A framework based on Hidden Markov Trees for multimodal PET/CT image co-segmentation

Houda Hanzouli-Ben, Jérôme Lapuyade-Lahorgue, Julien Bert, Didier Benoit, Philippe Lambin, Angela van Baardwijk, Emmanuel Monfrini, Wojciech Pieczynski, Dimitris Visvikis, Mathieu Hatt

▶ **To cite this version:**

Houda Hanzouli-Ben, Jérôme Lapuyade-Lahorgue, Julien Bert, Didier Benoit, Philippe Lambin, et al.. A framework based on Hidden Markov Trees for multimodal PET/CT image co-segmentation. Medical Physics, 2017. hal-01577785

HAL Id: hal-01577785
https://hal.science/hal-01577785v1

Submitted on 28 Aug 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A framework based on Hidden Markov Trees for multimodal PET/CT image co-segmentation

Houda HANZOULI-BEN SALAH[1], Jerome LAPUYADE-LAHORGUE[1], Julien BERT[1], Didier BENOIT[1], Philippe LAMBIN[2], Angela VAN BAARDWIJK[2], Emmanuel MONFRINI[3], Wojciech PIECZYNSKI[3], Dimitris VISVIKIS[1], Mathieu HATT[1]

[1] INSERM, UMR 1101, LaTIM, IBSAM, UBO, UBL, Brest, France

[2] Department of Radiation Oncology (MAASTRO), GROW - School for Oncology and Developmental Biology, Maastricht University Medical Center, Maastricht, The Netherlands

[3] SAMOVAR, Télécom SudParis, CNRS, Université Paris-Saclay, 9 rue Charles Fourier, 91000 Evry, France

Corresponding author:

Mathieu Hatt

INSERM, UMR 1101, LaTIM

IBRBS – faculté de médecine

22 rue Camille Desmoulins

29238 Brest

France

Tel: +33 2 98 01 81 11

Fax: +33 2 98 01 81 24

Email: hatt@univ-brest.fr

Wordcount: ~9800

Abstract

Purpose: To investigate the use of a probabilistic quad-tree graph (Hidden Markov Tree, HMT) to provide fast computation, robustness and an interpretational framework for multimodality image processing and to evaluate this framework for single gross tumor target (GTV) delineation from both positron emission tomography (PET) and computed tomography (CT) images.

Methods: We exploited joint statistical dependencies between hidden states to handle the data stack using multi-observation, multi-resolution of HMT and Bayesian inference. This framework was applied to segmentation of lung tumors in PET/CT datasets taking into consideration simultaneously the CT and the PET image information. PET and CT images were considered using either the original voxels intensities, or after wavelet/contourlet enhancement. The Dice similarity coefficient (DSC), sensitivity (SE), positive predictive value (PPV) were used to assess the performance of the proposed approach on one simulated and 15 clinical PET/CT datasets of non-small cell lung cancer (NSCLC) cases. The surrogate of truth was a statistical consensus (obtained with the Simultaneous Truth And Performance Level Estimation algorithm) of three manual delineations performed by experts on fused PET/CT images. The proposed framework was applied to PET-only, CT-only and PET/CT datasets, and were compared to standard and improved fuzzy c-means (FCM) multimodal implementations.

Results: A high agreement with the consensus of manual delineations was observed when using both PET and CT images. Contourlet-based HMT led to the best results with a DSC of 0.92±0.11 compared to 0.89±0.13 and 0.90±0.12 for Intensity-based HMT and Wavelet-based HMT respectively. Considering PET or CT only in the HMT led to much lower accuracy. Standard and improved FCM led to comparatively lower accuracy than HMT, even when considering multimodal implementations.

Conclusions: We evaluated the accuracy of the proposed HMT-based framework for PET/CT image segmentation. The proposed method reached good accuracy, especially with pre-processing in the contourlet domain.

Keywords: Positron emission tomography (PET), computed tomography (CT), Bayesian inference, hidden Markov trees (HMT), wavelet and contourlet analysis, segmentation

INTRODUCTION

The current trend in modern personalized medicine is to acquire and exploit sequential images for a patient during the course of its treatment (pre-, per-, post-treatment). These images are also more and more frequently coming from multiple image modalities including Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), Single Photon Emission Computed Tomography (SPECT) or UltraSound (US). In addition, the development of each modality has led to the use of several modes of acquisition, such as various MRI sequences or the development of several PET and SPECT radiotracers. In clinical oncology and radiotherapy, this has been especially true, due to the introduction of multimodality imaging integrated devices such as PET/CT and SPECT/CT, and now the rising development of sequential and simultaneous PET/MRI, which have also led to significant improvements regarding image fusion and anatomical/physiological data association.

On the one hand, multimodality imaging today provides the physicians with an unprecedented potential wealth of information, both morphological and functional regarding the pathology [1]. On the other hand, the manual and visual exploitation of the image datasets becomes more and more complex, tedious, subjective and time-consuming owing to increasing data volumes as well as image characteristics' variability. Automated analysis may thus potentially improve overall patient management based on the use of multimodality images.

From an image and data processing point of view, most recent developments have been focused on a single image modality, optimizing and adapting algorithms to deal with its specific characteristics. In clinical research, when multiple images (from one or several modalities) are being considered, they are most often analyzed independently and the results are then combined *a posteriori*, thereby not exploiting the full potential correlation between multimodal images. One of the main challenges for semi-automatically handling such multimodal datasets is the large variability in terms of spatial resolution and definition, noise and texture properties, across as well as within modalities.

Several papers have investigated the combination and association of functional and anatomical images for a specific purpose, such as denoising[2] or reconstruction[3] of PET data using anatomical priors coming from MRI or CT, correction of partial volume effects in PET or SPECT by exploiting the associated morphological information [4–6], the use of associated CT images to spatially register several low-resolution PET images, for instance during treatment [7, 8] or

several different radiotracers [9], or the definition of tumor target volumes in radiotherapy by considering both PET and CT images features [10, 11].

This last application in particular has led recently to the development of several (semi)automated methodologies aiming at determining a single gross tumor volume (GTV) from fused PET/CT datasets, assuming both modalities provide complementary information regarding a single contour of the tumor volume [10–16]. Although an alternative approach could consist in segmenting each image independently and then combine the obtained volumes *a posteriori* using some kind of consensus, these approaches have mostly considered the first approach and used a variety of tools including supervised learning using textural features of PET/CT images [10, 14], Markov random field [13], random walk segmentation [12], active contours [11, 16] and topology integrated within graph-models [15]. The idea underlying all these studies is that the delineation should benefit from the co-registered modalities' information in order to produce a single, more robust and accurate GTV. Most of these previous works have indeed highlighted the potential value in considering both modalities simultaneously in a common mathematical and algorithmic framework. The objective of most of these studies is to objectively automate the manual and subjective task of GTV delineation by a radiation oncologist visualizing the fused PET/CT images. We place our present work in the same context.

Hidden Markov fields (HMF) are well suited to image processing and have been exploited in a variety of applications, including medical imaging [17]. They have been used in PET segmentation [18] and two recent studies investigated the use of HMF to associate PET and CT datasets with the goal of co-segmentation [13, 19]. Markov random fields however require computationally expensive iterative methods for estimating parameters[17], contrary to hidden Markov chains (HMC) that have also been explored for PET segmentation [20]. Finally, hidden Markov trees (HMT) share similar favorable computational properties with chains, while being quite as robust as fields and providing a useful structure for the association of multiscale and multimodal images [17, 21]. HMT proved their efficiency in several fields. For example, phylogenetic models in biology are diagrams showing the evolutionary relationships between species [22]. Hierarchical models are also used for aerial image parsing [23]. HMT have found numerous applications in imaging, including astronomical [24], remote sensing [25, 26] and medical imaging for instance in MRI [27]. However, medical image segmentation using HMT is quite rare [28] contrary to Markov fields [29], and to the best of our knowledge, our work is the first to investigate their use for multimodal PET/CT image segmentation. The present paper is an extension of the first developments initially presented in a conference proceeding [30].

The scope of this work is thus to evaluate the value of a unified framework based on HMT for PET/CT segmentation with the objective of GTV delineation in radiotherapy planning. Results are provided on 1 simulated case and on 15 clinical datasets with comparison to a statistical consensus of three different manual delineations. Finally, we elaborate at several potential improvements regarding the framework itself, as well as other applications that may benefit from its use, and that will be investigated in future studies.

MATERIALS AND METHODS

1. Hidden Markov Tree framework

In this section, we present the mathematical methodology used to handle, associate and process multimodal medical images using different resolutions and modalities within a unified framework based on a HMT.

The multiresolution nature of the data can be managed by using the successive scales of the HMT in order to handle data represented at different resolutions. Another modeling approach consists in using the different levels in the HMT not to represent data of different resolutions, but rather as an algorithmic solution to model dependencies between neighboring observations at successive scales, while the data is associated only with the last scale of the HMT. In that case, different modalities are handled by considering observed data in a given voxel as a vector containing several values, each vector component representing a modality's observation. This setting was chosen for the present work: in radiotherapy planning clinical practice, the PET image is up-sampled to exactly the same dimension as the CT, so that the images can be overlaid for visualization and manual delineation. We used cubic B-spline interpolation to up-sample the PET image so that a PET voxel exactly corresponds to a CT voxel. Therefore both PET and CT voxels intensities are associated as a 2-value vector with the leaves of the tree (the level/scale at the bottom of the tree), as illustrated in figure 1. Note that all the scales/levels above in the tree are only used as calculus and parameters estimation tools and none of their nodes are associated with PET/CT voxels values.

The HMT will exploit conjointly within the same algorithmic framework each couple of voxels' intensities (one PET, one CT) to take a segmentation decision based on the probability of this couple to belong to a class (tumor) or another (background). This probability is calculated by combining two different probabilities: the first corresponds to the prior

probability of each node of the tree to belong to a given class, which will be based on the fathers-children statistical transitions in the tree structure (section 1.1). Regarding the segmentation task, the most important probability will be the one associated with each leaf. Indeed, the probabilities estimated in the above scales are only calculus and estimation tools to incorporate spatial neighborhood and multiscale information. The second corresponds to the observation probability calculated based on the voxel couple of values (PET and CT) with respect to the respective joint generalized normal distribution estimated for the entire PET/CT region of interest containing the tumor and background (section 1.2). Each case requires automatic estimation of all parameters defining the HMT model and these estimated parameters are then used to perform the actual segmentation (section 1.3).

