# Accurate automatic delineation of heterogeneous functional volumes in positron emission tomography for oncology applications.

Mathieu Hatt, Catherine Cheze Le Rest, Patrice Descourt, André Dekker, Dirk De Ruysscher, Michel Oellers, Philippe Lambin, Olivier Pradier, Dimitris Visvikis

# Accurate automatic delineation of heterogeneous functional volumes in positron emission tomography for oncology applications

Mathieu Hatt*, Ph.D.[1], Catherine Cheze le Rest, Ph.D., M.D.[1,2], Patrice Descourt, Ph.D.[1], André Dekker, Ph.D.[3], Dirk De Ruysscher, Ph.D., M.D.[3], Michel Oellers, Ph.D.[3], Philippe Lambin, Ph.D., M.D.[3], Olivier Pradier, Ph.D., M.D.[1,4], Dimitris Visvikis, Ph.D.[1].

**Running title**: Automatic PET image segmentation for oncology applications

[1] INSERM, U650, LaTIM

[2] Academic department of Nuclear Medicine

[3] MAASTRO Clinic

Dr. Tanslaan 12, 6229 ET Maastricht, The Netherlands

[4] Institute of Oncology

CHU Morvan, Brest F-29200, France

*Corresponding author:

Mathieu HATT

INSERM U650, LATIM, CHU MORVAN, 5 avenue Foch, 29609 Brest

e-mail: hatt@univ-brest.fr

Tel.: +33298018111

Fax: +33298018124

**CONFLICT OF INTEREST : none**

ABSTRACT

**Purpose**: Accurate contouring of PET functional volumes is now considered crucial in image-guided radiotherapy and other oncology applications, as the use of functional imaging allows biological target definition. In addition, the definition of variable uptake regions within the tumour itself may facilitate dose painting for dosimetry optimization.

**Material and methods**: Current state-of-the-art algorithms for functional volume segmentation use adaptive thresholding. We have developed an approach named Fuzzy Locally Adaptive Bayesian (FLAB), validated on homogeneous objects. We improved it for the delineation of inhomogeneous tumours by allowing the use of up to three classes. Simulated and real tumours with histology data containing homogeneous and heterogeneous activity distributions were used to assess the algorithm's accuracy.

**Results**: The new 3-FLAB algorithm is able to extract the overall tumour from the background tissues, as well as delineate variable uptake regions within the tumours, with higher accuracy and robustness compared to adaptive threshold ($T_{bckg}$) and fuzzy C-means (FCM). 3-FLAB performed with a mean classification error of less than 9±8% on the simulated tumours whereas binary-only implementation led to errors of 15±11%. $T_{bckg}$ and FCM lead to mean errors of 20±12% and 17±14% respectively. 3-FLAB also lead to more robust estimation of the maximum diameters of tumours with histology measurements, with less than 6% standard deviation whereas binary FLAB, $T_{bckg}$ and FCM lead to 10%, 12% and 13% respectively.

**Conclusion**: These encouraging results warrants further investigation in future studies that will investigate the impact of 3-FLAB in radiotherapy treatment planning, diagnosis and therapy response evaluation

**Key Words:** Heterogeneous functional volumes delineation; automatic segmentation; image-guided radiotherapy; dose painting.

## 1. Introduction

While most of positron emission tomography (PET) clinical applications rely on manual and visual analysis, accurate functional volume delineation in PET is crucial for numerous oncology applications. These include the use of tumour volume and associated determination of semi-quantitative indices of activity concentration for diagnosis and therapy response evaluation *(1)*, or the definition of target volumes in intensity-modulated radiation therapy (IMRT) *(2)*. Subjective *(1)* and tedious manual delineation cannot perform accurate and reproducible segmentation especially when considering complex shapes and non-homogeneous uptake. This results from the low quality of PET images notably due to statistical noise and partial volume effects (PVE) (3) arising from the scanner's limited spatial resolution.

Most of the previously proposed methods for PET volume definition are semi-automatic and threshold-based, using either fixed (30-75% of the maximum activity) *(2,4,5)* or adaptive approaches incorporating the background activity *(6-10)*. Unfortunately, these approaches often require additional *a priori* information and are user- and system-dependent. They require manual background regions of interest (ROIs) and their performance depend on parameters requiring optimization using phantom acquisitions for each scanner and reconstruction. Finally all of these approaches are strictly binary and were not validated considering heterogeneous volumes.

On the other hand, numerous works have addressed PET lesion segmentation using more advanced image segmentation methodologies *(11-19)*. However, the majority of these approaches often depend on pre- or post-processing steps like deconvolution and/or denoising, are often binary only and validated on phantom acquisitions or clinical data without rigorous ground-truth.

We have previously developed an algorithm for PET volume definition by combining a fuzzy measure with a locally adaptive Bayesian-based classification (FLAB) that has been shown to perform better with respect to fixed thresholding, Fuzzy C-Means (FCM) or Fuzzy Hidden Markov Chains (FHMC) for PET volume definition, as far as homogeneous spheres or slightly heterogeneous and non spherical tumours are concerned *(20)*. Preliminary results show that FLAB is also robust with respect to variability of the acquisition and reconstruction parameters *(24)*.

Clinical tumours may be characterized by heterogeneous uptake, thus demanding a non-binary approach for an accurate segmentation that may have a significant impact in defining biological target volumes for dose painting *(21)*. The goals of this work were to (a) improve the FLAB model by incorporating the use of three hard classes and three fuzzy transitions and (b) evaluate its accuracy on real (with known diameter measured in histology) and simulated (with known ground-truth) datasets containing inhomogeneous tumours.

