



HAL
open science

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

Houda Hanzouli-Ben, Jerome 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, Jerome 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-01577785>

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.

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

3

4 Houda HANZOULI-BEN SALAH¹, Jerome LAPUYADE-LAHORGUE¹, Julien BERT¹, Didier BENOIT¹, Philippe
5 LAMBIN², Angela VAN BAARDWIJK², Emmanuel MONFRINI³, Wojciech PIECZYNSKI³, Dimitris VISVIKIS¹,
6 Mathieu HATT¹

7 ¹ INSERM, UMR 1101, LaTIM, IBSAM, UBO, UBL, Brest, France

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

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

11

12 Corresponding author:

13 Mathieu Hatt

14 INSERM, UMR 1101, LaTIM

15 IBRBS – faculté de médecine

16 22 rue Camille Desmoulins

17 29238 Brest

18 France

19 Tel: +33 2 98 01 81 11

20 Fax: +33 2 98 01 81 24

21 Email: hatt@univ-brest.fr

22

23 Wordcount: ~9800

24

25 Abstract

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

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

40 Results: A high agreement with the consensus of manual delineations was observed when using both PET and CT
41 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
42 for Intensity-based HMT and Wavelet-based HMT respectively. Considering PET or CT only in the HMT led to much
43 lower accuracy. Standard and improved FCM led to comparatively lower accuracy than HMT, even when considering
44 multimodal implementations.

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

47

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

50

51

52 INTRODUCTION

53

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

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

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

74 Several papers have investigated the combination and association of functional and anatomical images for a specific
75 purpose, such as denoising² or reconstruction³ of PET data using anatomical priors coming from MRI or CT, correction
76 of partial volume effects in PET or SPECT by exploiting the associated morphological information ⁴⁻⁶, the use of
77 associated CT images to spatially register several low-resolution PET images, for instance during treatment ^{7, 8} or

78 several different radiotracers⁹, or the definition of tumor target volumes in radiotherapy by considering both PET and
79 CT images features^{10,11}.

80 This last application in particular has led recently to the development of several (semi)automated methodologies aiming
81 at determining a single gross tumor volume (GTV) from fused PET/CT datasets, assuming both modalities provide
82 complementary information regarding a single contour of the tumor volume¹⁰⁻¹⁶. Although an alternative approach
83 could consist in segmenting each image independently and then combine the obtained volumes *a posteriori* using some
84 kind of consensus, these approaches have mostly considered the first approach and used a variety of tools including
85 supervised learning using textural features of PET/CT images^{10,14}, Markov random field¹³, random walk segmentation
86¹², active contours^{11,16} and topology integrated within graph-models¹⁵. The idea underlying all these studies is that the
87 delineation should benefit from the co-registered modalities' information in order to produce a single, more robust and
88 accurate GTV. Most of these previous works have indeed highlighted the potential value in considering both modalities
89 simultaneously in a common mathematical and algorithmic framework. The objective of most of these studies is to
90 objectively automate the manual and subjective task of GTV delineation by a radiation oncologist visualizing the fused
91 PET/CT images. We place our present work in the same context.

92 Hidden Markov fields (HMF) are well suited to image processing and have been exploited in a variety of applications,
93 including medical imaging¹⁷. They have been used in PET segmentation¹⁸ and two recent studies investigated the use
94 of HMF to associate PET and CT datasets with the goal of co-segmentation^{13,19}. Markov random fields however
95 require computationally expensive iterative methods for estimating parameters¹⁷, contrary to hidden Markov chains
96 (HMC) that have also been explored for PET segmentation²⁰. Finally, hidden Markov trees (HMT) share similar
97 favorable computational properties with chains, while being quite as robust as fields and providing a useful structure
98 for the association of multiscale and multimodal images^{17,21}. HMT proved their efficiency in several fields. For
99 example, phylogenetic models in biology are diagrams showing the evolutionary relationships between species²².
100 Hierarchical models are also used for aerial image parsing²³. HMT have found numerous applications in imaging,
101 including astronomical²⁴, remote sensing^{25,26} and medical imaging for instance in MRI²⁷. However, medical image
102 segmentation using HMT is quite rare²⁸ contrary to Markov fields²⁹, and to the best of our knowledge, our work is the
103 first to investigate their use for multimodal PET/CT image segmentation. The present paper is an extension of the first
104 developments initially presented in a conference proceeding³⁰.

105

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

111 **MATERIALS AND METHODS**

112 **1. Hidden Markov Tree framework**

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

115 The multiresolution nature of the data can be managed by using the successive scales of the HMT in order to handle
116 data represented at different resolutions. Another modeling approach consists in using the different levels in the HMT
117 not to represent data of different resolutions, but rather as an algorithmic solution to model dependencies between
118 neighboring observations at successive scales, while the data is associated only with the last scale of the HMT. In that
119 case, different modalities are handled by considering observed data in a given voxel as a vector containing several
120 values, each vector component representing a modality's observation. This setting was chosen for the present work: in
121 radiotherapy planning clinical practice, the PET image is up-sampled to exactly the same dimension as the CT, so that
122 the images can be overlaid for visualization and manual delineation. We used cubic B-spline interpolation to up-sample
123 the PET image so that a PET voxel exactly corresponds to a CT voxel. Therefore both PET and CT voxels intensities
124 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
125 figure 1. Note that all the scales/levels above in the tree are only used as calculus and parameters estimation tools and
126 none of their nodes are associated with PET/CT voxels values.