1.1 Hidden Markov Tree model

Contrary to HMC, a HMT is a hierarchical hidden Markov model, which is able to take into account the probabilistic dependencies between scales directly. At a given scale, the probabilistic spatial dependencies among voxels are then taken into account via dependency on a common ancestor (Figure 1). Similarly to HMCs, HMTs allow fast data processing owing to their computational efficiency. Despite their simplicity, they are able to produce results with similar performance as hidden Markov fields [17].

The HMT model is defined as follows: let S be a finite set of points and $X = (X_s)_{s\in S}$, $Y = (Y_s)_{s\in S_0}$ two random vectors indexed respectively on S and $S_0 \subset S$ subset of observed voxels. Each X_s takes its values in the finite set of classes $\Omega=\{\omega_{1,\dots,}\omega_k\}$ (in this work tumor and background) and Y_s takes its values in the set of observations (real values corresponding to voxels' intensities). Let S^1, ..., S^R be a partition of S representing different "generations". Each $s\epsilon S^i$ admits $s^+ \subset S^{i+1}$ (called his "children") in such a way that every element of $t\ \epsilon S^{i+1}$ has a unique "father" $t^-\epsilon S^i$ such that $(s^+)^- = s$. We assume that S^1 is a singleton so as its element r is called "root". Setting $p(x,y)$ density of the distribution of (X, Y), the random vector (X, Y) is a HMT if:

$$p(x,y) = p(x_r)p(y_r|x_r)\prod_{j=2}^{N}\prod_{s\epsilon S^j} p(x_s|x_{s^-})\, p(y_s|x_s), \quad (1)$$

where $p(y_s|x_s) = 1$ if $s \notin S_0$

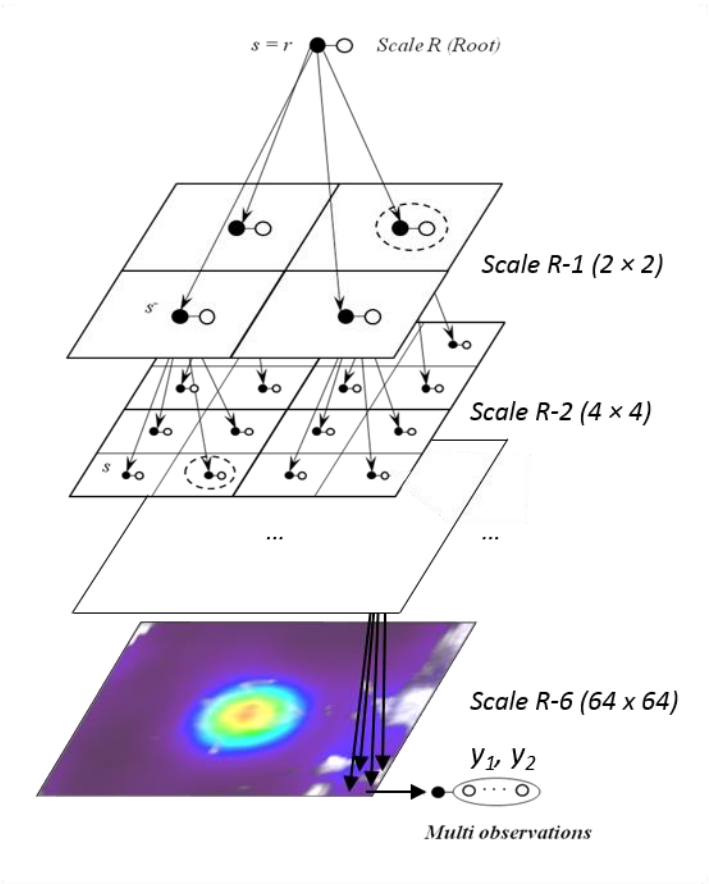


Figure 1: Hidden Markov Tree (HMT) Model with the associated PET/CT image (selection centered on the tumor): the root, scale R-1 with 4 children, scale 2 with 16, down to scale R-6 with 4096 "leaves" (64×64), each one being associated with a vector of 2 values (y_1 and y_2) corresponding to the PET and CT intensities. The classification as "tumor" or "background" is $x = (x_s)_{s \in S_0}$.

1.2 Observation modeling

With respect to image segmentation in a statistical setting, a Gaussian distribution is sufficient to model PET voxels' intensities and more complex distributions through the use of the Pearson's system (containing 7 different distributions)[31] does not improve accuracy [32]. However, in the present case the framework has to conjointly model PET and CT distributions that could be very different, hence calling for a more flexible statistical model. Therefore the observation's noise densities $p(y_s|x_s)$ (for $s \in S_0$) considered are of the generalized normal distribution form [33].

1.3 HMT parameters estimation

The parameters of the HMT that need to be estimated are listed in table 1. The table also lists how their values are initialized, as well as the aspects of these parameters that are not estimated but rather set *a priori*.

HMT parameter category	Parameter	Initialization	Parameter aspects that are not estimated but set *a priori*
Prior model	Root prior distribution	Equal probability for each class	Each node has 4 children. The number of scales and nodes/leaves are set depending on the size of the input image (*e.g.* 6 scales for a 64×64 image).
	Parent-child transition probabilities	Calculated using a first rough segmentation through a Fuzzy C-means algorithm	
Observation model (data likelihood)	Mean of distribution for each class		The distribution of each class is a generalized normal distribution.
	Standard deviation of distribution for each class		
	Shape parameter (generalized normal distribution) for each class		

Table 1: parameters of the HMT

We emphasize that parameters of the HMT are not set or chosen arbitrarily or empirically but all of them are automatically estimated from the observed PET, or CT, or the PET/CT intensities. To estimate the parameters we chose the iterative Stochastic Expectation Maximization (SEM) algorithm [34] which is a stochastic version of the EM algorithm [34]. The EM algorithm has been adapted to HMT [25], and in the present work we used a SEM version adapted to the HMT model. The parameters are first initialized using a simple fuzzy c-means. The loglikelihood and conditional probabilities involved in the (S)EM rely on the use of backward and forward probabilities, in a similar manner as in HMC. Segmentation (assigning a class to the hidden states based on the tree structure probabilities and the observed data) is then obtained using the estimated parameters with the Modes of Posterior Marginal (MPM) [35] inference, a Bayesian estimator which associates to each site the most probable class given all the data. This estimator requires the computation of the posterior marginals. These are computed through the Baum-Welsh algorithm [36] by propagating information first from the "leaves" to the "root", and then in the inverse direction, from the "root" to the "leaves". In this work we used the conditional version of Baum-Welsh algorithm [37] to avoid underflow issues. More details can be found in [25].

2. Pre-processing

The HMT model was applied to three different data representations. Observations in each node were defined as the intensity in each image voxel (as illustrated in figure 1) or its corresponding element in image obtained after pre-processing in the wavelet or contourlet domains.

The spatio-frequential (wavelet or contourlet) transforms provide the ability to exploit image information at different resolutions and directions, in order to better quantitatively describe the information provided by the observed data. The

use of images rebuilt after pre-processing in the spatio-frequential domain has been suggested recently for PET image segmentation [38]. We thus decided to investigate the potential benefit of such an approach within our context of PET/CT segmentation, considering the contourlets in addition to the wavelets.

2.1 Wavelets pre-processing

The wavelet transform (WT) is established as a powerful tool for statistical signal and image processing. In this work, we used the isotropic undecimated wavelet transform (IUWT) based on the 'à trous' algorithm [39]. The non decimation avoids the pseudo Gibbs phenomenon. The "à trous" algorithm decomposes an image $I_0(x,y,z)$ into a list of wavelet planes $\{w_j(x,y,z)\}$ with intermediate smoothed planes $\{F_j(x,y,z)\}$. This approximation sequence $\{F_j(x,y,z)\}$ is obtained through N successive filtering with a 3D low-pass filter h [5] related to a scaling function φ_{3D} which must satisfy a number of properties (compactly supported, regularity, symmetry, convergence in the Fourier space and derivability). B-spline interpolations were chosen in this work. At each iteration j, the approximation image I_{j+1} is given according to equation (2).

$$I_{j+1}(x,y,z) = \sum_{m,n,o \in [-2;2]} h(m,n,o) I_j(x + m2^j, y + n2^j, z + o2^j) \qquad (2)$$

The wavelet planes w_{j+1} are defined as the differences between two consecutive approximations I_j and I_{j+1} containing the details at a resolution level between I_j and I_{j+1} in such a way that the initial image $I_0(x,y)$ can be rebuilt from its details layers as in equation (3).

$$I_0(x,y,z) = I_N(x,y,z) + \sum_{j=1}^{j=N} w_j(x,y,z) \qquad (3)$$

2.2 Contourlets pre-processing

The Contourlet transform (CoT, to avoid confusion with CT, computed tomography) [40] is an extension of the WT using non-separable and directional filter banks. Thanks to its remarkable properties (multiresolution, localization, directionality and anisotropy), CoT is a more efficient tool than WT for capturing intrinsic geometrical structures of images. CoT is implemented by the pyramidal directional filter bank (PDFB). In a first step, a Laplacian pyramidal (LP) decomposition is applied to the image so as to obtain a number of radial subbands. In the second step, each LP subband is decomposed into a power of two's number of directional subbands through a filter banks. In this work, we used the non-subsampled contourlet transform (NSCT) [41]. First, the M-most significant coefficients from all subbands

are retained whereas the rest are set to 0. Second, the image is rebuilt using the selected coefficients. PET and CT images and their wavelet or contourlet-based pre-processing results are illustrated in figure 2. In a previous work[38], the value M was arbitrarily set at 200 for PET. In our case we determined these values empirically by exploring values between 100 and 3000. The best performance was reached for M=800 for PET images and M=2000 for CT images.