## 2. Materials and methods

### *2.1 The 3-class fuzzy Bayesian segmentation (3-FLAB)*

The 3-FLAB algorithm is an extension of the previous work considering only a binary segmentation *(20)*. FLAB automatically estimates parameters of interest from the image, maximizing the probability of each voxel to belong to one of the considered classes. This probability is estimated for each voxel as a function of its value and the values of its neighbours relative to the voxels' statistical distributions in the image which corresponds to an estimation of the noise within each class. Hence, each voxel of the volume is considered by the method as a random variable within a Bayesian framework:

$$P(\mathrm{X} \mid \mathrm{Y}) = \frac{P(\mathrm{X},\mathrm{Y})}{P(\mathrm{Y})} = \frac{P(\mathrm{Y} \mid \mathrm{X})P(\mathrm{X})}{P(\mathrm{Y})} \quad (1)$$

where, $P(\mathrm{X} \mid \mathrm{Y})$ is the probability to belong to class X knowing the observation Y. This probability is obtained by the product of $P(\mathrm{Y} \mid \mathrm{X})$ and $P(\mathrm{X})$, corresponding to the noise model and the spatial model respectively. $P(\mathrm{Y} \mid \mathrm{X})$ is estimated considering the statistical distribution of the voxels within each class, whereas $P(\mathrm{X})$ is estimated using a sliding cube of 3x3x3 voxels, hence each voxel's classification is influenced by its neighbours. The parameters to estimate are the mean and variance of each class and the spatial probabilities of each voxel with respect to its neighbours. This is performed iteratively using a stochastic version (SEM) *(25)* of the Expectation Maximization (EM) *(26)* initialized with K-Means *(27)* or Fuzzy C-Means *(28)*. In addition, a fuzzy measure between the

classes was added in order to account for the blur between regions, assuming each voxel may contain a mixture of classes.

The originality of 3-FLAB relative to the previously developed binary-only FLAB *(20)* is the use of three classes and three fuzzy transitions within the model (see fig.1), to deal with both homogeneous and heterogeneous activity distributions. Fig.2 demonstrates the inability of FLAB to handle highly non-uniform activity distributions, where the lower uptake part of the lesion is erroneously considered as part of the background (see fig.2(b)), emphasizing the need to better model heterogeneous activity distributions. 3-FLAB should retain the accuracy and robustness of the original model, being in addition able to accurately handle challenging heterogeneous activity distributions frequently characterizing clinical lesions. The 3-FLAB segmentation workflow is summarized below, while the implementation and mathematical details can be found in the Appendix:

(1). Initialization of both the spatial and noise models parameters. Means and variances of each class are obtained using the K-Means or Fuzzy C-Means. The prior probabilities are fixed at 1/3 for each class.

(2). Iterative estimation using the SEM by stochastic sampling for each voxel according to its posterior probability.

(3). Segmentation by selecting for each voxel the class or fuzzy level that maximizes its posterior probability and fusion of fuzzy levels with each hard class to generate a 2 or 3 class segmentation map.

2.2 Alternative segmentation methodologies used for comparison

We compared the results of the 3-FLAB algorithm with the binary FLAB approach and the Fuzzy C-Means (with 2 or 3 clusters) clustering introduced by Dunn *(28)* and used to segment PET brain tumours in *(13)*, as well as an adaptive thresholding *(6)* ($T_{bckg}$):

$$I_{threshold} = \alpha \times I_{mean} + I_{background} \qquad (2)$$

$I_{mean}$ was obtained by computing the mean of all voxels contained inside an initial threshold at 70% of the maximum and $I_{background}$ by computing the mean of the voxels inside a ROI manually drawn on the

background. $I_{mean}$ and $I_{background}$ were subsequently used to derive a first approximation of the source-to-background contrast. The parameter $\alpha$ was optimised using phantom acquisitions on each scanner used to obtain the data. The adaptive thresholding algorithm was implemented using a region growing approach with the maximum intensity voxel as a seed and iteratively adding 3D-neighbouring voxels if their value was above the threshold calculated using equation (2).

## 2.3 Validation studies

### *2.3.1 Datasets*

Dataset-1 was used to evaluate the performance of the algorithm under realistic imaging conditions. It consists of twenty three-dimensional simulated tumours with variable levels of irregular shape and homogeneous or non-homogeneous uptake distributions derived from tumours in patients undergoing 18F-FDG PET/CT investigations for radiotherapy treatment planning purposes. These images were acquired in 2D and 3D mode using the GE Discovery LS and Philips Gemini PET/CT scanners respectively. Three of these tumours illustrating the range of sizes, shapes and heterogeneities considered are shown in fig.4(a-c). The goal was to produce realistic images of PET tumours while retaining a voxel-based ground-truth in order to compute accurate voxel-based classification errors. Half of the tumours were simulated considering a homogeneous uptake distribution whereas the other half was simulated using significant heterogeneity within the tumour. The procedure followed to generate these images is illustrated in fig.3 and is detailed below.

Each clinical tumour is first manually delineated on the PET image by a nuclear medicine expert, thus creating a voxelized volume that represents the ground-truth of the simulation. The activity levels attributed to each of the tumour parts were derived based on the average activity measured in the same areas of the tumour in the corresponding patient images. This ground-truth tumour structure is subsequently transformed into a Non-Uniform Rational B-Splines (NURBS) volume via Rhinoceros™ (*CADLINK software*), for insertion into the NCAT phantom *(29)* attenuation maps at the same approximate position as where it was located in the patient *(30)*. No respiratory or cardiac motions were considered. Simulations using a model of the Philips PET/CT scanner previously

validated with GATE (Geant4 Application for Tomography Emission) *(31)* were carried out. A total of 45 million coincidences were simulated corresponding to the statistics of a clinical acquisition over a single axial 18 cm field of view *(31)*. Images were subsequently reconstructed using OPL-EM (7 iterations, 1 subset) *(31)* with two different voxel sizes (4x4x4 for the Philips Gemini and 2x2x5 $mm^3$ for the GE Discovery LS) in order to match those used in the corresponding clinical images.

Dataset-2 contains 18 images of lung tumours from patients with histologically proven Non-Small Cell Lung Cancer (clinical Stage Ib–IIIb), acquired on the Siemens Biograph PET/CT scanner and reconstructed using OSEM (4 iterations, 8 subsets), with scatter and CT-based attenuation correction, and 5.31x5.31x5.31 $mm^3$ voxels. These tumours were surgically extracted for a histology study in which their maximum diameter was measured by macroscopic examination *(32)*. These diameters range from 15 to 90 mm (44+/-21). One of these tumours is shown in fig.4(d).

2.3.2 Analysis

As our goal is not the detection of a lesion in the whole image but the accurate estimation of its volume and shape, we assume it has been detected and isolated by the clinician within a 3-D "box" well encompassing the tumour.