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

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

138 1.1 Hidden Markov Tree model

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

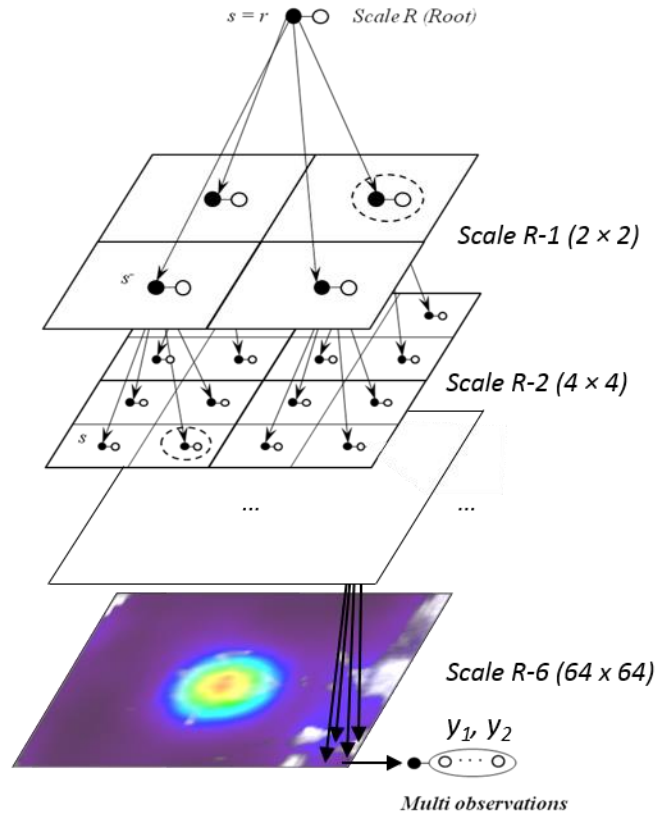
144 The HMT model is defined as follows: let S be a finite set of points and $X = (X_s)_{s \in S}$, $Y = (Y_s)_{s \in S_0}$ two random vectors
 145 indexed respectively on S and $S_0 \subset S$ subset of observed voxels. Each X_s takes its values in the finite set of classes
 146 $\Omega = \{\omega_1, \dots, \omega_k\}$ (in this work tumor and background) and Y_s takes its values in the set of observations (real values
 147 corresponding to voxels' intensities). Let S^1, \dots, S^R be a partition of S representing different "generations". Each $s \in S^i$
 148 admits $s^+ \subset S^{i+1}$ (called his "children") in such a way that every element of $t \in S^{i+1}$ has a unique "father" $t^- \in S^i$ such
 149 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
 150 distribution of (X, Y) , the random vector (X, Y) is a HMT if:

$$151 \quad p(x, y) = p(x_r) p(y_r | x_r) \prod_{j=2}^R \prod_{s \in S^j} p(x_s | x_{s^-}) p(y_s | x_s), \quad (1)$$

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

153

154



155

156 Figure 1: Hidden Markov Tree (HMT) Model with the associated PET/CT image (selection centered on the tumor):
 157 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
 158 associated with a vector of 2 values (y_1 and y_2) corresponding to the PET and CT intensities. The classification as
 159 “tumor” or “background” is $x = (x_s)_{s \in S_0}$.

160

161 1.2 Observation modeling

162 With respect to image segmentation in a statistical setting, a Gaussian distribution is sufficient to model PET voxels’
 163 intensities and more complex distributions through the use of the Pearson’s system (containing 7 different
 164 distributions)³¹ does not improve accuracy³². However, in the present case the framework has to conjointly model PET
 165 and CT distributions that could be very different, hence calling for a more flexible statistical model. Therefore the
 166 observation's noise densities $p(y_s | x_s)$ (for $s \in S_0$) considered are of the generalized normal distribution form³³.

167 1.3 HMT parameters estimation

168 The parameters of the HMT that need to be estimated are listed in table 1. The table also lists how their values are
 169 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 <i>a priori</i>
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		
	Standard deviation of distribution for each class		
	Shape parameter (generalized normal distribution) for each class	The distribution of each class is a generalized normal distribution.	

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) algorithm³⁴ which is a stochastic version of the EM algorithm³⁴. The EM algorithm has been adapted to HMT²⁵, 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 algorithm³⁶ 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 algorithm³⁷ to avoid underflow issues. More details can be found in²⁵.

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-frequency (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

192 use of images rebuilt after pre-processing in the spatio-frequency domain has been suggested recently for PET image
 193 segmentation³⁸. We thus decided to investigate the potential benefit of such an approach within our context of PET/CT
 194 segmentation, considering the contourlets in addition to the wavelets.