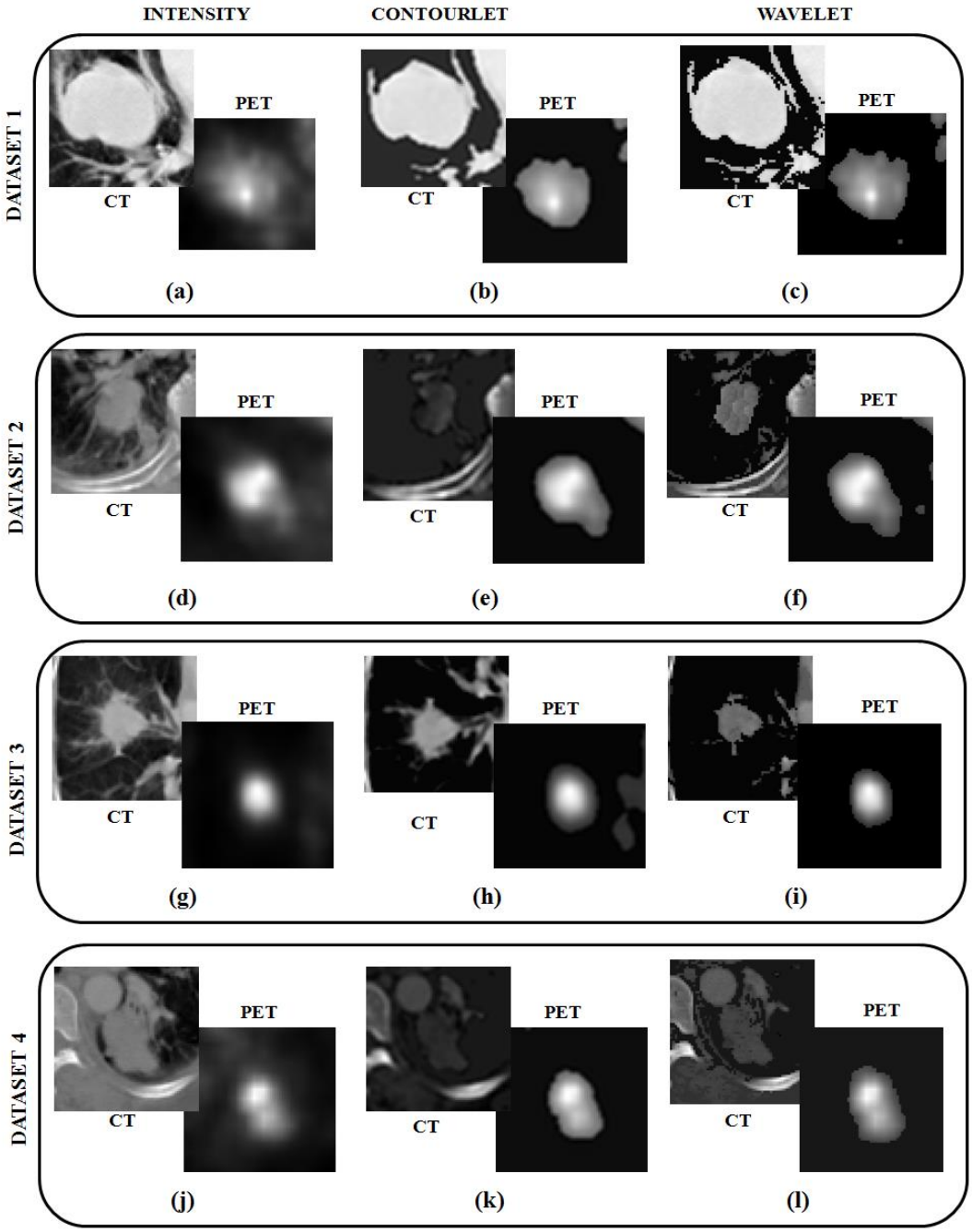
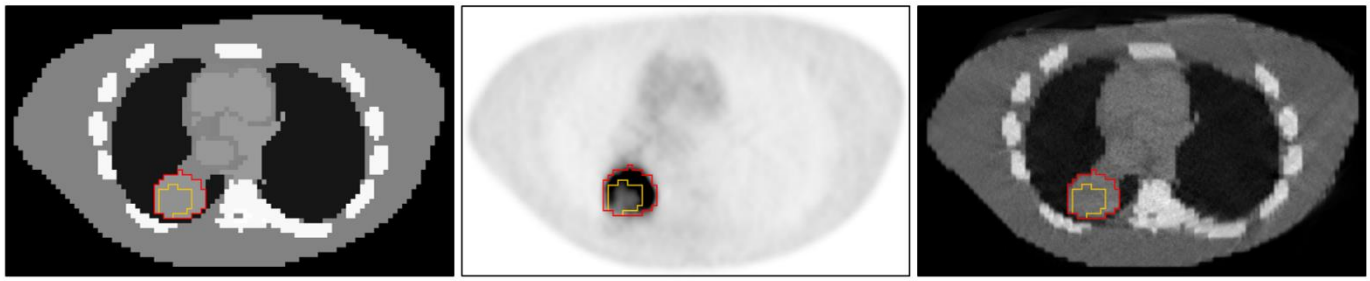


Figure 2: Four examples of PET/CT clinical datasets (a, d, g, j) preprocessed in the contourlets (b, e, h, k) or wavelets (c, f, i, l) domains. Note the increased contrast and loss of background details.

3. Evaluation study

3.1 Simulated dataset

We have been advocating for a long time the use of realistic simulated datasets with voxel-based ground-truth as part of a rigorous validation framework for PET segmentation [42, 43] and have contributed in making such datasets available to the community [44, 45]. However these past efforts were focused on the PET component and did not include CT simulated images. For the present work in order to investigate the complementary value of PET/CT *vs.* CT-only or PET-only HMT segmentation, a digital phantom of a lung tumor located close to the mediastinum and the chest wall was designed and both a PET image and CT image were simulated. No breathing was simulated. The simulated PET was generated with GATE (Geant4 Application for Tomography Emission) and reconstructed following the same process as described previously [44, 45]. PET spatial resolution was ~4mm and voxel size of the reconstructed PET image was 4×4×4 mm^3. The CT simulated image was generated using the same digital phantom used for the PET simulation. The CT simulation is based on a Monte-Carlo simulation where X-rays are created and tracked individually through the voxelized phantom using the Siddon projector [46] and attenuated following the Beer-Lambert Law. Compton and Rayleigh scattering were not computed. The simulation was performed using 360 projections around the phantom (1 projection per degree), and 5×10^9 particles were created to achieve ~10000 counts/voxel in the histogram. The source aperture was 20 degrees, a focal spot size of 0 mm was simulated, and a mono-energetic at 60 keV was simulated to avoid beam-hardening artifacts in the final reconstructed image. A line detector composed of 750 pixel elements and 1 mm pixel size, was placed behind the phantom. The phantom was centered in the image space, the position of the X-ray source was (-700, 0, 0) mm and the position of the simulated detector was (350, 0, 0) mm. Finally, the simulated histogram was reconstructed using a maximum likelihood for transmission tomography (MLTR) algorithm based on a gradient-ascent method [47]. For this study, the CT data was reconstructed using 5 iterations and 72 subsets, 1 mm voxel size, and the final image was blurred with 1 mm Gaussian filter. As explained in section 1, the reconstructed PET image was up-sampled to be of same size as the CT dataset to be associated within the HMT framework. Figure 3 shows the digital phantom and the associated simulated PET and CT images.



(a) (b) (c)

Figure 3. (a) Digital phantom with complete tumor ground-truth contour in red and sub-volume in orange, with a mean value of -5HU and 3 SUV. The rest of the tumor has a mean value of 20 HU and 12 SUV. Background uptake in the lung was set at 0.8 SUV whereas HU were set at -850. Bones were set at 700 HU. The uptake in the surrounding soft tissues (mediastinum) was set at 1.6 SUV and the HU at 40 and -80 (light and dark grey voxels respectively); (b-c) corresponding reconstructed and upsampled PET (b) and CT (c) images.

3.2 Clinical Datasets

The proposed methodology was evaluated on PET/CT clinical images. Fifteen patients with confirmed non–small cell lung cancer (NSCLC), stage Ib–IIIb, for which no obvious spatial mismatch between the CT and the PET images (due to respiratory motion and/or positioning differences between the two acquisitions) was observed, were analyzed. All patients underwent an ^{18}F-FDG PET/CT examination for staging purposes before treatment. Patients were instructed to fast for a minimum of 6 h before examination. Free-breathing PET and CT images were acquired 45–60 min after ^{18}F-FDG injection. A total of seven 5-min bed positions with overlap were used for whole-body PET (Biograph PET/CT; Siemens) acquisitions, which were corrected for attenuation using the CT data and iteratively reconstructed using the ordered-subsets expectation maximization algorithm (4 iterations, 8 subsets). The noise levels in PET and CT correspond to standard PET/CT clinical acquisitions. Voxel sizes of PET and CT datasets were $5.31\times5.31\times5$ mm^3 and $0.98\times0.98\times5$ mm^3 respectively. Spatial resolution of the PET scanner is estimated at 4.5 mm full-width at half-maximum (FWHM) in the center of the field of view, whereas spatial resolution of the CT scanner is estimated below 1 mm FWHM. The PET and CT images were assumed to be co-registered at reconstruction. The PET images were up-sampled using a cubic B-spline interpolation scheme [48] so that it corresponds to the dimension of the CT image.

3.3 Surrogate of truth

In the simulated case, the ground-truth of the phantom was used. In the clinical dataset, all patients underwent surgery and the maximum diameter was measured on the surgical specimen, as previously described [49]. In the absence of other surrogate of truth such as histopathological volumes, the validation of the approach consisted in comparing automatic segmentation of the tumor with a statistical consensus of manual delineations performed by three different experts on each slice of the fused PET/CT images. The experts were asked to draw contours with the goal of defining a GTV for radiotherapy planning. The consensus was obtained using the simultaneous truth and performance level estimation (STAPLE) algorithm [50], an expectation-maximization algorithm that computes a probabilistic estimate of the true

segmentation from a collection of segmentations. In order to strengthen the reliability of this consensus, the maximum diameter measured in histopathology was compared with the one measured on the STAPLE consensus. The maximum diameter was not further used for the validation of the proposed algorithm since we are interested here in spatial PET/CT overlap, which cannot be rigorously evaluated with the maximum diameter only.

3.4 PET, CT and PET/CT comparison

The proposed framework was first evaluated on the simulated PET/CT dataset, and then on the 15 clinical cases, by considering either the PET or the CT image only, then using both PET and CT images. Parameters of the HMT model as described in table 1 were in each case automatically estimated from the input images. They are therefore not set with the same values for each configuration (PET, CT and PET/CT) but iteratively estimated for each image and each configuration. Note however that the structure of the tree (1 father, 4 sons for instance) and type of statistical distributions (set as generalized normal distributions) are the same in all configurations since they are not estimated.

Note also that the segmentation algorithms are not applied to the entire whole-body datasets, but rather on a selected cropped volume around the tumor (as illustrated in figures 2 and 5 for instance).

3.5 Other methods for comparison

Segmentation results were also generated using the standard fuzzy c-means (FCM) algorithm and the fuzzy local information c-means (FLICM) [51] for comparison purposes. FLICM is an FCM implementation adding a weighted norm to account for outliers due to the noise and uses two parameters: a regularization parameter and the size of the surrounding kernel. We used the recommended parameters values (regularization equal to 1 and kernel radius equal to 3 voxels) [51]. Alternative values for these two parameters were explored without obtaining improved performances. Both methods were implemented as multimodal, *i.e.* using as input vectors containing the intensities of both up-sampled PET and CT images. The input regions (cropped from the entire image) were the same for all methods under comparison.

