficient contrast between normal tissue and tumor. It don't provide any useful information about the tumor biology. CT imaging may not show the viable extension of tumors and not localize isolated positive lymph node. To overcome these limitations, Positron Emission Tomography (PET) imaging has gained a fundamental impact in many fields of oncology: molecular imaging enables to visualize metabolic features of oncological

A. Petrosino (Ed.): ICIAP 2013, Part II, LNCS 8157, pp. 711–720, 2013.

lesions, providing an *in vivo* measure of the tumor biological processes. In addition, metabolic changes are often faster and more indicative of the effects of the therapy with respect to morphological changes [1]. The 18F-fluoro-2-deoxy-D-glucose (FDG) is the radiotracer commonly used in PET acquisition for oncology. FDG is an analogue of glucose and allows to evaluate tumor glucose metabolism that is altered in most of the neoplastic pathologies. FDG is then avidly accumulated by most malignant tumors, allowing to identify the location of the primary tumor and metastases. This offers the opportunity to radically change the patient treatment (i.e. from radiotherapy to chemotherapy) or the planning treatment volume (PTV) in radiotherapy field. Current generation PET/CT systems allow the integration of functional and morphological data. Within the CT gross tumor volume (GTV), defined on anatomical images, it is possible to define target volumes based on functional area (biological tumor volume - BTV) and to apply a strategy that will deliver radiation to these regions. BTV varies substantially depending on the algorithm used to delineate functional signal in PET images. Visual delineation is widely-used, but it is strongly operator-dependent, even if easily applicable. For this reason, the development and implementation of robust, fast, accurate, operator and scanner independent segmentation methods is mandatory.

In this paper a semi-automated approach based on Random Walks on graphs (RWg) is proposed [2, 3]. The method has been tested on phantom studies in order to assess the accuracy respect to region growing (RG) standard approach. To assess the applicability in a clinical environment, a pilot patient study was also considered.

The paper is organized as follows. In section 2 the current state of the art in PET image segmentation techniques is reviewed and the RWg algorithm is introduced. In section 3, PET phantom and patient protocols are described. In section 4, RWg delineation is evaluated in order to assess the accuracy and the applicability in clinical environment.

2 Related Works

Over the past years, various automatic or semi-automatic approaches, based on fixed, adaptive or iterative threshold, fuzzy c-means (FCM), region growing or watershed segmentation, have been proposed for PET image segmentation but few validation studies are available and there is no consensus for proper BTV delineation method with no clear guideline on how to incorporate metabolic data into target delineation [4]. Moreover, PET delineation approaches can be categorized on the basis of anatomical sites or used radiotracers. According to a comprehensive review of segmentation algorithms in PET imaging [5], four segmentation methodology categories can be identified. Image thresholding methods are the most widely used due to their simplicity to implement but they are too sensitive to image noise and heterogeneity [6]. Variational approaches based on gradient differences between target and background regions are mathematically efficient but sensitive to image noise and subject to numerical fluctuation [7]. Learning methods as artificial neural network, support vector machine, k-means algorithm, fuzzy C-means algorithm are efficient but require high computational steps and are sensitive to variability of PET radiotracer depending on

study protocol, as for example scanner characteristics, radiotracer injected dose and interval between radiotracer injection and exam start. Finally, stochastic models based on statistical differences in intensity distribution between lesion and normal tissues can be considered optimal for noisy images under condition that a proper noise model is used.

2.1 Random Walks for Image Segmentation

A graph-based approach provides a foreground and background recognition in images in which the seeds are specified by user input. An undirected graph G can be represented as a pair G = (V,E) with nodes v ∈ V and edges e ∈ E ⊆ VxV. A node v_i is a neighbor of another node v_j if they are connected by an edge e_{ij} with a weight w_{ij} (w_{ij}=w_{ji} being an undirected graph). The graph-based segmentation method represent DICOM (Digital Imaging and Communications in Medicine) image as a graph in which the voxels are its nodes and the edges are defined by a cost function which maps a change in image intensity to edge weights. The image is then converted into a lattice where some pixels are known (nodes with label specified by user input) and some pixels are not known. The delineation problem is to assign a label to unknown nodes. This is done by trying to find the minimum cost/energy among all possible scenarios in the graph to provide an optimal segmentation: RWg algorithm can be used to partition the nodes into two disjoint subsets representing lesion and background. RWg algorithm appeared in computer vision domain and then extended for image segmentation [3]. This approach is an efficient and accurate method in low contrast images characterized by noise and weak edges as PET images. RWg, respect to graph-cut algorithm [8], is less susceptible to the "small cut" behaviors.