195 2.1 Wavelets pre-processing

196 The wavelet transform (WT) is established as a powerful tool for statistical signal and image processing. In this work,
 197 we used the isotropic undecimated wavelet transform (IUWT) based on the ‘à trous’ algorithm³⁹. The non decimation
 198 avoids the pseudo Gibbs phenomenon. The “à trous” algorithm decomposes an image $I_0(x,y,z)$ into a list of wavelet
 199 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
 200 through N successive filtering with a 3D low-pass filter h^5 related to a scaling function ϕ_{3D} which must satisfy a number
 201 of properties (compactly supported, regularity, symmetry, convergence in the Fourier space and derivability). B-spline
 202 interpolations were chosen in this work. At each iteration j , the approximation image I_{j+1} is given according to equation
 203 (2).

$$204 \quad I_{j+1}(x, y, z) = \sum_{m,n,o \in [-2;2]} h(m, n, o) I_j(x + m2^j, y + n2^j, z + o2^j) \quad (2)$$

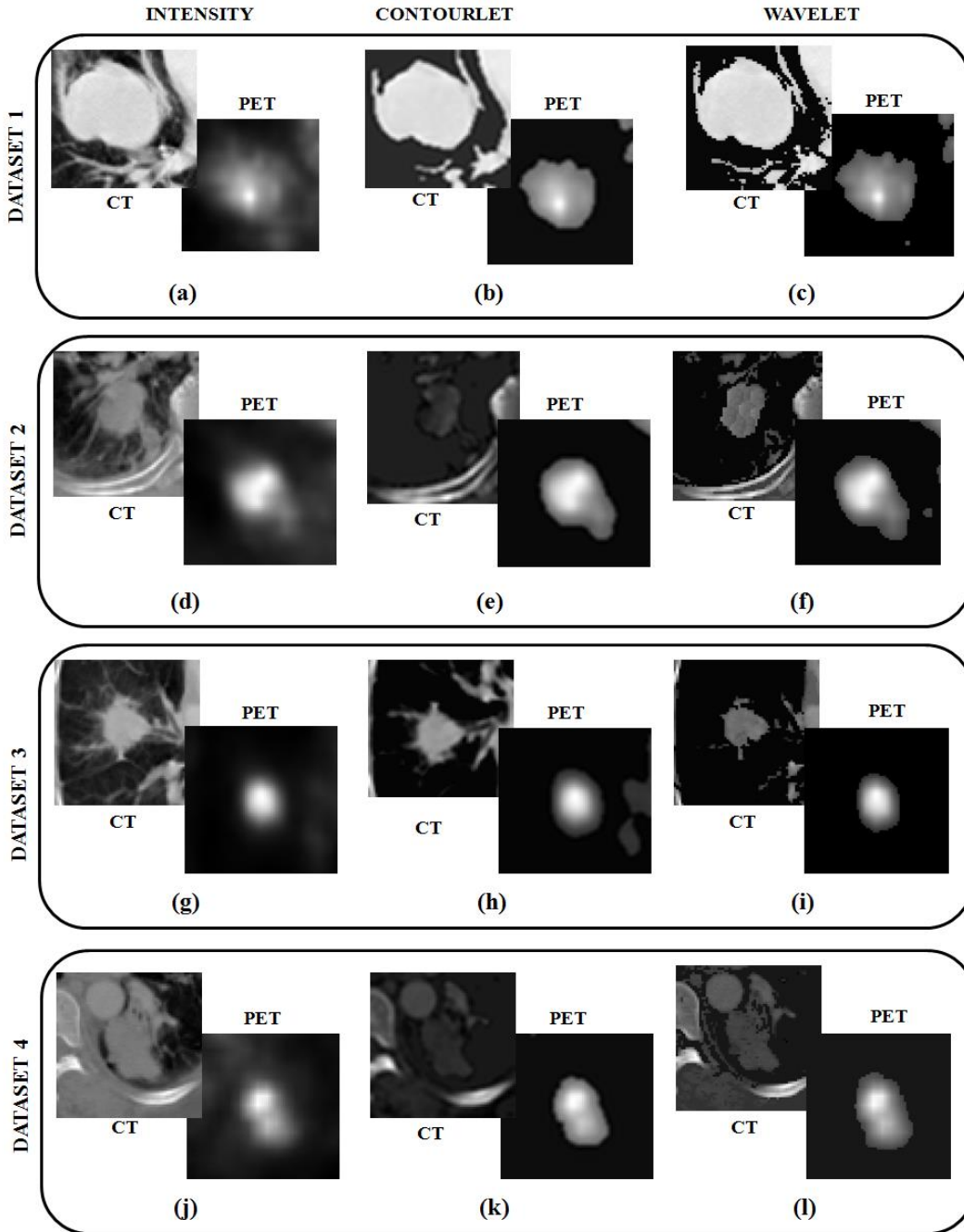
205 The wavelet planes w_{j+1} are defined as the differences between two consecutive approximations I_j and I_{j+1} containing
 206 the details at a resolution level between I_j and I_{j+1} in such a way that the initial image $I_0(x,y)$ can be rebuilt from its
 207 details layers as in equation (3).

$$208 \quad I_0(x, y, z) = I_N(x, y, z) + \sum_{j=1}^{j=N} w_j(x, y, z) \quad (3)$$

209 2.2 Contourlets pre-processing

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

217 are retained whereas the rest are set to 0. Second, the image is rebuilt using the selected coefficients. PET and CT
 218 images and their wavelet or contourlet-based pre-processing results are illustrated in figure 2. In a previous work³⁸, the
 219 value M was arbitrarily set at 200 for PET. In our case we determined these values empirically by exploring values
 220 between 100 and 3000. The best performance was reached for M=800 for PET images and M=2000 for CT images.