Since a ground-truth was available, Classification Errors (CE) were computed. In the case of a 2-class ground-truth, the CE is:

$$CE = \frac{card\{t \mid c_t \neq x_t\}}{card\{t \mid x_t = 1\}} \times 100 \qquad (3)$$

where, $c_t$ is the classification of voxel $t$ and $x_t$ is the true class. Card is the number of elements. This error measurement takes into consideration the spatial distribution of the tumor by considering both background voxels classified as object and object voxels classified as background. Consequently this measure is more appropriate than simple volume estimation which could lead to overall small volume errors associated with largely inaccurate segmentations. In addition, the errors are computed relatively to the size of the object, to avoid biases relative to the size of the processing box. In the case of a 3-class ground-truth, CE may be computed for each of the three classes using equation (4) or with respect to a binarized ground-truth ($2^{nd}$ and $3^{rd}$ class merged) using equation (3).

$$CE_c = \frac{card\{t \mid x_t = c, c_t \neq c\} + card\{t \mid x_t \neq c, c_t = c\}}{card\{t \mid x_t = c\}} \times 100 \quad (4)$$

where, $CE_c$ stands for the Classification Error associated with a given class c.

Two different analyses were conducted using dataset-1: The first considered the entire dataset (both homogeneous and heterogeneous tumours) and CE computed using equation (3), in order to compare overall performances of FLAB (binary only), 3-FLAB, FCM and $T_{bckg}$. The second considered only the ten heterogeneous tumours to compute $CE_2$ and $CE_3$ using equation (4) for 3-FLAB and FCM with three clusters.

The segmentation accuracy on the tumours with histology (dataset-2) was assessed by segmenting the clinical image and subsequently measuring the maximum diameter on the segmented volumes, to compare it with the histology measurement.

**3. Results**

Fig.5 contains one axial slice of the segmentations obtained on three simulated tumours of dataset-1 and one tumour of dataset-2. Fig.6(a) contains the mean classification errors and standard deviation obtained by all the methods on the 20 tumours of dataset-1. FLAB (binary only) performed well on homogeneous tumours but failed as expected on strongly heterogeneous lesions leading to overall errors of 15±11%. 3-FLAB on the other hand produced segmentation maps closer to the ground-truth, both visually and quantitatively, with errors between 5% and 15% (9±8%). FCM (with 2 or 3 clusters) was competitive with respect to 3-FLAB for some tumours, but showed a higher variability (10-40%) and mean error (20±12%). This translated qualitatively in FCM being unable to differentiate two different regions within the tumour as well as being unable to detect discontinuities in the contours (e.g. fig.5(d), 1st row). In addition, for the regions where a transition was present between the high uptake region and the background (e.g. fig.4(d)) the 3-FLAB approach was the only one giving accurate representation of this transition (fig.5(c) versus fig.5(d), last row). $T_{bckg}$ was not able to produce satisfactory segmentation in several cases. Tumours with high overall contrast were approximately extracted from the background (e.g. fig.5(e), rows 2-4). However, as a binary method it is unable to delineate uptake distributions within the tumour. In several cases the heterogeneity was

significant and $T_{bckg}$ lead to significant under-evaluation of the tumour volume (CE up to 60% with a mean of 17%±14%) as it tends to extract the high activity region or parts of the reduced uptake region only (e.g. fig.5(e), 1st row).

Fig.6(b) compares 3-FCM (using three clusters) and 3-FLAB concerning the three-class segmentation of the ten heterogeneous simulated tumours of dataset-1. 3-FCM is less accurate and robust compared to 3-FLAB especially in the case of the delineation of higher activity regions (3rd class), with about twice the mean error and standard deviation (24±20%) of 3-FLAB (11±8%).

Fig.7 contains the mean error and standard deviation with respect to the maximum diameter, computed on the tumour histology database (dataset-2). Whereas all methods gave relatively low mean errors (<-3%), the standard deviation associated with FCM and $T_{bckg}$ (13% and 12% respectively) is about twice the one of 3-FLAB (<6%), while binary FLAB lead to almost 10% standard deviation. The low mean error for all these algorithms is explained by the fact that there were about the same amount of under- and over-estimation of the diameters in this dataset, resulting in an overall low mean error. The standard deviation is here a better indicator of the accuracy obtained on the dataset and demonstrates higher accuracy and robustness for 3-FLAB.

**4. Discussion**

Functional volume delineation represents today an area of interest for multiple clinical (routine and research) applications of PET. Such areas include response to therapy studies and the use of biological tumour volumes in radiotherapy treatment planning. Although several fully automatic algorithms have been recently proposed *(11-20)*, segmentation methodologies currently used in clinical practice are based on the use of fixed and adaptive thresholding *(4-10)*. These algorithms have been shown to accurately determine functional volumes under specific imaging conditions of spherical and homogeneous activity distribution object in phantom studies, as well as been evaluated on clinical images where the ground-truth is unknown. In clinical practice lesions are often heterogeneous in shape and uptake. In order to address these issues we have extended a previously developed algorithm to accurately handle lesions with non-uniform uptake and non-spherical forms. In addition, we have proposed an evaluation framework including both realistic simulated patient lesions and histological

assessment of tumour diameters, allowing the evaluation of segmentation algorithms under standard imaging conditions with the added advantage of knowing the ground-truth.

The inability of the adaptive thresholding considered in this study to accurately segment complex tumours is demonstrated by its poor performance. This is explained by the fact that in case of heterogeneous uptake, the 70% threshold used for the initial estimation of the tumour-to-background contrast may retain only the high uptake region, thus leading to incorrect contrast estimation. On the other hand, if the lesion is small and/or has a small contrast, the 70% threshold may lead to an initial overestimation of the volume of the tumour, hence an underestimation of its uptake and therefore an incorrect estimation of the contrast which the subsequent adaptive thresholding may not be able to compensate for. In addition, the background ROI is user-dependent with a potentially high impact on the result, especially with heterogeneous background. When such a case occurred we systematically selected the ROI which resulted in the lowest error. Finally, the region growing implementation avoids incorporating false positives of the background if they are not connected to the main tumour, especially when the contrast is low and/or if the background is noisy and heterogeneous. However it also makes the algorithm dependent on the seed location and can lead to missing parts of the tumour when several high uptake regions are connected by low uptake regions. FCM can produce binary or 3-class segmentations but its robustness and accuracy are much lower compared to FLAB because it incorporates neither spatial correlation nor noise modeling. One advantage of the $T_{bckg}$ over FCM is its region growing implementation that makes it less susceptible than FCM to the inclusion of high intensity voxels of the background. Therefore FCM usually performs poorer than $T_{bckg}$ for low contrast lesions and noisy images but better for heterogeneous activity distributions within the tumour. On the other hand, 3-FLAB performed accurately even under challenging contrast, noise and heterogeneity conditions, with overall superior performance than the other algorithms considered here.