The RWg problem is to determine the highest probabilities for each pixel to reach the target node (target seed) and has the same solution as the combinatorial Dirichlet problem [3]:

$$D[x] = \frac{1}{2} x^T L x \tag{1}$$

where L indicates the graph's Laplacian matrix and x the vector of the probabilities that each voxel is included in target region. RW measures the "betweenness" through starting pixel (foreground seed) to the un-labeled pixel determining the highest probabilities for assigning labels to the nodes [2].

3 Materials and Methods

3.1 Phantom Study

Both quality control and anthropomorphic phantoms were used to estimate the accuracy of the PET segmentation algorithms: 1) The NEMA IEC body phantom consisting of an elliptical cylinder (D1 = 24 cm, D2 =30 cm, h= 21cm) including six spheres of different diameters (d1= 10 mm, d2 = 13 mm, d3 = 17 mm, d4 = 22 mm,

d5 = 26 mm, d6 = 37 mm) positioned at 5,5 cm from the center of the phantom; 2) an anthropomorphic oncological phantom (elliptical cylinder of D1 = 20 cm, D2 =30 cm, h= 21cm) simulating thorax body regions and oncological lesions (three different spheres of d2 = 15.6 mm, d3 = 25.8 mm, d4 = 31.3 mm and two irregular inserts of 50 and 130 cm3, respectively). Spheres, inserts and background were filled with FDG with a ratio between measured sphere radioactivity concentration and measured background radioactivity concentration (S/B) that ranged from 1.5 to 11 for ten independent experiments. In particular, seven experiments were performed with the NEMA IEC body phantom with S/B that ranged from 1.5 to 11 at two different matrix sizes and three experiments were performed with the anthropomorphic oncological phantom with S/B that ranged from 5 to 10.

3.2 Clinical Study

A clinical case was selected to evaluate clinical applicability of the RWg segmentation algorithm. A 80 years old male with HNC that had been referred for a diagnostic PET/CT scan before radiotherapy treatment was enrolled. Patient fasted for twelve hours before PET exam and was intravenous injected with FDG. The PET/CT oncological protocol began 60 minutes after the injection. PET/CT scan was performed in diagnostic position with patient on a flat carbon bed similar to the radiotherapy treatment couch (in replacement of the concave couch usually used for patient comfort in diagnostic PET/CT exam) and a thermoplastic mask was used for immobilization of the head. The patient breathed normally during the PET and CT exams, and scanning was executed from the top of the skull to the middle of the thigh with the arms along the body.

3.3 Data Acquisition

Phantoms and the patient acquisition were performed on Discovery 690 with time of flight and Discovery STE PET/CT scanners by General Electric Medical Systems [9].

PET images of five experiments with NEMA IEC body phantom consisted into a matrix of 256x256 voxels of 2.73x2.73x3.27 mm^3 voxel size, while CT images consisted into a matrix of 512x512 voxels of 1.36x1.36x3.75 mm^3 voxel size (Discovery 690). PET images of two experiments with NEMA IEC body phantom, of three experiments with the anthropomorphic oncological phantom, and of clinical study consisted into a matrix of 128x128 voxels of 4.7x4.7x3.27 mm^3 voxel size, while CT images consisted into a matrix of 512x512 voxels of 0.97x0.97x3.27 mm^3 voxel size (Discovery STE).

3.4 Manual Segmentation

In phantom studies, the actual sizes of the used spheres were known and a manual segmentation was not required. In patient study, the PET/CT images were reported by a nuclear medicine physician for diagnostic and staging purposes. The CT gross tumor volume (GTV) was manually outlined by the radiation oncologist on each CT

transverse section of lesion. Metabolic information was used to modulate the target volume delineation. The active tumor volume was manually defined by the radiation oncologist in consensus with the nuclear medicine physician: BTV included the tumor volume with an intense tracer uptake respect to background FDG activity level.