3.6 Evaluation metrics and figures of merit

The performance of segmentation was assessed using Dice similarity coefficient (DSC), sensitivity (SE) and positive predictive value (PPV). Let us denote by V_T the true lesion volume, V_S the segmented volume, $V_T \cap V_S$ the intersection between volumes V_T and V_S and size (V) the size of volume V.

The sensitivity (SE) is defined as the ratio between the size of the segmented volume intersecting the true volume, and the true volume:

$$SE = \frac{size(V_T \cap V_S)}{size(V_T)} \qquad (4)$$

The positive predictive value (PPV) is defined as the ratio between the size of the segmented volume intersecting the true volume and the measured size of the segmented volume:

$$PPV = \frac{size(V_T \cap V_S)}{size(V_S)} \qquad (5)$$

The combination of a high PPV and a low SE indicates an under-evaluation of the true volume, whereas the inverse indicates an over-evaluation of the true volume.

The Dice similarity coefficient (DSC) is a statistical measure of accuracy in image segmentation. This metric assesses the spatial overlap between the segmented volume (VS) and the true one (VT) and is defined as the ratio between twice the size of the segmented volume intersecting the true volume and the size of the sum of V_S and V_T.

$$DSC = \frac{2 \times size(V_T \cap V_S)}{size(V_T + V_S)} \qquad (6)$$

The Kruskal–Wallis rank test was used (Medcalc™, Medcalc software, Belgium) to statistically compare the performance of each methods.

RESULTS

Validity of the manual consensus

The consensus of three manual delineations led to volumes with a maximum diameter very close to the histopathology measurement, with a correlation of 0.91, and differences of 9±7%, which is in line with previous observations on the

same datasets by the original authors [49] and other studies [55]. This consensus of manual delineations can therefore be considered a reliable surrogate of truth for evaluating the proposed methodology.

HMT segmentation results

Simulated dataset

Figures 4 and 5 show the IHMT segmentation results considering only the PET, only the CT, or the joint PET/CT. The PET-only segmentation led to very high PPV (0.96) but low SE (0.65) (DSC of 0.77), due to the heterogeneous uptake of the tumor, although with no issue differentiating the rest of the tumor uptake from the surrounding normal tissues uptake. On the contrary the CT-only segmentation led to very high SE (0.98) but low PPV (0.64) (DSC of 0.77) due to the lack of contrast between tumor and surrounding normal soft tissues. The PET/CT combination led to a better coverage of the tumor with high SE (0.90) and PPV (0.97) (DSC of 0.93).

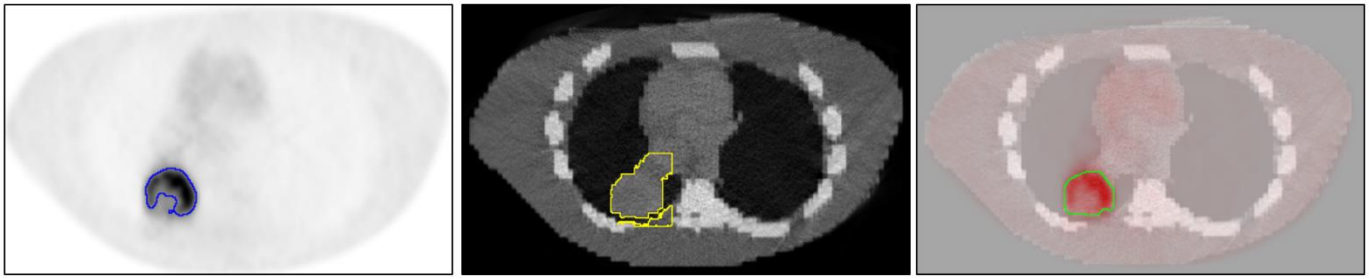


Figure 4. Simulated (a) PET, (b) CT and (c) overlaid PET/CT images with HMT segmentation on (a) PET only in blue, (b) CT only in yellow and PET/CT in green.

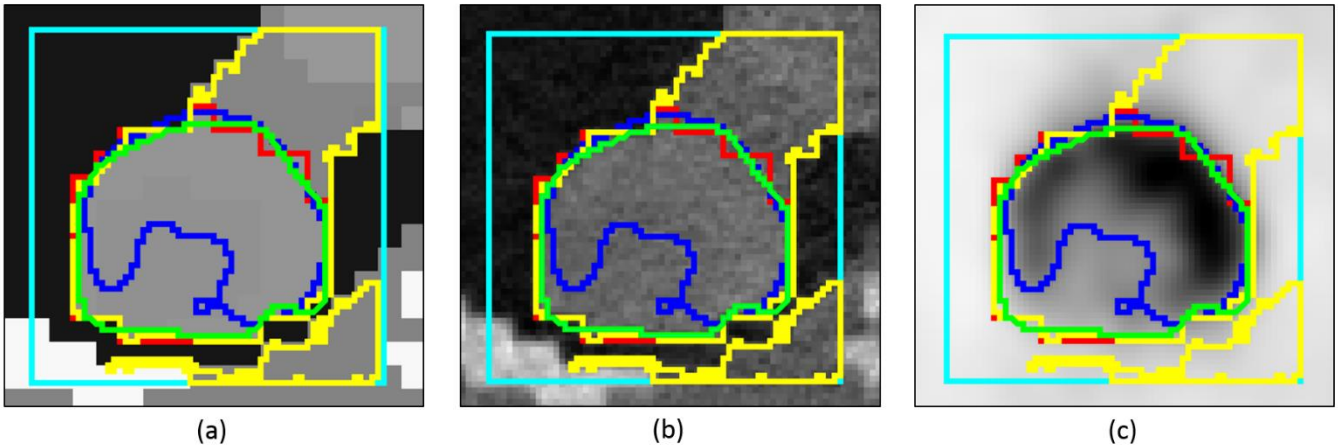


Figure 5. All contours reported to the digital phantom (a), the simulated CT (b) and the simulated PET (c). Cyan box corresponds to the initial selection used as input. Red is the ground-truth, blue is PET-only segmentation, yellow is CT-only segmentation and green is PET/CT segmentation.

Clinical datasets

Table 2 summarizes the DSC calculated respectively across the 15 cases, comparing IHMT applied to either PET or CT only, or on PET/CT. The HMT on PET/CT obtained the best accuracy with DSC of 0.89±0.13 compared to 0.50±0.28 on PET only and 0.48±0.35 on CT only. Using WHMT and CHMT the conclusions were similar with the same hierarchy, with slightly better DSC, and a small advantage for CHMT over WHMT. The differences between IMHT, WHMT and CHMT were not significant ($p>0.1$).

	PET-only	CT-only	PET/CT
IHMT	0.50±0.28	0.48±0.35	0.89±0.13
WHMT	0.61±0.29	0.56±0.31	0.90±0.12
CHMT	0.64±0.23	0.59±0.29	0.92±0.11

Table 2: Performance of IHMT, WHMT and CHMT according to DSC for the 15 cases exploiting either PET or CT only, or PET/CT.

Figure 6 illustrates two examples of IHMT segmentation on PET only, CT only and PET/CT, showing a similar behavior on a clinical case as it was shown on the simulated case.

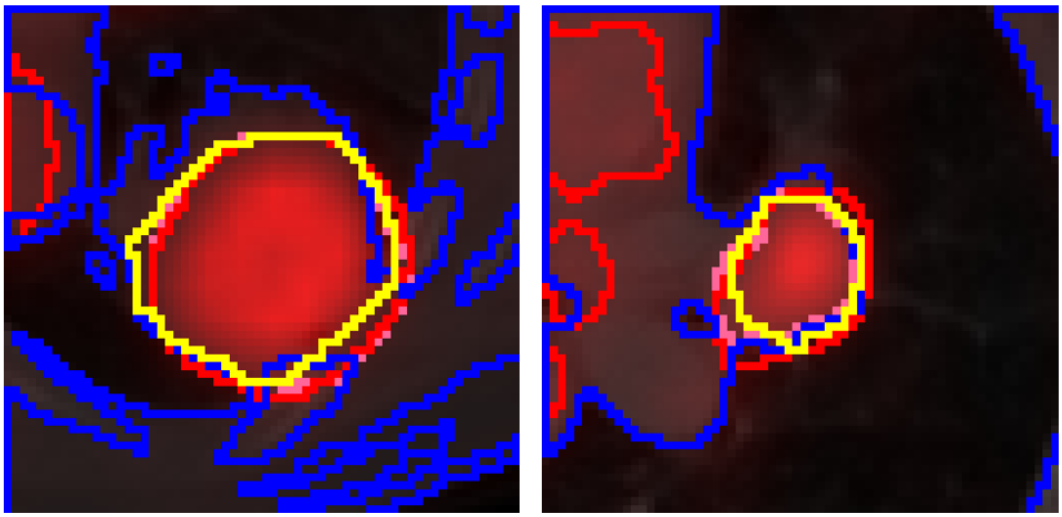


Figure 6: Comparison of monomodal (PET or CT) segmentation with multimodal PET/CT segmentation using IHMT on two cases. Yellow contour is the STAPLE consensus, whereas red, blue and pink contours are IHMT segmentations using PET only, CT only and PET/CT respectively.

Figure 7 shows 4 examples of HMT segmentations compared to the consensus of manual delineations for the PET/CT cases illustrated in figure 2.