221

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.

222

223 3. Evaluation study

3.1 Simulated dataset

224 We have been advocating for a long time the use of realistic simulated datasets with voxel-based ground-truth as part of a rigorous
225 validation framework for PET segmentation^{42, 43} and have contributed in making such datasets available to the community^{44, 45}.
226 However these past efforts were focused on the PET component and did not include CT simulated images. For the present work
227 in order to investigate the complementary value of PET/CT vs. CT-only or PET-only HMT segmentation, a digital phantom of a
228 lung tumor located close to the mediastinum and the chest wall was designed and both a PET image and CT image were simulated.
229 No breathing was simulated. The simulated PET was generated with GATE (Geant4 Application for Tomography Emission) and
230 reconstructed following the same process as described previously^{44, 45}. PET spatial resolution was ~4mm and voxel size of the
231 reconstructed PET image was $4 \times 4 \times 4$ mm³. The CT simulated image was generated using the same digital phantom used for the
232 PET simulation. The CT simulation is based on a Monte-Carlo simulation where X-rays are created and tracked individually
233 through the voxelized phantom using the Siddon projector⁴⁶ and attenuated following the Beer-Lambert Law. Compton and
234 Rayleigh scattering were not computed. The simulation was performed using 360 projections around the phantom (1 projection
235 per degree), and 5×10^9 particles were created to achieve ~10000 counts/voxel in the histogram. The source aperture was 20 degrees,
236 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
237 final reconstructed image. A line detector composed of 750 pixel elements and 1 mm pixel size, was placed behind the phantom.
238 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
239 simulated detector was (350, 0, 0) mm. Finally, the simulated histogram was reconstructed using a maximum likelihood for
240 transmission tomography (MLTR) algorithm based on a gradient-ascent method⁴⁷. For this study, the CT data was reconstructed
241 using 5 iterations and 72 subsets, 1 mm voxel size, and the final image was blurred with 1 mm Gaussian filter. As explained in
242 section 1, the reconstructed PET image was up-sampled to be of same size as the CT dataset to be associated within the HMT
243 framework. Figure 3 shows the digital phantom and the associated simulated PET and CT images.

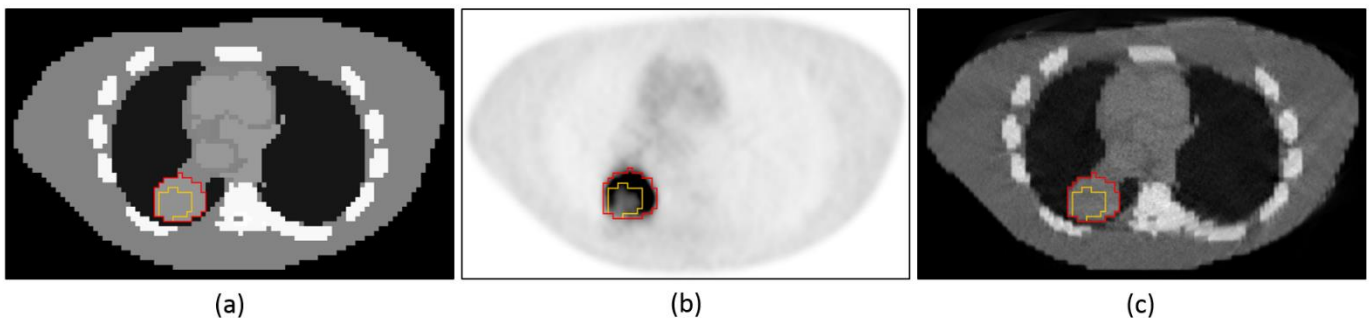


Figure 3. (a) Digital phantom with complete tumor ground-truth contour in red and sub-volume in orange, with a mean value of -5HU and 3 SUV. The rest of the tumor has a mean value of 20 HU and 12 SUV. Background uptake in the lung was set at 0.8 SUV whereas HU were set at -850. Bones were set at 700 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

244 The proposed methodology was evaluated on PET/CT clinical images. Fifteen patients with confirmed non-small cell
245 lung cancer (NSCLC), stage Ib–IIIb, for which no obvious spatial mismatch between the CT and the PET images (due
246 to respiratory motion and/or positioning differences between the two acquisitions) was observed, were analyzed. All
247 patients underwent an ^{18}F -FDG PET/CT examination for staging purposes before treatment. Patients were instructed
248 to fast for a minimum of 6 h before examination. Free-breathing PET and CT images were acquired 45–60 min after
249 ^{18}F -FDG injection. A total of seven 5-min bed positions with overlap were used for whole-body PET (Biograph
250 PET/CT; Siemens) acquisitions, which were corrected for attenuation using the CT data and iteratively reconstructed
251 using the ordered-subsets expectation maximization algorithm (4 iterations, 8 subsets). The noise levels in PET and
252 CT correspond to standard PET/CT clinical acquisitions. Voxel sizes of PET and CT datasets were $5.31 \times 5.31 \times 5 \text{ mm}^3$
253 and $0.98 \times 0.98 \times 5 \text{ mm}^3$ respectively. Spatial resolution of the PET scanner is estimated at 4.5 mm full-width at half-
254 maximum (FWHM) in the center of the field of view, whereas spatial resolution of the CT scanner is estimated below
255 1 mm FWHM. The PET and CT images were assumed to be co-registered at reconstruction. The PET images were up-
256 sampled using a cubic B-spline interpolation scheme⁴⁸ so that it corresponds to the dimension of the CT image.