3.5 Automatic Segmentation

The method based on graph theory proposed by Grady [3] was used. RWg parameters have been modulated to adapt to PET application domain in order to incorporate metabolic information in radiotherapy treatment. The Gaussian weighting function for PET image was defined as:

$$w_{ij} = \exp\left(-\beta\left(SUV_i - SUV_j\right)\text{^}2\right) \tag{2}$$

where

$$SUV_i = \frac{Concentration}{\left(\frac{Injected\ dose}{Patient's\ weight}\right)} \tag{3}$$

indicates the body-weight Standardized Uptake Value (SUV) of voxel i. SUV is a widely used PET semi-quantitative parameter, calculated as a ratio of tissue radioactivity concentration (kBq/ml) and FDG injected dose (MBq) at the time of injection divided by body weight (kg). In our experiments, the β weighting factor was set to 50. Hence the PET image is converted into a lattice where SUV of each voxel is mapped to wij. To create a semi-automatic delineation method with RWg embedded, an automatic background region localization to identify background seed is implemented. Then, the operator chooses the best slice containing the target lesion in order to identify the target seed with a single mouse click. This approach offers a greater flexibility and ability to properly delineate the PET lesion, excluding false positives.

3.6 Evaluation

The effectiveness of the proposed method has been evaluated comparing to well-known and commonly used Region Growing (RG) method, calculating the difference between actual sphere (or insert) sizes and semi-automatic PET segmentations. In addition, we used the precision, recall, and dice similarity coefficient (DSC) to prove the effectiveness of the proposed approach [10, 11]. In a phantom study the morphological region must match with metabolic region. This is not true in patient studies: in fact metabolic volume cannot match the tumor anatomic extension, showing different and additional information. Hence the patient study was mainly used to assess the applicability in a clinical environment of the RWg algorithm. Furthermore, we will be make a comparison between manual and semi-automatic BTV segmentations. The average of the time for delineating spheres, irregular inserts and oncological lesions was recorded to assess algorithm performances. RWg and RG algorithms were implemented on the Matlab R2012b simulation environment, running on a general

purpose PC with a 3.00GHz Intel R CoreTM i5-2320 processor, 4 GB memory, and 64-bit Windows 7 Professional.

4 Experimental Results

4.1 Trials and Results on Phantoms

For each sphere of the phantoms, the percentage difference was calculated as:

$$E\% = ABS\left(100 * \left(\frac{Measured\ Size - Actual\ Size}{Actual\ Size}\right)\right) \quad (4)$$

at different S/B. To avoid negative values in the relative differences as results of an under sizing in PET delineation, absolute value (ABS) was applied. Table 1 shows the precision (P), recall (R), and DSC averaged on different spheres in each NEMA IEC body phantom experiment. Phantoms (a-e) were studied using the Discovery 690 with a sampling matrix of 256x256 voxels and a S/B of 1.5-2 for the NEMA (a), 2-3 for the NEMA (b), 3-5 for the NEMA (c), 5-6 for the NEMA (d), 6-7 for the NEMA (e). Phantoms (f-g) were studied using the Discovery STE with a sampling matrix of 128x128 voxels and a S/B range of 3.5-9 for the NEMA (f), 9-11 for the NEMA (g).

Table 1. Precision (P), Recall (R), and DSC in NEMA experiments

	RWg				RG		
	P	R	DSC	-:-	P	R	DSC
(a)	0.976	0.934	0.952		0.853	1.000	0.921
(b)	0.988	0.949	0.968		0.899	0.929	0.908
(c)	0.994	0.960	0.976		0,951	0,899	0,918
(d)	0.982	0.975	0.978		0,916	0,987	0,946
(e)	0.972	0.983	0.977		0,963	0,964	0,961
(f)	0.961	0.968	0.963		0.913	0.897	0.886
(g)	0.950	0.965	0.957		0.907	0.920	0.893

The E% range of RWg algorithm was found to be from 0.52% up to 16.85% without any restriction in sphere size and in S/B. The range reduced from 0.52% up to 5.38% for the spheres with a diameter > 1.7 cm. The minimum error (0.52%) was obtained in the sphere with a diameter of 3.7 cm and with a S/B of 5 (c). For the spheres with a diameter < 1.7 cm the range was found to be from 4.55% up to 16.85%. The maximum error (16.85%) was obtained in the smaller sphere of NEMA (g) experiment with a S/B of 3.5. RWg algorithm failed in the smaller sphere segmentation at very low S/B ((a) and (b)) where P, R, DSC and E% were obtained by considering the five spheres with a diameter > 1 cm. The average of the time for sphere volume segmentation was around 0.8 seconds in 128x128 PET images (one slice in around 0.1 seconds) and around 2 seconds in 256x256 PET images (one slice in around 0.2 seconds). The E% range of PET delineation using RG algorithm was

found to be from 9.35% up to 36.32% for the sphere < 17 mm diameter and from 0.66% up to 20.40% for the sphere > 17 mm diameter, respectively. RG algorithm failed to delineate spheres with a diameter of 1 cm. In the experiment with a S/B<2 (a) only the sphere with a diameter of 3.7 cm was segmented. The average of the time for sphere volume delineation by the RG algorithm was around 0.9 seconds (one slice in around 0.11 seconds) in both 128x128 and 256x256 PET images. As a representative example of above mentioned results, figure 1 shows PET delineations reported in PET/CT image for one of the NEMA experiments (c).