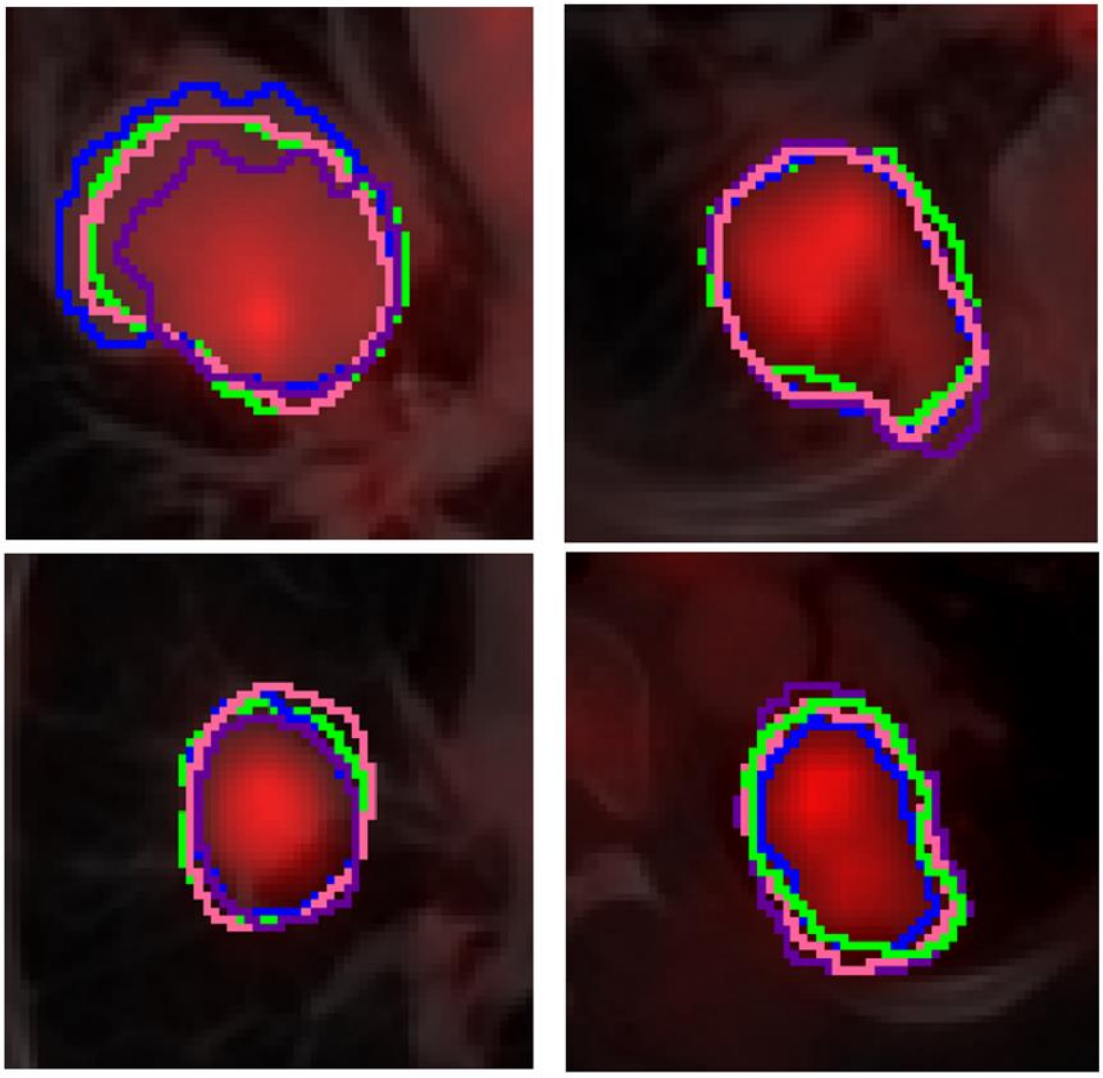


Figure 7: Comparison of IHMT (blue), WHMT (purple) and CHMT (pink) PET/CT segmentations obtained on the 4 cases from figure 2. Green contours are the STAPLE consensus.

Quantitative results of all figures of merit calculated across the 15 cases are summarized in table 3, for IHMT, WHMT CHMT, FCM and FLICM. The three HMT methods led to an accurate tumor segmentation, characterized by high PPV,

SE and DSC values. CHMT slightly outperformed the two other methods with higher mean PPV, SE and DSC values along with smaller standard deviations, thanks mostly to an improvement in sensitivity. There was however no statistically significant differences between the three approaches. On the contrary, FCM and FLIMC led to significantly lower accuracy ($p<0.01$), with slightly higher PPV but at the cost of a much lower SE (<70%) resulting in mean DSC below 0.8. The difference between FCM and FLICM was on the other hand not significant ($p=0.2$).

	FCM	FLICM	IHMT	WHMT	CHMT
SE (%)	68.3 ± 15.1	65.1 ± 14.0	85.7 ± 12.0	86.9 ± 13.1	92.7 ± 5.2
PPV (%)	89.9 ± 7.7	92.6 ± 8.0	85.2 ± 9.3	86.3 ± 16.4	86.1 ± 7.4
DSC	0.73 ± 0.19	0.75 ± 0.16	0.89 ± 0.13	0.90 ± 0.12	0.92 ± 0.11

Table 3: Performance of FCM, FLICM, IHMT, WHMT and CHMT according to SE, PPV and DSC for the 15 patients cases.

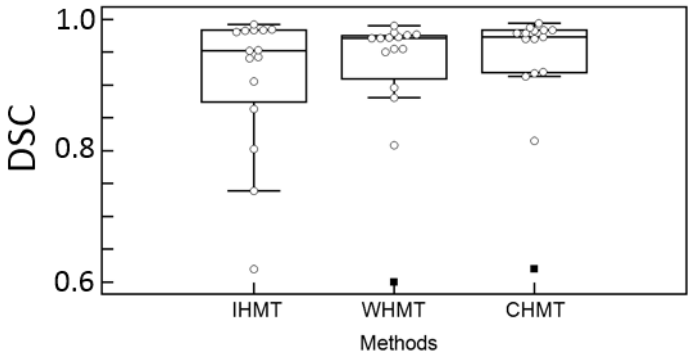


Figure 8 shows the distribution of the DSC across the entire patient datasets for the three HMT methods. Although the differences between the three methods were not significant, smaller spread and higher values were obtained with WHMT and CHMT compared to IHMT.

Figure 8: Distributions of the DSC values obtained by IHMT, WHMT and CHMT on the 15 cases.

In figure 9, we show an outlier case for which the DSC is 0.62, 0.60 and 0.62 for IHMT, WHMT and CHMT respectively. These results may be influenced by the tumor's neighborhood to the ribs and soft tissues in the lung.

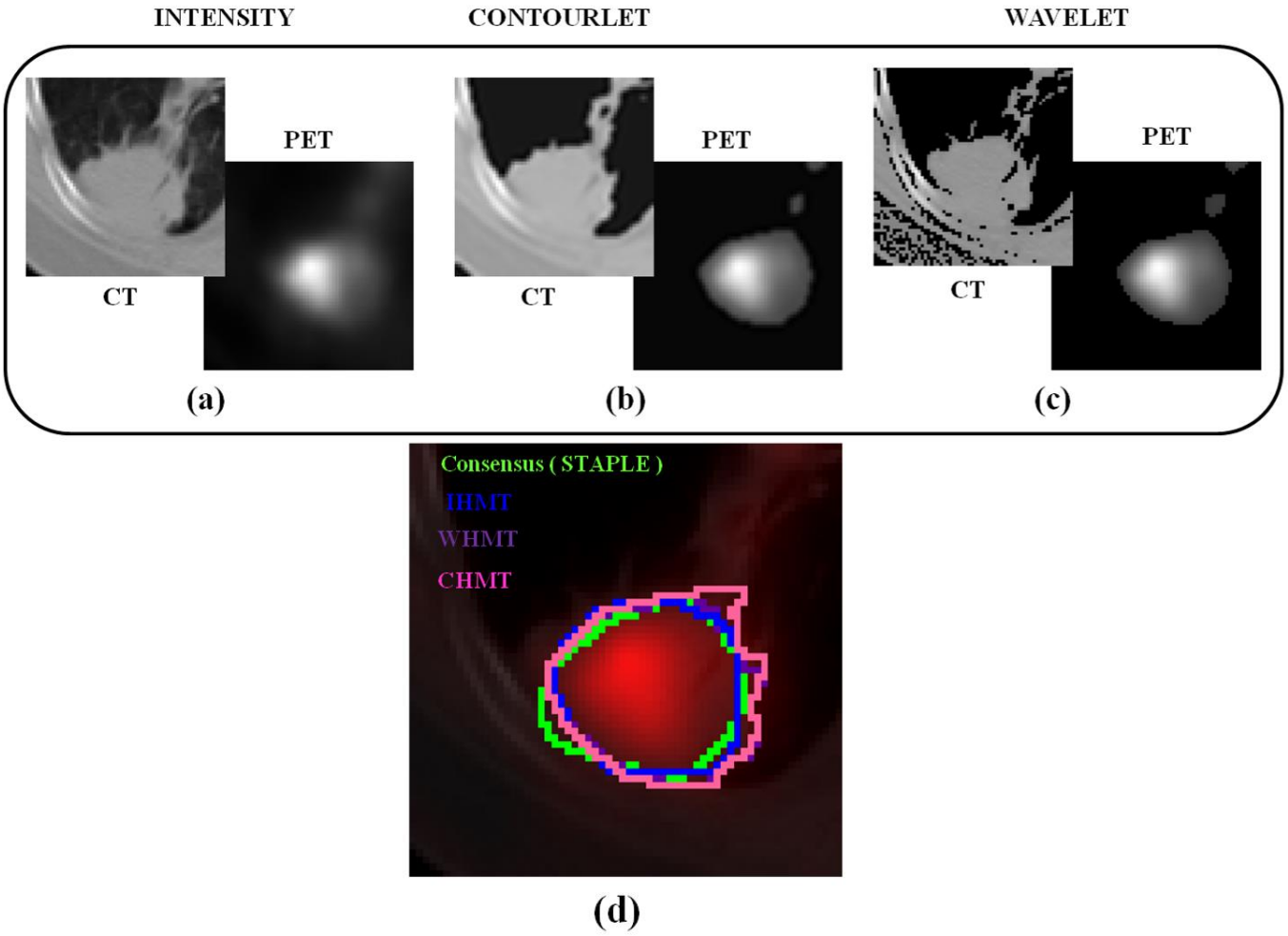


Figure 9: Illustration of the outlier case: (a) PET/CT clinical dataset image preprocessed in the (b) contourlet or (c) wavelet domain and (d) comparison of IHMT, WHMT and CHMT contours.

DISCUSSION

The accuracy of GTV definition becomes crucial in order to exploit fully image-guided, intensity-modulated, motion-compensated radiotherapy treatment, for which margins that are added to the GTV to derive the clinical (CTV) and

planning target volume (PTV) are continuously reduced with each technical improvement. The manual definition of the GTV by radiation oncologists, which is time-consuming and prone to very high inter- and intra- observer variability therefore appears to be a major limiting factor amongst the automation efforts occurring in all other steps of radiotherapy treatment planning and delivery. It has been standard practice for decades to perform manual delineation of GTV on CT scans. However, over the last few years there has been a growing exploitation of ^{18}F-FDG PET functional imaging in the radiotherapy practice. The addition of FDG PET as a complementary information within the treatment planning step by fusion of CT and PET datasets has been performed in order to improve tumor visualization and delineation by radiation oncologists. This has led to important changes in practice and guidelines for treatment planning in several cancer models including lung cancer [56, 57], including smaller or larger irradiation volumes, higher inter- and intra- observer reproducibility, and faster delineation [49].