The need for more than three classes may arise for heterogeneous tumours on a heterogeneous background. However, all the clinical tumours considered in this study were correctly delineated using two or three classes because the contrasts between the heterogeneities within the tumour are usually much higher than those occurring in the background, hence only one hard class may be sufficient to deal with the background whereas two are required to correctly handle the significantly different

uptakes occurring inside the tumour. Eventually the 3-FLAB algorithm could be extended to more than three classes assuming that only pairs of hard classes generate fuzzy transitions. One also has to keep in mind that using more classes will lead to smaller regions, but those regions within the tumor will subsequently be used for quantification or radiotherapy dose boosting and/or painting and should therefore be kept reasonably large. The potential impact of using three classes proposed by 3-FLAB should therefore be investigated before more complex segmentations using additional classes can be considered.

We have already demonstrated that FLAB performs well for small lesions down to 13 mm in diameter *(20)* and this study was not designed to specifically investigate the ability of 3-FLAB to deal with small tumours since these rarely exhibit heterogeneous uptake that can be detected on the PET image considering the existing resolution limits. 3-FLAB retains all the characteristics of FLAB but adds the ability to consider a third class and therefore handle non-uniform lesion activity distributions. Thus 3-FLAB does not as such improve the delineation of small (<2cm) lesions. However the higher/lower uptake regions within the larger tumours are often of small size comparable to that of small lesions, with PVE affecting them with respect to their "background" which is in fact the other part of the tumour with different uptake. As fig.6(b) demonstrates 3-FLAB is capable to accurately segment these regions.

An application that could greatly benefit from the use of FLAB is radiotherapy treatment planning *(33)*. It is now acknowledged that planning based on PET/CT volumes improves tumour delineation by reducing inter- and intra-observer variability *(32,34)*. It can also lead to the inclusion of regions not visible on CT, or the exclusion of regions without significant uptake *(35)*. Using the 3-FLAB algorithm could help lower the inter- and intra-observer variability , as well as shorten the overall time consuming delineation process associated with currently implemented algorithms considering the need of multiple phantom studies for the use of adaptive thresholding. 3-FLAB takes a few seconds per iteration even for the largest tumours considered in this study (on a single 2Ghz core processor in C++ implementation).Secondly, "dose painting" can be facilitated considering the non-binary nature of the proposed segmentation, allowing the automatic definition of ROIs inside the tumour, for example in dose escalation studies *(36)*, in addition to the external contour information for

optimized dosimetry, potentially reducing the dose delivered to healthy surrounding tissues and organs. The impact of such improved accuracy on overall patient's outcome remains to be demonstrated in clinical studies which are planned for the future. Finally, FLAB robustness with respect to the noise characteristics associated with the use of different scanners, acquisition protocols and reconstruction algorithms have been demonstrated in a preliminary study *(24)* and should allow its use with any type of PET images without the need of time consuming pre-processing optimization.

The proposed algorithm may also have an impact in diagnosis and therapy response assessment when combined with PVE correction (PVC) for accurate quantification. Considering different PVC approaches, anatomical information from MRI or CT is used to improve the quantitative and qualitative accuracy of functional images *(37,38)*. Unfortunately, when no anatomical image is available or no correlation exists between the anatomical and functional structures, such approaches cannot be easily used *(3)*. This is especially true in the case of large heterogeneous tumours for which there is little to no correlation between the anatomical and functional information. A potential solution will be the use of the FLAB result instead of the anatomical image in combination with one of the previously proposed PVC algorithms. This should lead to improved contrast at the object's borders as well as improved quantification in the regions within the tumour. Such combination recently demonstrated encouraging results *(39)* and warrants further investigation in terms of the potential impact in clinical therapy response studies.

## 5. Conclusion

A modified version of the FLAB algorithm has been developed in order to include the estimation of three hard classes and three fuzzy transitions. This automatic approach combines statistical and fuzzy modeling in order to address specific issues such as noise and PVE associated with 3D PET images. Its accuracy has been assessed on both simulated and clinical images of complex shapes, containing inhomogeneous activities and small regions. The results demonstrate the ability 3-FLAB to accurately delineate such lesions, for which threshold-based methodologies suggested until now fail.

Acknowledgments

-Brittany Region grant program 1202-2004

-French National Research Agency (ANR-06-CIS6-004-03,ANR-08-ETEC-005-01)

-Cancéropôle Grand Ouest (R05014NG).

References

1. N.C. Krak, R. Boellaard, et al., "Effects of ROI definition and reconstruction method on quantitative outcome and applicability in a response monitoring trial" *Eur. J. Nucl. Med. Mol. Im.,* 2005;32:294-301.

2. H. Jarritt, K. Carson, A.R. Hounsel, D. Visvikis, "The role of PET/CT scanning in radiotherapy planning", *Brit. J. Rad.,* 2006;79(S):27-35.

3. M. Soret, S. L. Bacharach, I. Buvat, "Partial-Volume Effect in PET Tumour Imaging", *J. Nucl. Med.*, 2007;48:932-945.

4. Y.E. Erdi, O. Mawlawi, S.M. Larson, et al, "Segmentation of Lung Lesion Volume by Adaptative Positron Emission Tomography Image Thresholding", *Cancer,* 1997;80(S12):2505-2509.

5. C. Greco, K. Rosenzweig, G. L. Cascini, et al, "Current status of PET/CT for tumour volume definition in radiotherapy treatment planning for non-small cell lung cancer (NSCLC), *Lung Cancer*, 2007;57(2):125-134.

6. U. Nestle, S. Kremp, A. Schaefer-Schuler, et al, "Comparison of Different Methods for Delineation of 18F-FDG PET-Positive Tissue for Target Volume Definition in Radiotherapy of Patients with Non-Small Cell Lung Cancer", *Jour. Nucl. Med.,* 2005;46(8):1342-8.

7. Q.C. Black, I.S. Grills et al., "Defining a radiotherapy target with positron emission tomography", *Int. J. Radiat. Oncol. Biol. Phys.*, 2004;60:1272-1282.