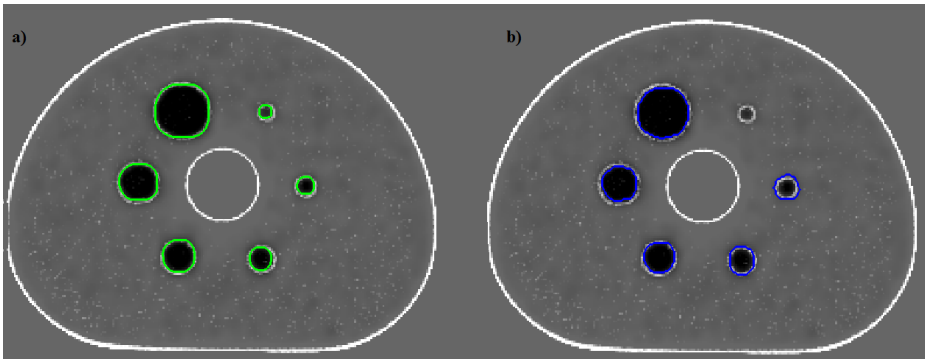

Fig. 1. Segmentations of spheres showed in PET/CT image: (a) RWg DSC = 0.976±0.023 (averaged on six spheres), (b) RG DSC = 0.918±0.034 (averaged on five spheres)

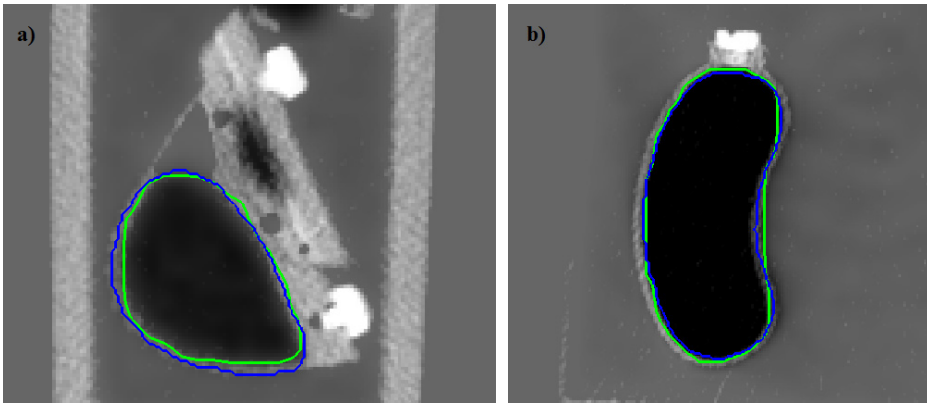

Fig. 2. Irregular insert segmentations in PET/CT images: (a) RWg DSC (green line) = 0.981 versus RG DSC (blue line) = 0.964 (b) RWg DSC = 0.962 versus RG DSC = 0.990

Table 2 shows the P, R, and DSC indexes averaged on different spheres and inserts in the three anthropomorphic oncological phantom experiments with a S/B of 5, 7, and 10. Results showed that RWg technique estimate the actual size with an error < 10% for spherical lesions and with an error < 4% for the two large irregular inserts (volume > 50 cm3). Figure 2 shows two examples of insert delineation using RWg

(green line) and RG (blue line) methods. The average of the time for segmentation of irregular inserts was around 0.1 seconds (RWg) versus 0.4 seconds (RG) for single slice.

Table 2. P, R, and DSC in anthropomorphic phantom studies

	RWg				RG		
	P	R	DSC	-:-	P	R	DSC
Spheres	0,974	0,979	0,975		0,998	0.844	0,913
Inserts	0,984	0,976	0,979		0,958	0,969	0,962

4.2 Trials and Results on the Pilot Patient Study

Figure 3.a shows lesions in CT image: semiautomatic BTV was located outside the GTV domain in a metabolic positive lymph node. In this clinical case, manual BTV radically changed the treatment volume because uptake was found outside the GTV in a involved lymph node (not CT visible). The figure 3.b shows the corresponding PET image with manual BTV (blue line) and semiautomatic BTV (green line): manual BTV was larger than the semiautomatic BTV (<10%, see table 3). An analysis of the time performance of the considered technique showed that the algorithm is fast: the volume, consisting of 13 slices, was segmented in around 1.4 seconds.