Today radiation oncologists are therefore required to take into consideration complementary visual information using multimodality imaging when defining GTV. In addition to being used for dose calculation due to the correlation between Hounsfield units and electron density, CT has always been a modality of choice for manual delineation due to a high spatial resolution and rather homogeneous tissue densities. PET is intrinsically less well suited to the task as it suffers from limited spatial resolution, small reconstruction grids (and thus large voxels and limited spatial sampling), low signal-to-noise ratio and complex uptake distributions shapes and heterogeneities. PET however provides an often higher contrast between the tumor and healthy background, especially in challenging cases where the anatomical tumor volume is attached to the chest wall or mediastinum, or in cases of atelectasis. Manual delineation of PET/CT datasets remains a complex, tedious, time-consuming and less than ideally reproducible task [58]. Efforts to provide automated PET/CT tumor segmentations are therefore still necessary. This is going to become even more crucial as additional hybrid imaging is increasingly introduced in clinical therapy with the development of PET/MRI scanners. Despite these recent developments, the majority of commercially available automated and semi-automated segmentation algorithms do not usually consider more than one modality at a time.

Several works have recently addressed the PET/CT multimodal segmentation [12–14, 19]. A method based on a decision tree with K-Nearest Neighbors classifiers relying on textural features from both PET and CT images and originally developed for head and neck [10] was retrained for NSCLC [14]. This approach requires a training dataset to build the decision tree and is clearly dependent on it as it requires different training for each localization. Compared to a

consensus of 3 manual contours generated by STAPLE (similarly as in our study), it demonstrated very high specificity of almost 1, at the cost of low specificity (<0.7) for a mean DSC of 0.61 only.

Another framework for the segmentation of multimodal PET/CT/MRI images models functional and anatomical data in a product lattice and applies simultaneous delineations of tumor regions based on a random walk algorithm [12]. Results in term of DSC were compared for PET-only, CT-only, MRI-only, PET-CT, PET-MRI and PET-CT-MRI segmentations, with mean values of 0.832, 0.878, 0.882, 0.914, 0.931 and 0.934 respectively, demonstrating increasing accuracy with each additional image modality being exploited by the algorithm. Our results comparing segmentations using either PET or CT only, or PET/CT, are in line with these observations, higher accuracy being obtained when exploiting both modalities.

Another approach used a HMF based segmentation of PET and CT image separately with a regularized term penalizing the segmentation difference between the two modalities [19]. PET/CT segmentation led to an accuracy of 0.86±0.05 in terms of DSC which is comparable to our results, with however likely lower computational efficiency: HMT ensures better computational properties and faster estimation of parameters than HMF [17]. A similar approach was evaluated by comparing PET-only, CT-only and the association of PET and CT images through optimization of a HMF based segmentation with the graph cut algorithm, where each node is associated with a pair of corresponding voxels in PET and CT images [13]. More importantly, the method allows obtaining different segmentations on the PET and the CT through this co-segmentation framework. This original approach was compared with a standard fused PET/CT segmentation, the surrogate of truth being a manual delineation by a radiation oncologist on CT images with the guidance of PET images. The proposed method combining PET and CT led to a much higher accuracy than PET-only and CT-only in a set of 23 lung tumors (DSC of 0.81±0.08 compared to 0.66±0.13 and 0.48±0.27 with PET-only and CT-only). However, the improvement of the proposed approach compared to a segmentation on fused PET/CT, comparable to our approach, (DSC of 0.79±0.08) was not significant.

Note that the comparison with our results is challenging since these methods were evaluated on different datasets. However, we can also observe that the range of overlap values we obtained in the present work compared with manual delineation, is competitive or even better than those previously reported in similar evaluation settings.

In the present work, we developed a HMT based framework allowing the exploitation of several image modalities of the same object of interest (here, tumors). This framework was implemented under the assumption that the

segmentation would benefit from combining both modalities. We aimed at automating the manual task of the radiation oncologist in defining a single GTV from PET/CT images, contrary to a CT segmentation guided by PET (or the opposite). The proposed framework was evaluated NSCLC cases previously considered as a reference comparison set by several other studies [16, 43, 59]. Given our goal (a single GTV derived from both PET and CT images, as is performed by radiation oncologists), the validation protocol consisted in comparing automatic results with the surrogate of truth defined by a statistical consensus of three manual delineations, in order to address inter-observer variability. The accuracy was assessed using 3 complementary metrics in order to provide a full picture of the performance. In addition, we investigated the potential benefit of a pre-processing filtering step of the PET/CT images using either the contourlet (CHMT) or wavelet (WHMT) transforms.

Our proposed multimodal HMT framework achieved high overlap with the surrogate of truth HMT, providing balanced segmentation with high PPV and high SE (both >0.80). Overall, the results were slightly improved when applying the HMT framework on images filtered in the contourlet domain, although the difference was not significant. All HMT methods significantly outperformed ($p<0.01$) multimodal FCM and FLICM, which tended to under-estimate volumes (high PPV, low SE).

In this work, we proposed a joint-segmentation framework for anatomical and functional images that has the potential to combine more than two images (in the present study PET and CT), since extra modalities can be handled and analyzed simultaneously. As an extension of this work, we thus aim to test the proposed framework on PET/CT rectal cancer datasets [60, 61], as well as MRI-PET and MRI-CT-PET datasets.

Our work has several limitations. The proposed framework assumes an appropriate co-registration of the considered images, with an accurate correspondence between PET and CT images. Data were acquired in free-breathing acquisitions and no correction for respiratory motion was applied. Datasets with obvious mismatches between PET and CT data were excluded prior to the present evaluation, as they would lead to strong bias in both manual and automatic segmentations. The inclusion of datasets strongly affected by respiratory motion would require pre-processing corrections outside the scope of the present study [62, 63]. In our NSCLC cases, bias due to breathing motion cannot be excluded, even after exclusion of the obvious cases. Other tumor locations with less/no motion could be considered such as head and neck or brain tumors, however we do not possess PET/CT images for such tumors for which histopathology measurements are available, the only dataset available to us consisting of head and neck PET images

only [64]. We did not use the maximum diameter measurement as an evaluation metric, because it would not allow differentiating spatial overlap differences between methods. It was however used to check the reliability of the manual consensus. Another limitation concerns the wavelet and contourlet pre-processing steps: we empirically set the threshold under which significant wavelets and contourlets coefficients were excluded. In the previous work that proposed the approach, the value was set arbitrarily at 200 without justification [38]. We found this parameter is quite robust, as choosing different values only lowered the accuracy of WHMT and CHMT to the level of IHMT. We plan to further improve the robustness of this step by implementing a case-by-case automated adaptive selection of the most significant coefficients. Our HMT-based approach could not be directly compared to other previously published advanced multimodal segmentation methods. These are indeed not freely available and are complex to implement, which was beyond the scope of the present work. The current efforts of the AAPM Task Group 211 for PET image segmentation benchmarking [42] are going to be expanded to address PET/CT segmentation. When ready, such a benchmark will allow a more comprehensive and direct comparison of various multimodal algorithms. We did not investigate our approach within the context of dose painting or dose redistribution/boosting, as our goal was to automate the definition of the GTV. Such a GTV could however constitute a starting point for a dose painting/boosting/redistribution strategy analysis within this GTV.

Although the HMT used to build the proposed framework showed interesting properties in terms of overall performance, robustness and stability, we will further investigate the use of more sophisticated HMT models, among which we can cite pairwise Markov trees (PMTs) that extend the classic HMT [65]. Another rich recent family of models are the triplet Markov trees (TMTs) [66], which extend PMTs and combined with the theory of evidence (Dempster-Shafer), can be useful when dealing with non-stationary data [67]. These future developments could be exploited to deal with other applications of the framework such as partial volume effects correction, therapy follow-up using sequential PET/CT datasets, and fusion of multi tracer PET/CT datasets, or even motion characterization and correction.

CONCLUSIONS

In this work, we developed a framework based on HMT for multimodal image processing and analysis and investigated its relevance for multimodal PET/CT segmentation. Tumor delineation was performed by exploiting the hierarchical and multi-observation properties of a HMT to exploit the CT and the PET image simultaneously. The impact of pre-

processing in the wavelet and contourlet domains on the segmentation accuracy was also investigated. The proposed method led to high accuracy, the best results being obtained after filtering in the contourlet domain, with higher DSC and the best trade-off between sensitivity and positive predictive value. HMT models could be further applied to PET/MR and PET multitracer data. Validation of the PET/CT segmentation on datasets with full volume histopathological reference will also be investigated.

Acknowledgements: This work was funded by a grant from the French Ministry of Research. This research is also supported by the Dutch technology Foundation STW (grant n° 10696 DuCAT & n° P14-19 Radiomics STRaTegy), which is the applied science division of NWO, and the Technology Programme of the Ministry of Economic Affairs.

Disclosure of conflict of interest: The authors have no relevant conflicts of interest to disclose.