3.3 Surrogate of truth

257 In the simulated case, the ground-truth of the phantom was used. In the clinical dataset, all patients underwent surgery
258 and the maximum diameter was measured on the surgical specimen, as previously described⁴⁹. In the absence of other
259 surrogate of truth such as histopathological volumes, the validation of the approach consisted in comparing automatic
260 segmentation of the tumor with a statistical consensus of manual delineations performed by three different experts on
261 each slice of the fused PET/CT images. The experts were asked to draw contours with the goal of defining a GTV for
262 radiotherapy planning. The consensus was obtained using the simultaneous truth and performance level estimation
263 (STAPLE) algorithm⁵⁰, an expectation-maximization algorithm that computes a probabilistic estimate of the true

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

3.4 PET, CT and PET/CT comparison

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

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

3.5 Other methods for comparison

276 Segmentation results were also generated using the standard fuzzy c-means (FCM) algorithm and the fuzzy local
277 information c-means (FLICM)⁵¹ for comparison purposes. FLICM is an FCM implementation adding a weighted norm
278 to account for outliers due to the noise and uses two parameters: a regularization parameter and the size of the
279 surrounding kernel. We used the recommended parameters values (regularization equal to 1 and kernel radius equal to
280 3 voxels)⁵¹. Alternative values for these two parameters were explored without obtaining improved performances.
281 Both methods were implemented as multimodal, *i.e.* using as input vectors containing the intensities of both up-
282 sampled PET and CT images. The input regions (cropped from the entire image) were the same for all methods under
283 comparison.

3.6 Evaluation metrics and figures of merit

284 The performance of segmentation was assessed using Dice similarity coefficient (DSC), sensitivity (SE) and positive
285 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
286 between volumes V_T and V_S and size (V) the size of volume V .

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

$$289 \quad SE = \frac{\text{size}(V_T \cap V_S)}{\text{size}(V_T)} \quad (4)$$

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

$$292 \quad PPV = \frac{\text{size}(V_T \cap V_S)}{\text{size}(V_S)} \quad (5)$$

293
294 The combination of a high PPV and a low SE indicates an under-evaluation of the true volume, whereas the inverse
295 indicates an over-evaluation of the true volume.

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

$$299 \quad DSC = \frac{2 \times \text{size}(V_T \cap V_S)}{\text{size}(V_T + V_S)} \quad (6)$$

300

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

303

304 **RESULTS**

305 ***Validity of the manual consensus***

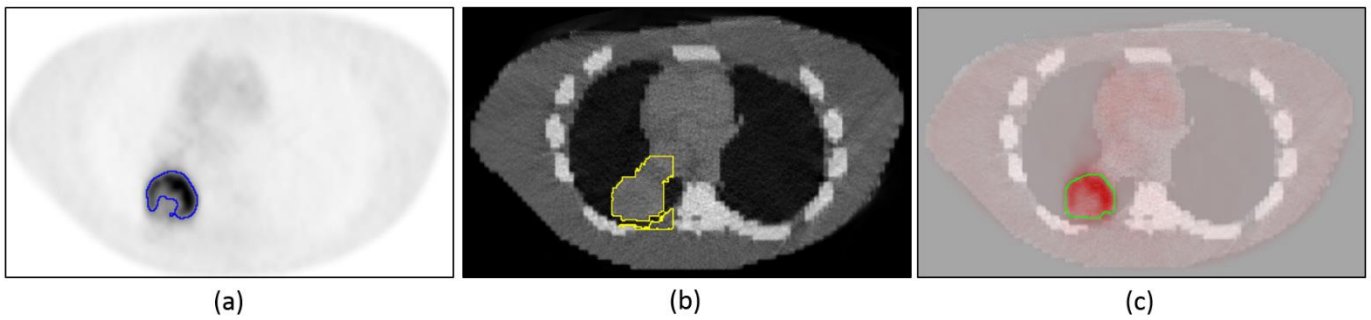
306 The consensus of three manual delineations led to volumes with a maximum diameter very close to the histopathology
307 measurement, with a correlation of 0.91, and differences of $9 \pm 7\%$, which is in line with previous observations on the