8. J.B. Davis, B. Reiner, et al., "Assessment of 18(F) PET signals for automatic target volume definition in radiotherapy treatment planning", *Radiother. Oncol.*, 2006;80:43-50.

9. J-F. Daisne, M. Sibomana, A. Bol, et al, Tri-dimensional automatic segmentation of PET volumes based on measured source-to-background ratios: influence of reconstruction algorithms, *Radioth. Oncol.*, 2003;69:247-250.

10. J. A. Van Dalen, A. L. Hoffman, et al., "A novel iterative method for lesion delineation and volumetric quantification with FDG PET", *Nuclear Medicine Communications*, 2007;28:485-493.

11. M.B. White, "A semi-automatic approach to the delineation of tumour boundaries from PET data using Level Sets", *SNM annual meeting*, 2005;314.

12. P. Tylski, G. Bonniaud, E. Decenciere, et al, "18F-FDG PET images segmentation using morphological watershed : a phantom study", *IEEE NSS-MIC*, 2006:2063-2067.

13. W. Zhu, T.Jiang, "Automation Segmentation of PET Image for Brain Tumours" *IEEE NSS-MIC,* 2003;4:2627- 2629.

14. D. W. G. Montgomery, A. Amira, and H. Zaidi, “Fully automated segmentation of oncological PET volumes using a combined multiscale and statistical model”, *Med. Phys.*, 2007;34(2):722-736.

15. O. Demirkaya, “Lesion segmentation in wholebody images of PET”, *IEEE NSS-MIC*, 2003:2873-2876.

16. X. Geets, J. A. Lee, A. Bol, et al, “A gradient-based method for segmenting FDG-PET images: methodology and validation”, *Eur. J. Nucl. Med. Mol. Im.*, 2007;34:1427-1438.

17. H. Li, W. L. Thorstad, K. J. Biehl, et al, “A novel PET tumor delineation method based on adaptive region-growing and dual-front active contours”, *Med. Phys.*, 2008;35(8):3711-3721.

18. H. Yu, C. Caldwell, K. Mah, et al, “Co-registered FDG PET/CT Based Textural Characterization of Head and Neck Cancer for Radiation Treatment Planning”, *IEEE Trans. Med. Im.*, in press, 2008.

19. M. Hatt, F. Lamare, N. Boussion, et al, “Fuzzy hidden Markov chains segmentation for volume determination and quantitation in PET”, *Phys. Med. Biol.,* 2007;52:3467-3491.

20. M. Hatt, A. Turzo, C. Roux, et al, “A fuzzy Bayesian locally adaptive segmentation approach for volume determination in PET”, *IEEE Trans. Med. Im.,* 2008;in press.

21. C.C. Ling, J. Humm, S. Larson, et al, “Towards multi-dimensional radiotherapy (MD-CRT): biological imaging and biological conformality”, *Int. J. Radiat. Oncol. Biol. Phys.*, 2000;47:551-560.

22. H. Caillol, W. Pieczynski, A. Hillon, “Estimation of Fuzzy Gaussian Mixture and Unsupervised Statistical Image Segmentation”, *IEEE Trans. Im. Proc.*, 1997;6(3):425-440.

23. F. Salzenstein, W. Pieczynski, “Parameter Estimation in hidden fuzzy Markov random fields and image segmentation”, *CVGIP,* 1997;59(4):205-220.

24. M. Hatt, A. Turzo, P. Bailly, I. Murray, C. Roux, D. Visvikis, “Automatic delineation of functional volumes in PET: a robustness study”, SNM annual meeting, 2009.

25. G. Celeux, J. Diebolt, “L’algorithme SEM : un algorithme d’apprentissage probabiliste pour la reconnaissance de mélanges de densités”, *Revue de statistique appliquée,* 1986;34(2):35-52.

26. A.P. Dempster, N.M. Laird, D.B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm”, *J. R. Stat. Soc. B.,* 1977;39:1-38.

27. J. McQueen, “Some methods for classification and analysis of multivariate observations”, *Proc. 5th Symp. Math. Stat. Prob.,* 1967;1:281-297.

28. J.C. Dunn, “A Fuzzy relative of the Isodata process and its use in detecting compact well-separeted clusters”, *J. Cybernet*., 1974;31:32-57.

29. W.P. Segars, “Development and Application of the New Dynamic NURBS-based Cardiac-Torso (NCAT) Phantom”, thesis, NC University;2001.

30. A. Le Maitre, W.P. Segars, S. Marache, A. Reilhac, M. Hatt, S. Tomei, C. Lartizien, D. Visvikis, “Incorporating patient specific variability in the simulation of realistic whole body 18F-FDG distributions for oncology applications”, *Proceedings of the IEEE*, 2009;in press.

31. F. Lamare, A. Turzo, Y. Bizais, et al, “Validation of a Monte Carlo simulation of the Philips Allegro/Gemini PET systems using GATE”, *Phys. Med. Biol.*, 2006;51:943-962.

32. A. Van Baardwijk, G. Bosmans, L. Boersma, et al, “PET-CT-based auto-contouring in non-small-cell lung cancer correlates with pathology and reduces interobserver variability in the delineation of the primary tumour and involved nodal volumes”. *Int. J. Radiat. Oncol. Biol. Phys.*, 2007;68(3):771-778.

33. T. Pan, O. Mawlawi, “PET/CT in radiation oncology”, *Med. Phys.*, 2008;35(11):4955-4966.

34. J.L. Fox, R. Rengan, W. O'meara, et al, “Does registration of PET and planning CT images decrease interobserver and intraobserver variation in delineating tumor volumes for non-small-cell lung cancer?”, *Int. J. Radiat. Oncol. Biol. Phys.*, 2005;62(1):70-75.

35. H. Ashamallaa, S. Raa, K. Parikh, et al, “The contribution of integrated PET/CT to the evolving definition of treatment volumes in radiation treatment planning in lung cancer”, *Int. J. Radiat. Oncol. Biol. Phys.*, 2005;63(4):1016-1023.

36. A. Sovik, E. Malinen, D. R. Olsen, "Strategies for biologic image-guided dose escalation: a review", *Int. J. Radiat. Oncol. Biol. Phys.*, 2009;73(3):650-658.

37. O.G. Rousset, Y. Ma, A.C. Evans, “Correction for partial volume effects in PET: principle and validation”. *J. Nucl. Med.*, 1998;39:904-911.