REFERENCES

1. D. De Ruysscher, J. Belderbos, B. Reymen, W. van Elmpt, A. van Baardwijk, R. Wanders, F. Hoebers, M. Vooijs, M. Ollers, and P. Lambin, "State of the art radiation therapy for lung cancer 2012: a glimpse of the future," Clin. Lung Cancer **14**(2), 89–95 (2013).
2. F.E. Turkheimer, N. Boussion, A.N. Anderson, N. Pavese, P. Piccini, and D. Visvikis, "PET image denoising using a synergistic multiresolution analysis of structural (MRI/CT) and functional datasets," J. Nucl. Med. Off. Publ. Soc. Nucl. Med. **49**(4), 657–666 (2008).
3. P. Novosad and A.J. Reader, "MR-guided dynamic PET reconstruction with the kernel method and spectral temporal basis functions," Phys. Med. Biol. **61**(12), 4624–4644 (2016).
4. C. Coello, F. Willoch, P. Selnes, L. Gjerstad, T. Fladby, and A. Skretting, "Correction of partial volume effect in (18)F-FDG PET brain studies using coregistered MR volumes: voxel based analysis of tracer uptake in the white matter," NeuroImage **72**, 183–192 (2013).
5. A. Le Pogam, M. Hatt, P. Descourt, N. Boussion, C. Tsoumpas, F.E. Turkheimer, C. Prunier-Aesch, J.-L. Baulieu, D. Guilloteau, and D. Visvikis, "Evaluation of a 3D local multiresolution algorithm for the correction of partial volume effects in positron emission tomography," Med. Phys. **38**(9), 4920–4923 (2011).
6. A. Bousse, S. Pedemonte, B.A. Thomas, K. Erlandsson, S. Ourselin, S. Arridge, and B.F. Hutton, "Markov random field and Gaussian mixture for segmented MRI-based partial volume correction in PET," Phys. Med. Biol. **57**(20), 6681–6705 (2012).
7. S. David, D. Visvikis, G. Quellec, C.C. Le Rest, P. Fernandez, M. Allard, C. Roux, and M. Hatt, "Image change detection using paradoxical theory for patient follow-up quantitation and therapy assessment," IEEE Trans. Med. Imaging **31**(9), 1743–1753 (2012).
8. H. Necib, C. Garcia, A. Wagner, B. Vanderlinden, P. Emonts, A. Hendlisz, P. Flamen, and I. Buvat, "Detection and characterization of tumor changes in 18F-FDG PET patient monitoring using parametric imaging," J Nucl Med **52**(3), 354–61 (2011).
9. B. Lelandais, S. Ruan, T. Denœux, P. Vera, and I. Gardin, "Fusion of multi-tracer PET images for dose painting," Med. Image Anal. **18**(7), 1247–1259 (2014).
10. H. Yu, C. Caldwell, K. Mah, I. Poon, J. Balogh, R. MacKenzie, N. Khaouam, and R. Tirona, "Automated radiation targeting in head-and-neck cancer using region-based texture analysis of PET and CT images," Int. J. Radiat. Oncol. Biol. Phys. **75**(2), 618–625 (2009).
11. I. El Naqa, D. Yang, A. Apte, D. Khullar, S. Mutic, J. Zheng, J.D. Bradley, P. Grigsby, and J.O. Deasy, "Concurrent multimodality image segmentation by active contours for radiotherapy treatment planning," Med Phys **34**(12), 4738–49 (2007).
12. U. Bagci, J.K. Udupa, N. Mendhiratta, B. Foster, Z. Xu, J. Yao, X. Chen, and D.J. Mollura, "Joint segmentation of anatomical and functional images: Applications in quantification of lesions from PET, PET-CT, MRI-PET, and MRI-PET-CT images," Med. Image Anal. **17**(8), 929–945 (2013).
13. Q. Song, J. Bai, D. Han, S. Bhatia, W. Sun, W. Rockey, J.E. Bayouth, J.M. Buatti, and X. Wu, "Optimal co-segmentation of tumor in PET-CT images with context information," IEEE Trans. Med. Imaging **32**(9), 1685–1697 (2013).
14. D. Markel, C. Caldwell, H. Alasti, H. Soliman, Y. Ung, J. Lee, and A. Sun, "Automatic Segmentation of Lung Carcinoma Using 3D Texture Features in 18-FDG PET/CT," Int. J. Mol. Imaging **2013**, 980769 (2013).
15. H. Cui, X. Wang, J. Zhou, S. Eberl, Y. Yin, D. Feng, and M. Fulham, "Topology polymorphism graph for lung tumor segmentation in PET-CT images," Phys. Med. Biol. **60**(12), 4893–4914 (2015).
16. D. Markel, H. Zaidi, and I. El Naqa, "Novel multimodality segmentation using level sets and Jensen-Rényi divergence," Med. Phys. **40**(12), 121908 (2013).
17. W. Pieczynski, "Modèles de Markov en traitement d'images," Trait. Signal **20**(30), 255–278 (2003).
18. D.W. Montgomery, A. Amira, and H. Zaidi, "Fully automated segmentation of oncological PET volumes using a combined multiscale and statistical model," Med Phys **34**(2), 722–36 (2007).
19. D. Han, J. Bayouth, Q. Song, A. Taurani, M. Sonka, J. Buatti, and X. Wu, "Globally Optimal Tumor Segmentation in PET-CT Images: A Graph-based Co-segmentation Method," in *Proc. 22Nd Int. Conf. Inf. Process. Med. Imaging*(Springer-Verlag, Berlin, Heidelberg, 2011), pp. 245–256.
20. M. Hatt, F. Lamare, N. Boussion, A. Turzo, C. Collet, F. Salzenstein, C. Roux, P. Jarritt, K. Carson, C. Cheze-Le Rest, and D. Visvikis, "Fuzzy hidden Markov chains segmentation for volume determination and quantitation in PET," Phys Med Biol **52**(12), 3467–91 (2007).
21. A.S. Willsky, "Multiresolution Markov models for signal and image processing," Proc. IEEE **90**(8), 1396–1458 (2002).
22. G. Guénard, P. Legendre, and P. Peres-Neto, "Phylogenetic eigenvector maps: a framework to model and predict species traits," Methods Ecol. Evol. **4**(12), 1120–1131 (2013).
23. J. Porway, Q. Wang, and S.C. Zhu, "A Hierarchical and Contextual Model for Aerial Image Parsing," Int. J. Comput. Vis. **88**(2), 254–283 (2010).
24. F. Flitti, C. Collet, B. Vollmer, and F. Bonnarel, "Multiband Segmentation of a Spectroscopic Line Data Cube: Application to the HI Data Cube of the Spiral Galaxy NGC 4254," EURASIP J Appl Signal Process **2005**, 2546–2558 (2005).

[25] J.M. Laferté, P. Pérez, and F. Heitz, "Discrete Markov image modeling and inference on the quadtree," IEEE Trans. Image Process. Publ. IEEE Signal Process. Soc. **9**(3), 390–404 (2000).
[26] F. Li, X. Jia, D. Fraser, and A. Lambert, "Super Resolution for Remote Sensing Images Based on a Universal Hidden Markov Tree Model," IEEE Trans. Geosci. Remote Sens. **48**(3), 1270–1278 (2010).
[27] C. Li and Q. Hao, "Functional MR image statistical restoration for neural activity detection using Hidden Markov Tree model," Int. J. Comput. Biol. Drug Des. **6**(3), 190–209 (2013).
[28] J. Wu and A.C.S. Chung, "A novel framework for segmentation of deep brain structures based on Markov dependence tree," NeuroImage **46**(4), 1027–1036 (2009).
[29] M. Wels, Y. Zheng, M. Huber, J. Hornegger, and D. Comaniciu, "A discriminative model-constrained EM approach to 3D MRI brain tissue classification and intensity non-uniformity correction," Phys. Med. Biol. **56**(11), 3269–3300 (2011).
[30] H. Hanzouli, J. Lapuyade-Lahorgue, E. Monfrini, G. Delso, W. Pieczynski, D. Visvikis, and M. Hatt, "PET/CT image denoising and segmentation based on a multi observation and a multi scale Markov tree model," in *2013 IEEE Nucl. Sci. Symp. Med. Imaging Conf. NSSMIC*(2013), pp. 1–4.
[31] Y. Delignon, A. Marzouki, and W. Pieczynski, "Estimation of generalized mixtures and its application in image segmentation," IEEE Trans Image Process **6**(10), 1364–75 (1997).
[32] M. Hatt, C. Cheze le Rest, A. Turzo, C. Roux, and D. Visvikis, "A fuzzy locally adaptive Bayesian segmentation approach for volume determination in PET," IEEE Trans Med Imaging **28**(6), 881–93 (2009).
[33] T.K. Pogány and S. Nadarajah, "On the characteristic function of the generalized normal distribution," Comptes Rendus Math. **348**(3–4), 203–206 (2010).
[34] G. Celeux and J. Diebolt, "A stochastic approximation type EM algorithm for the mixture problem," (1991).
[35] J.-B. Durand and P. Gonçalves, "Statistical Inference for Hidden Markov Tree Models and Application to Wavelet Trees," (2001).
[36] L. Baum, "{An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes}," Inequalities **3**, 1–8 (1972).
[37] P.A. Devijver, "Baum's forward-backward algorithm revisited," Pattern Recognit. Lett. **3**(6), 369–373 (1985).
[38] M. Abdoli, R.A.J.O. Dierckx, and H. Zaidi, "Contourlet-based active contour model for PET image segmentation," Med. Phys. **40**(8), 082507 (2013).
[39] P. Dutilleux, "An Implementation of the 'algorithme à trous' to Compute the Wavelet Transform," in *Wavelets*, edited by P.J.-M. Combes, P.A. Grossmann and P.P. Tchamitchian (Springer Berlin Heidelberg, 1989), pp. 298–304.
[40] M.N. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," IEEE Trans. Image Process. **14**(12), 2091–2106 (2005).
[41] A.L. da Cunha, J. Zhou, and M.N. Do, "The Nonsubsampled Contourlet Transform: Theory, Design, and Applications," IEEE Trans. Image Process. **15**(10), 3089–3101 (2006).
[42] M. Hatt, J. Lee, C.R. Schmidtlein, I. El Naqa, C. Caldwell, E. De Bernardi, W. Lu, S. Das, X. Geets, V. Gregoire, R. Jeraj, M.P. MacManus, O.R. Mawlawi, U. Nestle, A.B. Pugachev, H. Schöder, T. Shepherd, E. Spezi, D. Visvikis, H. Zaidi, and A.S. Kirov, "Classification and evaluation strategies of auto-segmentation approaches for PET: Report of AAPM Task Group No. 211," Med. Phys. (2017).
[43] M. Hatt, C. Cheze le Rest, P. Descourt, A. Dekker, D. De Ruysscher, M. Oellers, P. Lambin, O. Pradier, and D. Visvikis, "Accurate automatic delineation of heterogeneous functional volumes in positron emission tomography for oncology applications," Int J Radiat Oncol Biol Phys **77**(1), 301–8 (2010).
[44] A. Le Maitre, W. Segars, S. Marache, A. Reilhac, M. Hatt, S. Tomei, C. Lartizien, and D. Visvikis, "Incorporating Patient-Specific Variability in the Simulation of Realistic Whole-Body 18F-FDG Distributions for Oncology Applications," Proc. IEEE **9**(12), 2026–2038 (2009).
[45] P. Papadimitroulas, G. Loudos, A. Le Maitre, M. Hatt, F. Tixier, N. Efthimiou, G.C. Nikiforidis, D. Visvikis, and G.C. Kagadis, "Investigation of realistic PET simulations incorporating tumor patient's specificity using anthropomorphic models: creation of an oncology database," Med. Phys. **40**(11), 112506 (2013).
[46] R.L. Siddon, "Fast calculation of the exact radiological path for a three-dimensional CT array," Med. Phys. **12**(2), 252–255 (1985).
[47] K. Van Slambrouck and J. Nuyts, "Reconstruction Scheme for Accelerated Maximum Likelihood Reconstruction: The Patchwork Structure," IEEE Trans. Nucl. Sci. **61**(1), 173–181 (2014).
[48] P. Thevenaz, T. Blu, and M. Unser, "Interpolation revisited," IEEE Trans Med Imaging **19**(7), 739–58 (2000).
[49] A. van Baardwijk, G. Bosmans, L. Boersma, J. Buijsen, S. Wanders, M. Hochstenbag, R.-J. van Suylen, A. Dekker, C. Dehing-Oberije, R. Houben, S.M. Bentzen, M. van Kroonenburgh, P. Lambin, and D. De Ruysscher, "PET-CT-based auto-contouring in non-small-cell lung cancer correlates with pathology and reduces interobserver variability in the delineation of the primary tumor and involved nodal volumes," Int. J. Radiat. Oncol. Biol. Phys. **68**(3), 771–778 (2007).
[50] S.K. Warfield, K.H. Zou, and W.M. Wells, "Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation," IEEE Trans. Med. Imaging **23**(7), 903–921 (2004).
[51] S. Krinidis and V. Chatzis, "A robust fuzzy local information C-Means clustering algorithm," IEEE Trans. Image Process. Publ. IEEE Signal Process. Soc. **19**(5), 1328–1337 (2010).
[52] L.R. Dice, "Measures of the Amount of Ecologic Association Between Species," Ecology **26**(3), 297 (1945).