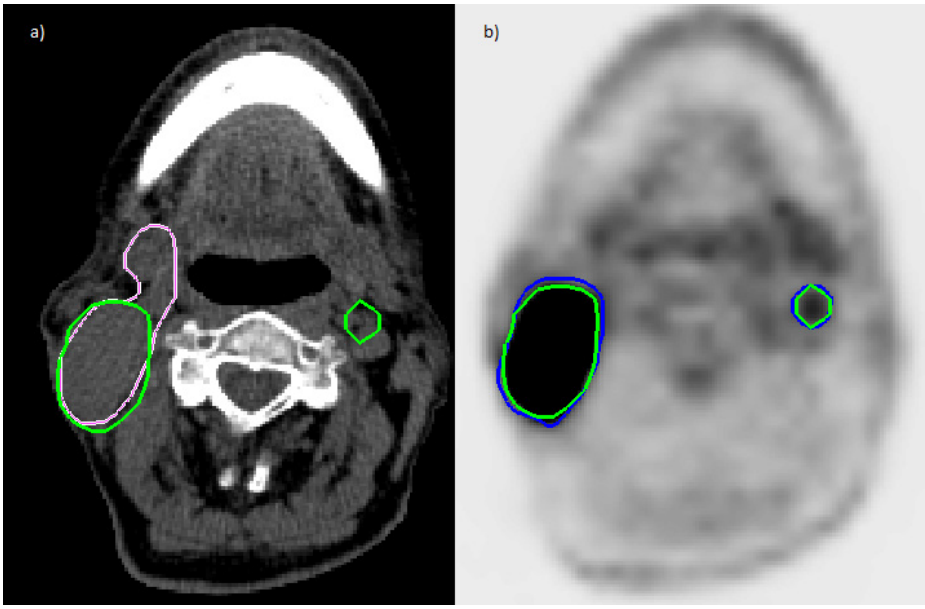

Fig. 3. Differences between GTV (magenta line), semiautomatic BTV (green line) and manual BTV (blue line)

Table 3. Comparison between manual and RWg segmentations

	Primary lesion	Lymph node
Manual	31.75 cm^3	2.73 cm^3
RWg	29.57 cm^3	2.48 cm^3

5 Discussion and Conclusions

PET segmentation in radiotherapy is a critical task due to the low image resolution. To date, qualitative visual interpretation is the most commonly used method. Unfortunately, the manual contouring limits the accuracy and reproducibility of measurement. Several automatic methods have been presented, like threshold-based or gradient-based algorithms, but few clinical studies are available and there is no consensus for proper BTV determination.

The aim of this work was to validate a FDG-PET image segmentation method based on RWg. A pilot patient study was used to assess its applicability in clinical practice. The algorithm segments PET images from SUV and it is very fast (one slice in 0.1 - 0.2 seconds) in a clinical environment if compared against the time needed for manual segmentation. The accuracy of RWg segmentation method was higher than RG segmentation method in phantom studies. This was evident for the smaller spheres despite a drop in the RWg precision for lesions with diameter < 1cm and at low contrast. This was compatible with the severe errors in the volume estimation reported for small tumor volume [12]. Partial volume effect is one of the most important factors impacting the quality and the quantitative accuracy in PET imaging [13]. The images are blurred due to the limited spatial resolution of PET scanner and small lesions appear larger. Several corrective techniques have been developed and a partial volume correction method could be included in the algorithm, such as that described in [14]. Increasing the target size, RWg time performance and accuracy remain steady, while RG accuracy increases and time performance decreases. Moreover, RWg method provided resolution independent results considering the two set of images tested. At last, no false positive regions were segmented, because of the possibility to manually identify the target lesion. In conclusion, the proposed method is very powerful in terms of noisy image segmentation and time performance.