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

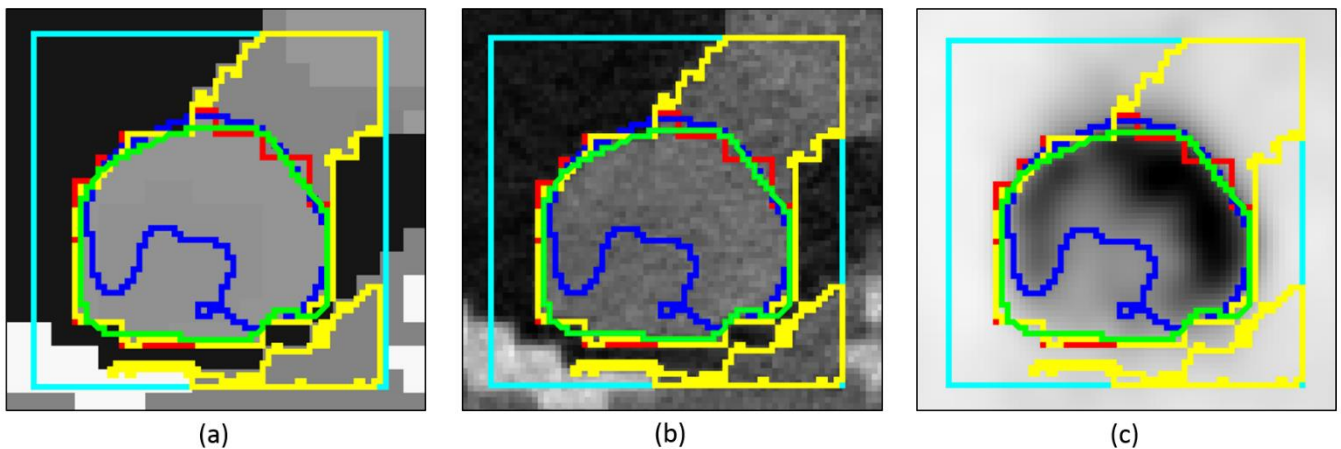
310 ***HMT segmentation results***

311 ***Simulated dataset***

312 Figures 4 and 5 show the IHMT segmentation results considering only the PET, only the CT, or the joint PET/CT. The
313 PET-only segmentation led to very high PPV (0.96) but low SE (0.65) (DSC of 0.77), due to the heterogeneous uptake
314 of the tumor, although with no issue differentiating the rest of the tumor uptake from the surrounding normal tissues
315 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
316 the lack of contrast between tumor and surrounding normal soft tissues. The PET/CT combination led to a better
317 coverage of the tumor with high SE (0.90) and PPV (0.97) (DSC of 0.93).



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



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

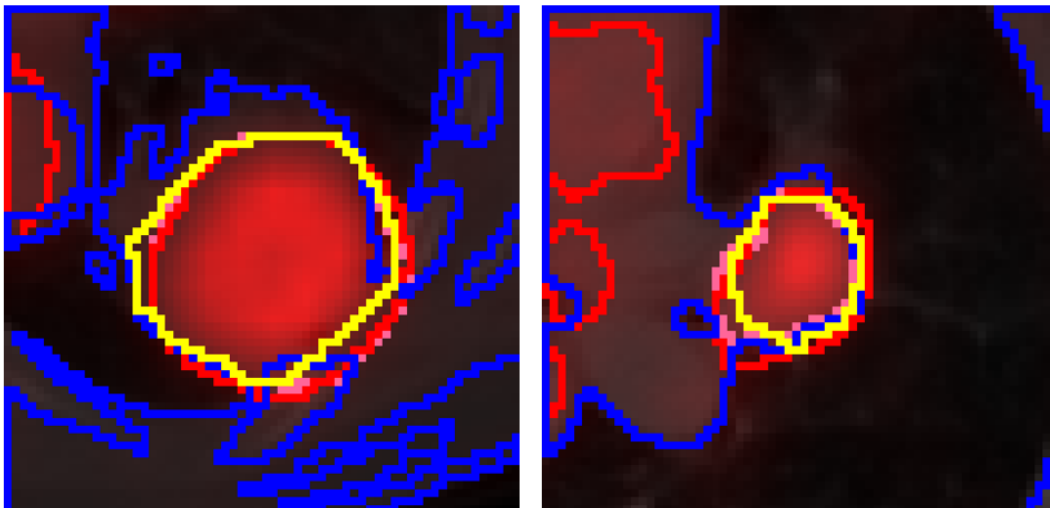
325 ***Clinical datasets***

326 Table 2 summarizes the DSC calculated respectively across the 15 cases, comparing IHMT applied to either PET or
 327 CT only, or on PET/CT. The HMT on PET/CT obtained the best accuracy with DSC of 0.89 ± 0.13 compared to
 328 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
 329 same hierarchy, with slightly better DSC, and a small advantage for CHMT over WHMT. The differences between
 330 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