38. N. Boussion, M. Hatt, F. Lamare, et al, “A multiresolution image based approach for correction of partial volume effects in emission tomography”, *Phys. Med. Biol.*, 2006;51:1857-1876.

39. N. Boussion, M. Hatt, D. Visvikis, “Partial volume correction in PET based on functional volumes”, *J. Nucl. Med.*, 2008;49(S1):388.

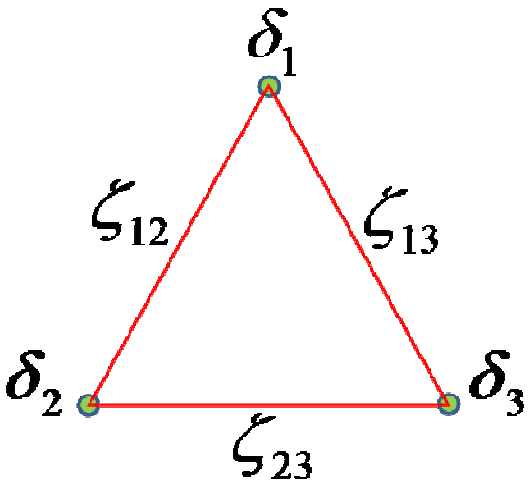
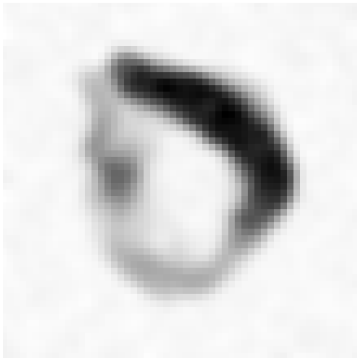



Figure 1



(a)

(b)

Figure 2

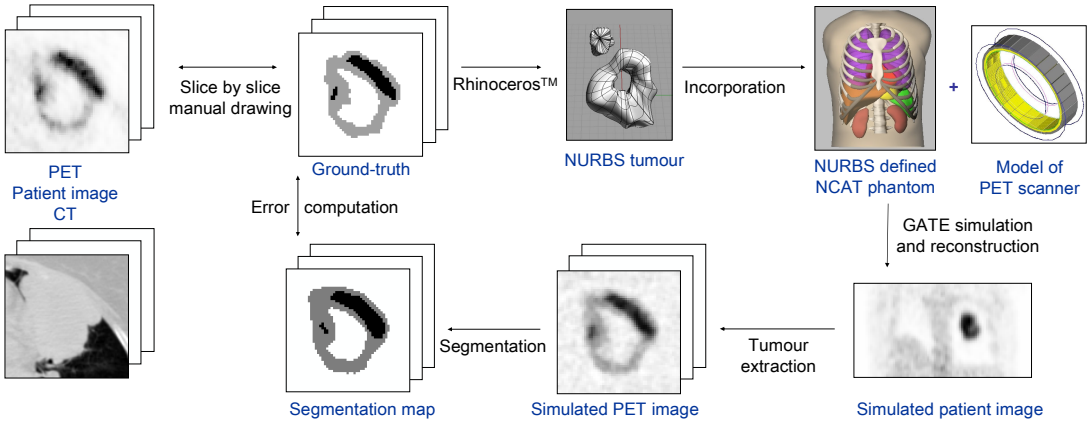

PET
Patient image
CT
Slice by slice
manual drawing
Ground-truth
Rhinoceros™
NURBS tumour
Incorporation
NURBS defined
NCAT phantom
+
Model of
PET scanner
GATE simulation
and reconstruction
Simulated patient image
Tumour
extraction
Simulated PET image
Segmentation
Segmentation map
Error
computation