[53] T.F. Heston, "Standardizing predictive values in diagnostic imaging research," J. Magn. Reson. Imaging JMRI **33**(2), 505; author reply 506-507 (2011).
[54] R. Parikh, A. Mathai, S. Parikh, G. Chandra Sekhar, and R. Thomas, "Understanding and using sensitivity, specificity and predictive values," Indian J. Ophthalmol. **56**(1), 45–50 (2008).
[55] K. Wu, Y.C. Ung, D. Hwang, M.S. Tsao, G. Darling, D.E. Maziak, R. Tirona, K. Mah, and C.S. Wong, "Autocontouring and manual contouring: which is the better method for target delineation using 18F-FDG PET/CT in non-small cell lung cancer?," J. Nucl. Med. Off. Publ. Soc. Nucl. Med. **51**(10), 1517–1523 (2010).
[56] M.P. Mac Manus and R.J. Hicks, "The role of positron emission tomography/computed tomography in radiation therapy planning for patients with lung cancer," Semin. Nucl. Med. **42**(5), 308–319 (2012).
[57] J. van Loon, A. van Baardwijk, L. Boersma, M. Ollers, P. Lambin, and D. De Ruysscher, "Therapeutic implications of molecular imaging with PET in the combined modality treatment of lung cancer," Cancer Treat. Rev. **37**(5), 331–343 (2011).
[58] M. Hatt, C.C.-L. Rest, E.O. Aboagye, L.M. Kenny, L. Rosso, F.E. Turkheimer, N.M. Albarghach, J.-P. Metges, O. Pradier, and D. Visvikis, "Reproducibility of 18F-FDG and 3′-Deoxy-3′-18F-Fluorothymidine PET Tumor Volume Measurements," J. Nucl. Med. **51**(9), 1368–1376 (2010).
[59] M. Hatt, C. Cheze-le Rest, A. van Baardwijk, P. Lambin, O. Pradier, and D. Visvikis, "Impact of tumor size and tracer uptake heterogeneity in (18)F-FDG PET and CT non-small cell lung cancer tumor delineation," J Nucl Med **52**(11), 1690–7 (2011).
[60] J. Buijsen, J. van den Bogaard, H. van der Weide, S. Engelsman, R. van Stiphout, M. Janssen, G. Beets, R. Beets-Tan, P. Lambin, and G. Lammering, "FDG-PET-CT reduces the interobserver variability in rectal tumor delineation," Radiother Oncol **102**(3), 371–6 (2012).
[61] J. Buijsen, J. van den Bogaard, M.H. Janssen, F.C. Bakers, S. Engelsman, M. Ollers, R.G. Beets-Tan, M. Nap, G.L. Beets, P. Lambin, and G. Lammering, "FDG-PET provides the best correlation with the tumor specimen compared to MRI and CT in rectal cancer," Radiother Oncol **98**(2), 270–6 (2011).
[62] F. Lamare, M.J. Ledesma Carbayo, T. Cresson, G. Kontaxakis, A. Santos, C.C. Le Rest, A.J. Reader, and D. Visvikis, "List-mode-based reconstruction for respiratory motion correction in PET using non-rigid body transformations," Phys Med Biol **52**(17), 5187–204 (2007).
[63] J. Orban de Xivry, G. Janssens, G. Bosmans, M. De Craene, A. Dekker, J. Buijsen, A. van Baardwijk, D. De Ruysscher, B. Macq, and P. Lambin, "Tumour delineation and cumulative dose computation in radiotherapy based on deformable registration of respiratory correlated CT images of lung cancer patients," Radiother Oncol **85**(2), 232–8 (2007).
[64] X. Geets, J.A. Lee, A. Bol, M. Lonneux, and V. Gregoire, "A gradient-based method for segmenting FDG-PET images: methodology and validation," Eur J Nucl Med Mol Imaging **34**(9), 1427–38 (2007).
[65] E. Monfrini, J. Lecomte, F. Desbouvries, and W. Pieczynski, "Image and signal restoration using pairwise Markov trees," in *2003 IEEE Workshop Stat. Signal Process.*(2003), pp. 174–177.
[66] W. Pieczynski, "Arbres de Markov Triplet et fusion de Dempster–Shafer," Comptes Rendus Math. **336**(10), 869–872 (2003).
[67] P. Lanchantin and W. Pieczynski, "Evidential Markov chains and trees with applications to non stationary processes segmentation," Trait. Signal **22**(1), 15–26 (2005).

Appendix

Annex

Let S_0 , S_1 , … , S_N be successive scales, with $S_0 = \{r\}$ containing just the root, S_1 containing four children of r , and so on, as showed in Figure 1. Let $S = S_0 \cup S_1 \cup ... \cup S_N$ be the set of all nodes. For $s \in S_n$, with $0 \le n < N$, we will denote with s^+ the four children of s (which are in S_{n+1}), and for $s \in S_n$, with $0 < n \le N$ we will denote with s^- the unique parent of s (which is in S_{n-1}). The distribution of the classic HMT $(X,Y) = (X_s, Y_s)_{s \in S}$ used in the paper is defined by

$$p(x,y) = [\, p(x_r) \prod_{n=0}^{N-1} \prod_{s_n \in S_n} \prod_{s_{n+1} \in s_n^+} p(x_{s_{n+1}} | x_{s_n})] \prod_{s_N \in S_N} p(y_{s_N} | x_{s_N}), \qquad (1)$$

$p(y_{s_N} | x_{s_N})$ are generalized Gaussian distributions modeling PET/CT images $y_{s_N} = (y_{s_N}^{PET}, y_{s_N}^{CT})$. Then the problem is to compute $p(x_s | y_{s_N})$ for each $s \in S_N$, which is the probability to have $x_s = tumor$, $x_s = notumor$. Then co-segmenting images $y = y_{s_N} = (y_{s_N}^{PET}, y_{s_N}^{CT})$ consists of setting at each node $s \in S_N$ "tumor" if $p(x_s = tumor | y) > p(x_s = nontumor | y)$, and "no tumor" if not. Setting ω_1 as "tumor" and ω_2 as "no tumor" the co-segmentation $y = y_{s_N} = (y_{s_N}^{PET}, y_{s_N}^{CT}) \to \hat{x}_{s_N}$ is thus given with

$$\text{For each node } s \in S_N : \hat{x}_s = \begin{cases} \omega_1 & \text{if } p(x_s = \omega_1 | y) > p(x_s = \omega_2 | y) \\ \omega_2 & \text{if } p(x_s = \omega_1 | y) < p(x_s = \omega_2 | y) \end{cases} \qquad (2)$$

The computation of $p(x_s | y)$ for each $s \in S_N$ is made on two steps: Backward - Up step, and Forward - Down step. For each S_n and $s \in S_n$ let s^{++} be the set of all the descendent of s (which is the set of leaves of the sub-tree having s for root), and let $\beta^n(x_s) = p(y_{s^{++}} | x_s)$ for $s \in S - S_N$, and $\beta^N(x_s) = 1$. Then one shows:

$$\beta^n(x_s) = \prod_{u \in s^+} [\sum_{x_u \in \Omega} \beta^{n+1}(u_s)\, p(x_u | x_s)], \qquad (3)$$

which makes $\beta^n(x_s)$ computable for each $1 \le n \le N$, $x_s \in S_n$.

The core point is that the distribution of the whole tree X_S conditional on both observed images $Y_{S_N} = y = (y_{s_N}^{PET}, y_{s_N}^{CT})$, denoted as p^y, also is a Markov tree distribution, with $p^y(x_r)$ and transitions $p^y(x_s | x_{s^-})$ given by

$$p^y(x_r) = \frac{p(x_r)\beta^1(x_s)}{\sum_{x_r \in \Omega} p(x_r)\beta^1(x_s)}, \quad p^y(x_s | x_{s^-}) = \frac{\beta^{n+1}(x_s)\, p(x_s | x_{s^-})}{\sum_{x_s \in \Omega} \beta^{n+1}(x_s)\, p(x_s | x_{s^-})} \qquad (4)$$

Having calculated $p^y(x_r)$ and $p^y(x_s | x_{s^-})$ with (3), (4), probabilities $p(x_s | y)$ for each $s \in S_N$ are computed by the following Forward – Down procedure.

Let $s \in S_N$. There exists an unique path $s_1, s_2, \ldots, s_N$ such that $s_N = s$, $s_{N-1} = s^-, \ldots, s_1 = s_r = s_2^-$. The distribution of $X_{s_1}, X_{s_2}, \ldots, X_{s_N}$ conditional on $Y = y$ is then a Markov chain distribution, and $p^y(x_s)$ is classically computable by:

$p^y(x_{s_1}) = p^y(x_r)$ given with (4) ; for $n = 1, \ldots, N-1$: $p^y(x_{s_{n+1}}) = \sum_{x_{s_n} \in \Omega} p^y(x_{s_n}) p^y(x_{s_{n+1}} | x_{s_n})$ (5)

Finally the co-segmentation algorithm runs as follows:

(1) For each for $s \in S$ and for $x_s \in \Omega$ compute $\beta^n(x_s) = p(y_{s^{++}} | x_s)$ par Backward – Up procedure (3);

(2) For each $s \in S$ and $x_s \in \Omega$ compute $p^y(x_r)$, $p^y(x_s | x_{s^-})$ with (4);

(3) For each $s \in S_N$ compute $p(x_s | y_{s_N})$ with the Forward – Down procedure (5);

(4) Perform co-segmentation using (2).