In the pilot patient exam, FDG-PET has been proved to modify GTV size, location, and shape changing radically radiation treatment strategy and leading to the opportunity to prevent potential disease progression. In many cases, such as the one presented in the paper, qualitative interpretation and manual contouring are sufficient to obtain fundamental information for patient care, including invisible metastases using traditional radiologic techniques. However, a more objective assessment to estimate BTV for radiotherapy treatment with the implementation of automatic algorithms is mandatory and the RWg algorithm meets the requirements in a radiotherapy environment. The results shown in table 3 underline a difference of 10% between manual and RWg area sizes. However, the manual segmentation is high operator dependent.

As this was only a pilot study, further investigations are required with a larger number of patients in order to assess the prognostic usefulness and long-term clinical impact to correlate BTV segmentation with clinical outcomes, progression-free survival and overall survival.

References

1. Wahl, R.L., Jacene, H., Kasamon, Y., Lodge, M.A.: From RECIST to PERCIST: Evolving Considerations for PET Response Criteria in Solid Tumors. Journal of Nuclear Medicine 50, 122S–150S (2009)
2. Bagci, U., Yao, J., Caban, J., Turkbey, E., Aras, O., Mollura, D.J.: A graph-theoretic approach for segmentation of PET images. In: Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 8479–8482 (2011)
3. Grady, L.: Random walks for image segmentation. Ieee Transactions on Pattern Analysis and Machine Intelligence 28, 1768–1783 (2006)
4. Schinagl, D.A.X., Vogel, W.V., Hoffmann, A.L., Van Dalen, J.A., Oyen, W.J., Kaanders, J.H.A.M.: Comparison of five segmentation tools for 18 F-FLUORO-DEOXYGLUCOSE-POSITRON emission tomography-based target volume definition in head and neck cancer. International Journal of Radiation Oncology Biology Physics 69, 1282–1289 (2007)
5. Zaidi, H., El Naqa, I.: PET-guided delineation of radiation therapy treatment volumes: a survey of image segmentation techniques. European Journal of Nuclear Medicine and Molecular Imaging 37, 2165–2187 (2010)
6. Kao, C.-H., Hsieh, T.-C., Yu, C.-Y., Yen, K.-Y., Yang, S.-N., Wang, Y.-C., Liang, J.-A., Chien, C.-R., Chen, S.-W.: F-18-FDG PET/CT-based gross tumor volume definition for radiotherapy in head and neck Cancer: a correlation study between suitable uptake value threshold and tumor parameters. Radiation Oncology 5 (2010)
7. Wanet, M., Lee, J.A., Weynand, B., De Bast, M., Poncelet, A., Lacroix, V., Coche, E., Gregoire, V., Geets, X.: Gradient-based delineation of the primary GTV on FDG-PET in non-small cell lung cancer: A comparison with threshold-based approaches, CT and surgical specimens. Radiotherapy and Oncology 98, 117–125 (2011)
8. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. Ieee Transactions on Pattern Analysis and Machine Intelligence 23, 1222–1239 (2001)
9. Bettinardi, V., Mancosu, P., Danna, M., Giovacchini, G., Landoni, C., Picchio, M., Gilardi, M.C., Savi, A., Castiglioni, I., Lecchi, M., Fazio, F.: Two-dimensional vs three-dimensional imaging in whole body oncologic PET/CT: a Discovery-STE phantom and patient study. Quarterly Journal of Nuclear Medicine and Molecular Imaging 51, 214–223
10. Han, D., Bayouth, J., Song, Q., Taurani, A., Sonka, M., Buatti, J., Wu, X.: Globally optimal tumor segmentation in PET-CT images: a graph-based co-segmentation method. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 245–256. Springer, Heidelberg (2011)
11. Parker, B.J., Feng, D.G.: Graph-based Mumford-Shah segmentation of dynamic PET with application to input function estimation. Ieee Transactions on Nuclear Science 52, 79–89
12. Tylski, P., Stute, S., Grotus, N., Doyeux, K., Hapdey, S., Gardin, I., Vanderlinden, B., Buvat, I.: Comparative Assessment of Methods for Estimating Tumor Volume and Standardized Uptake Value in F-18-FDG PET. Journal of Nuclear Medicine 51, 268–276
13. Soret, M., Bacharach, S.L., Buvat, I.: Partial-volume effect in PET tumor imaging. Journal of Nuclear Medicine 48, 932–945 (2007)
14. Gallivanone, F., Stefano, A., Grosso, E., Canevari, C., Gianolli, L., Messa, C., Gilardi, M.C., Castiglioni, I.: PVE Correction in PET-CT Whole-Body Oncological Studies From PVE-Affected Images. Ieee Transactions on Nuclear Science 58, 736–747 (2011)