Figure 3

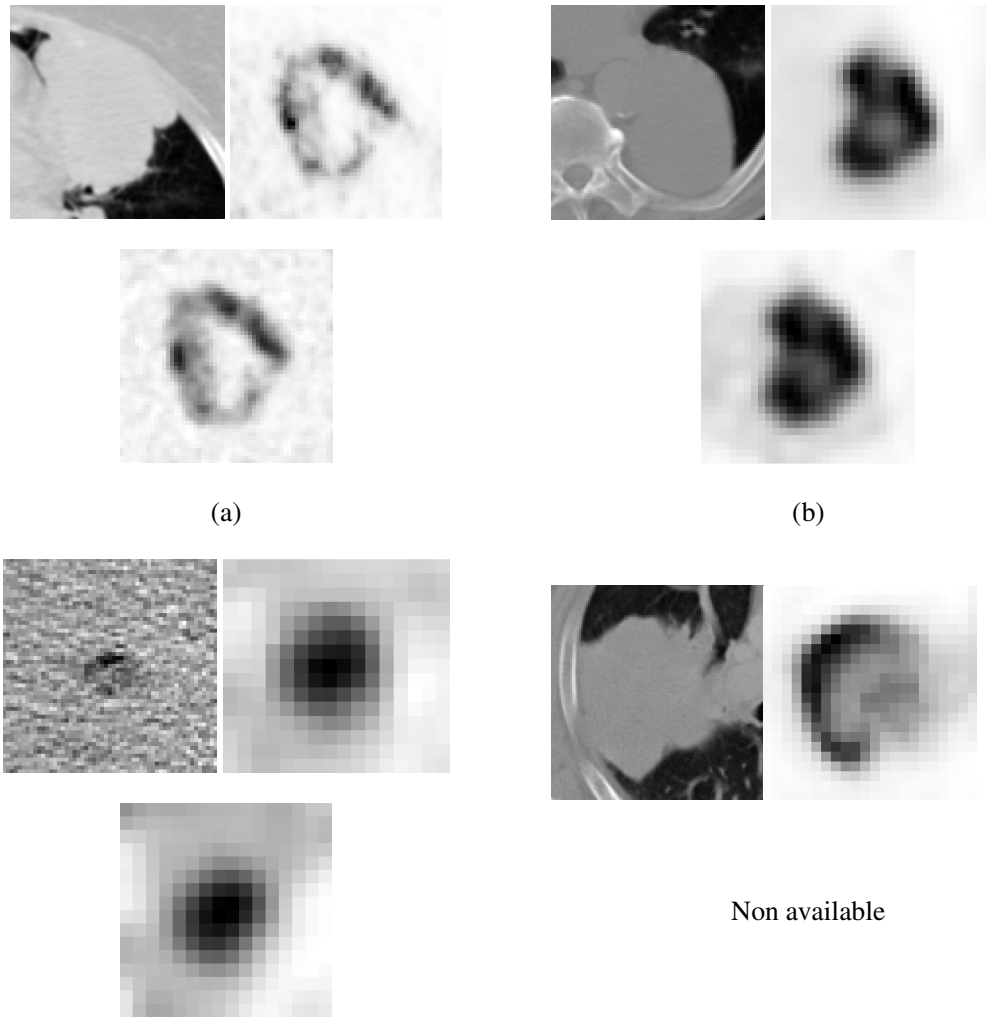
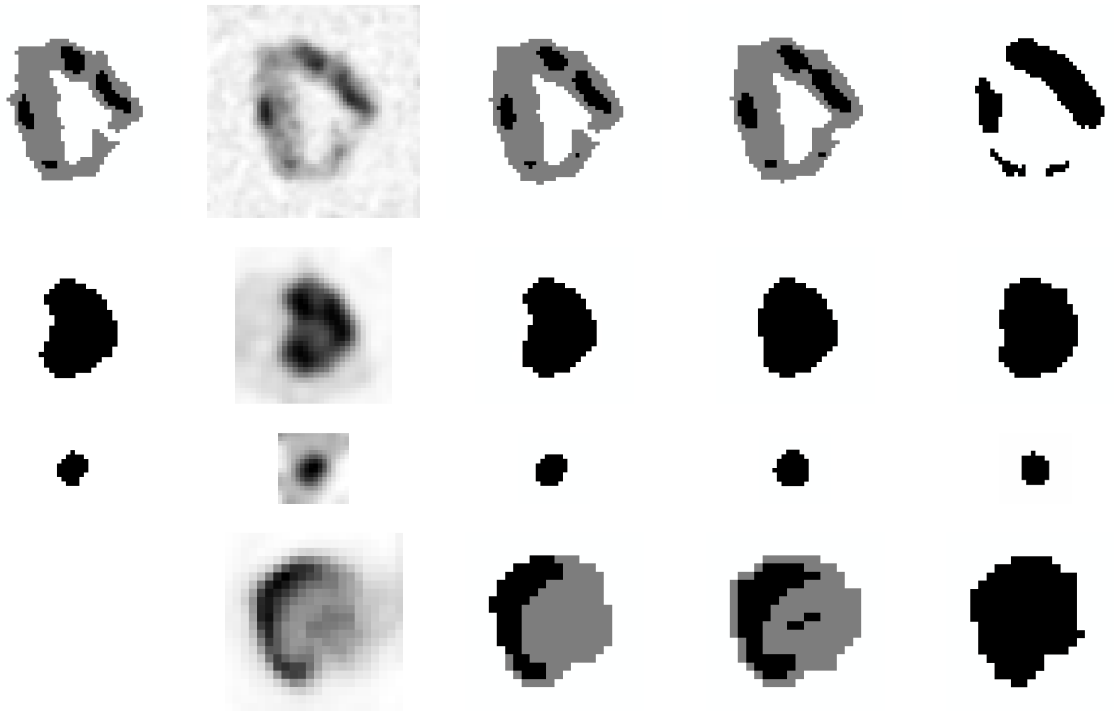
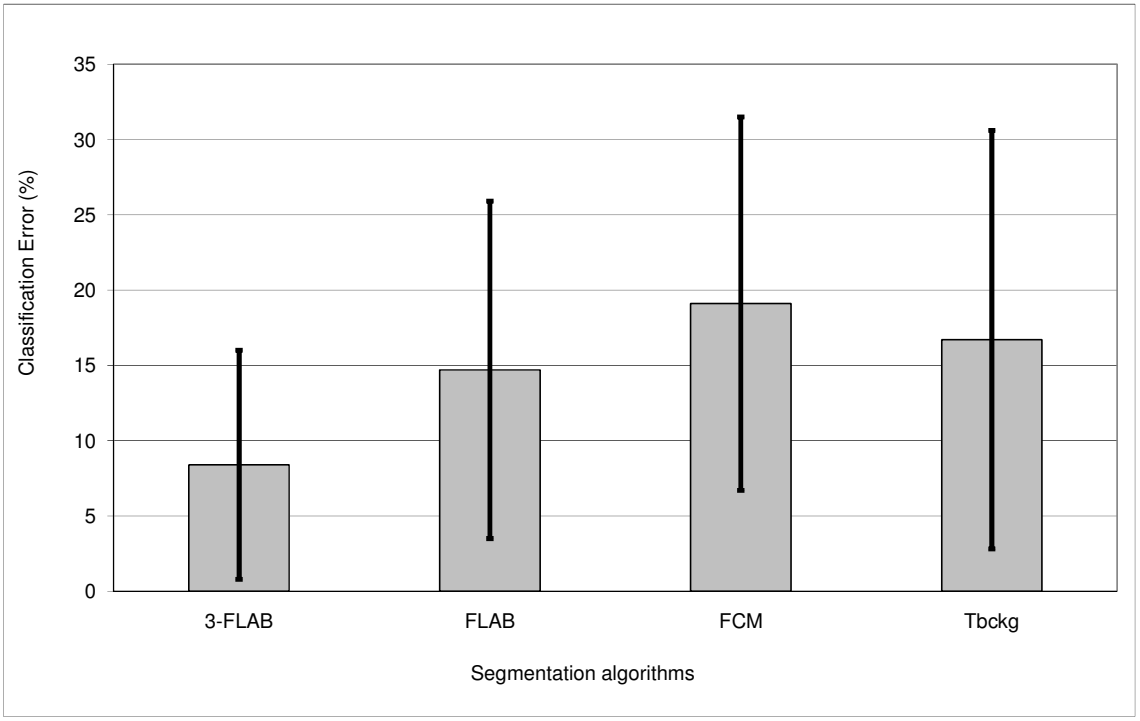