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

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

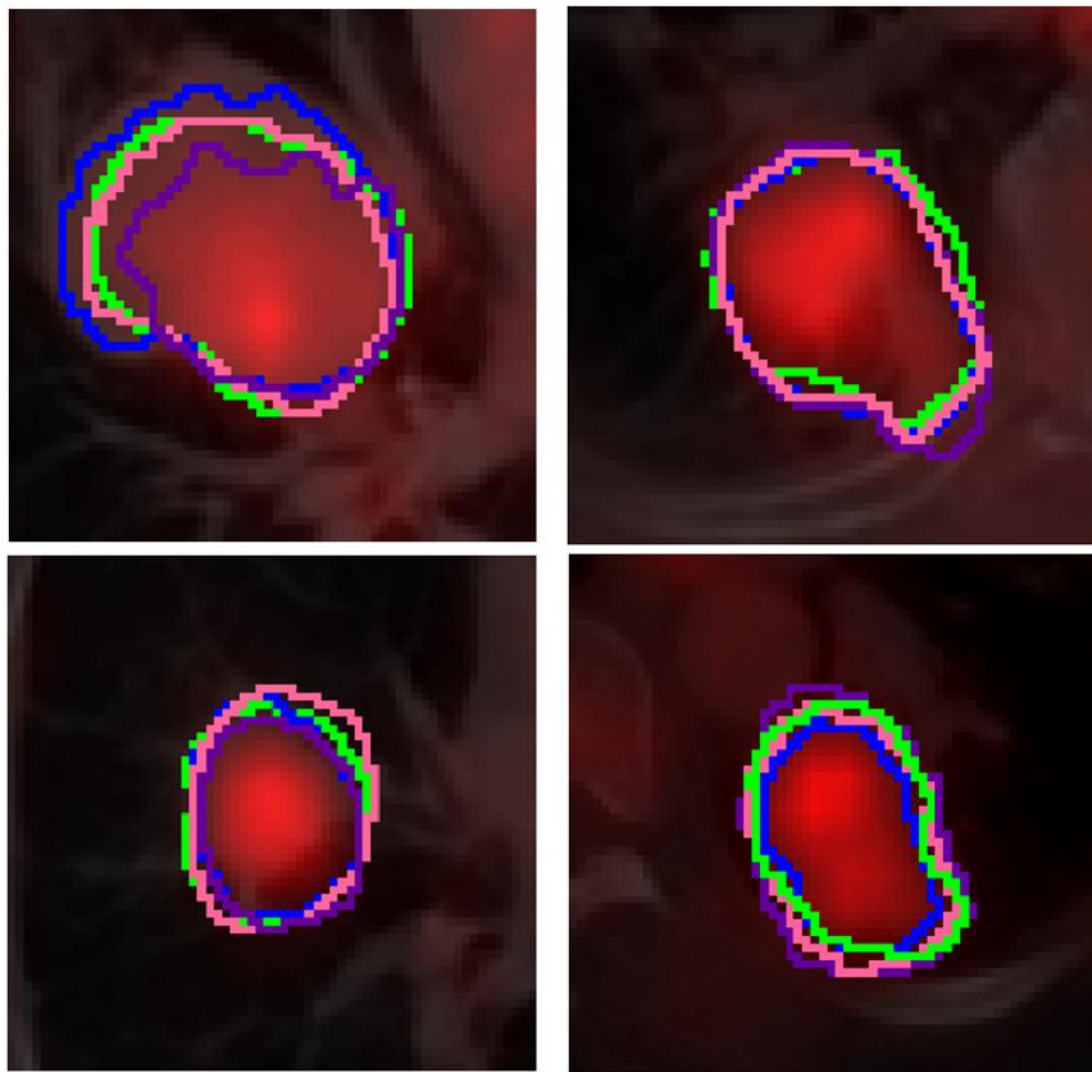


335

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

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

341



342

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

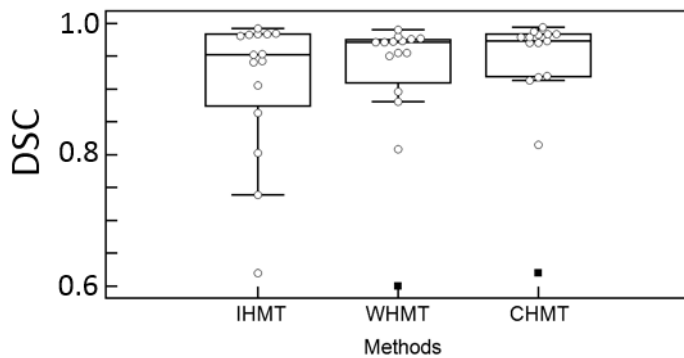
345 Quantitative results of all figures of merit calculated across the 15 cases are summarized in table 3, for IHMT, WHMT
346 CHMT, FCM and FLICM. The three HMT methods led to an accurate tumor segmentation, characterized by high PPV,