Figure 4

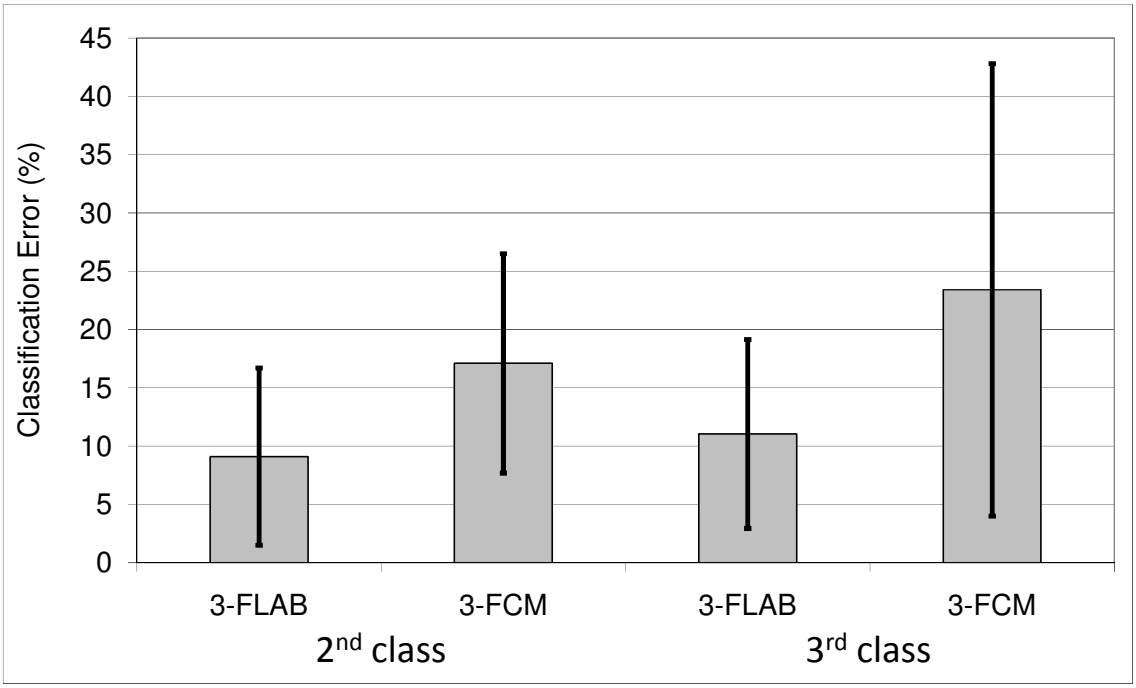


(a) (b) (c) (d) (e)

Figure 5


35
30
25
20
15
10
5
0
Classification Error (%)
3-FLAB
FLAB
FCM
Tbckg
Segmentation algorithms

(a)


45
40
35
30
25
20
15
10
5
0
Classification Error (%)
3-FLAB
3-FCM
3-FLAB
3-FCM
2nd class
3rd class

(b)

Figure 6

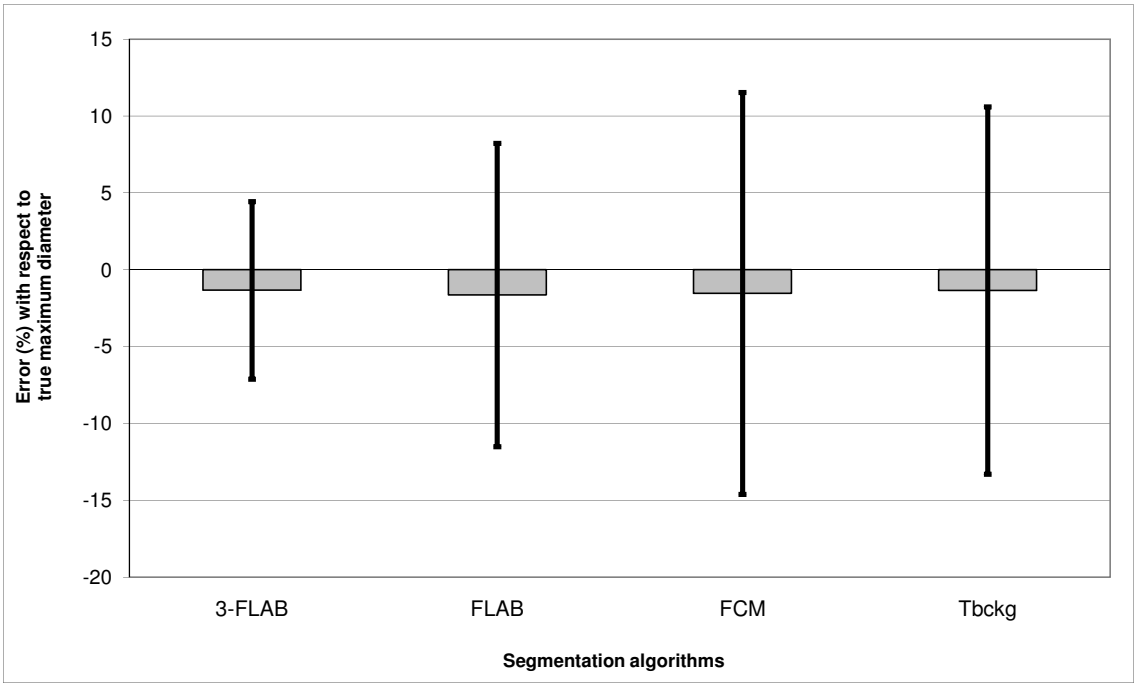

15
10
5
0
-5
-10
-15
-20
Error (%) with respect to true maximum diameter
3-FLAB
FLAB
FCM
Tbckg
Segmentation algorithms

Figure 7

**Figure Captions**

**Figure 1**: The fuzzy scheme in the 3-FLAB implementation.

**Figure 2:** Binary FLAB model applied to a heterogeneous simulated tumour (a). The segmentation result (b) clearly misses parts of the tumour.

**Figure 3**: The simulation of realistic PET images.

**Figure 4**: Datasets illustration. (a-d) Examples of clinical tumours (up) with CT (left) and PET (right), and the corresponding simulated PET (down). (a)-(c) dataset-1; (d) dataset-2.

**Figure 5:** Segmentations of the tumours in fig.4(a-d): (a) ground-truth; (b) PET image; (c-d) segmentations for (c) 3-FLAB, (d) FCM, (e) $T_{bckg}$.

**Figure 6**: Mean Classification Errors and standard deviation for (a) all methodologies considering all twenty tumours of dataset-1, (b) 3-FLAB and 3-FCM considering $2^{nd}$ and $3^{rd}$ classes of the ten heterogeneous tumours of dataset-1.

**Figure 7**: Mean errors and standard deviation for each methodology, with respect to known maximum diameter of dataset-2 tumours.