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

352

	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

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



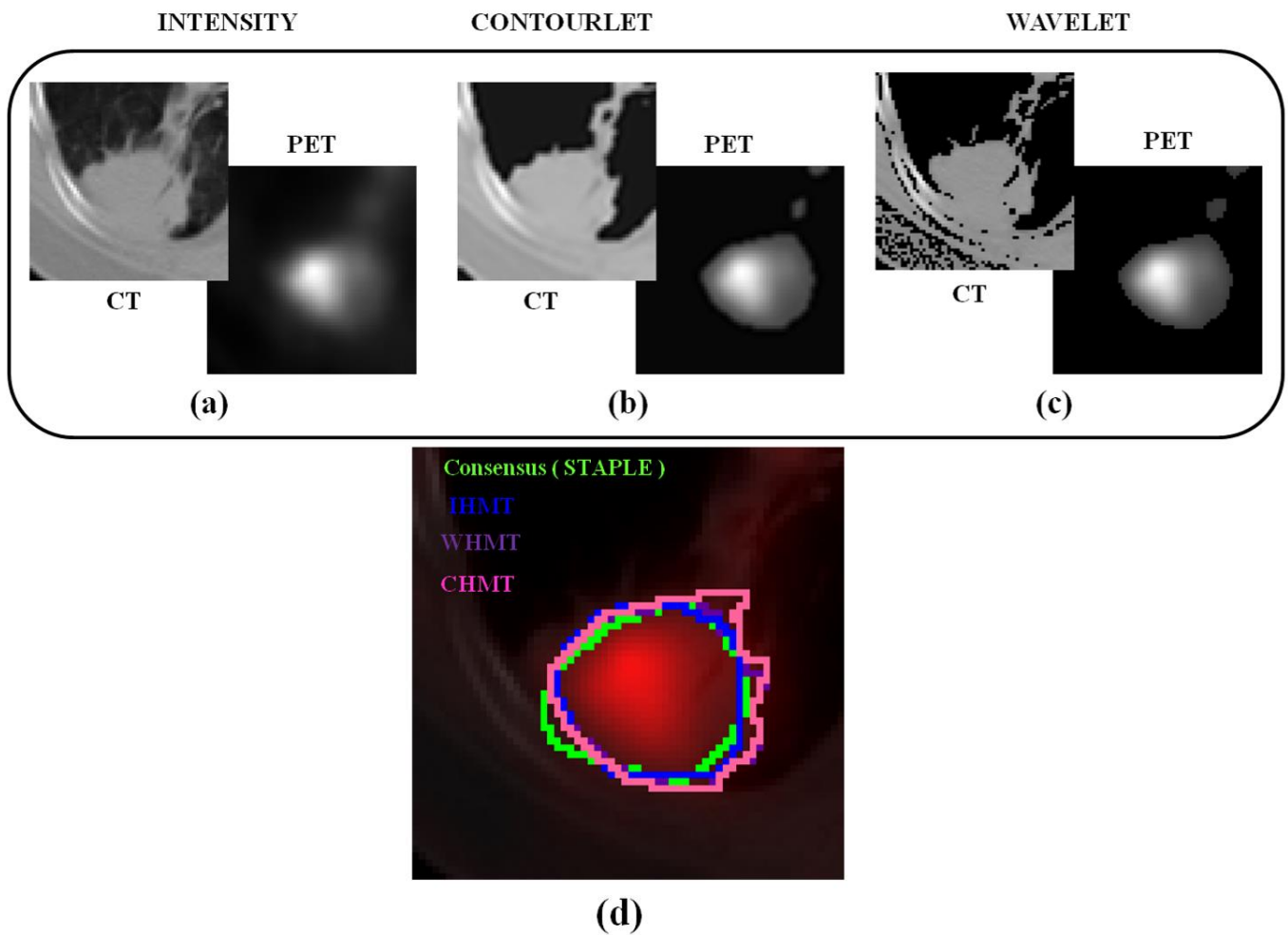
355

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

359

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

361 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
362 respectively. These results may be influenced by the tumor's neighborhood to the ribs and soft tissues in the lung.



363

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.

364

365 DISCUSSION

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

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

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

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

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

397 Another framework for the segmentation of multimodal PET/CT/MRI images models functional and anatomical data
398 in a product lattice and applies simultaneous delineations of tumor regions based on a random walk algorithm¹². Results
399 in term of DSC were compared for PET-only, CT-only, MRI-only, PET-CT, PET-MRI and PET-CT-MRI
400 segmentations, with mean values of 0.832, 0.878, 0.882, 0.914, 0.931 and 0.934 respectively, demonstrating increasing
401 accuracy with each additional image modality being exploited by the algorithm. Our results comparing segmentations
402 using either PET or CT only, or PET/CT, are in line with these observations, higher accuracy being obtained when
403 exploiting both modalities.

404 Another approach used a HMF based segmentation of PET and CT image separately with a regularized term penalizing
405 the segmentation difference between the two modalities¹⁹. PET/CT segmentation led to an accuracy of 0.86 ± 0.05 in
406 terms of DSC which is comparable to our results, with however likely lower computational efficiency: HMT ensures
407 better computational properties and faster estimation of parameters than HMF¹⁷. A similar approach was evaluated by
408 comparing PET-only, CT-only and the association of PET and CT images through optimization of a HMF based
409 segmentation with the graph cut algorithm, where each node is associated with a pair of corresponding voxels in PET
410 and CT images¹³. More importantly, the method allows obtaining different segmentations on the PET and the CT
411 through this co-segmentation framework. This original approach was compared with a standard fused PET/CT
412 segmentation, the surrogate of truth being a manual delineation by a radiation oncologist on CT images with the
413 guidance of PET images. The proposed method combining PET and CT led to a much higher accuracy than PET-only
414 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
415 CT-only). However, the improvement of the proposed approach compared to a segmentation on fused PET/CT,
416 comparable to our approach, (DSC of 0.79 ± 0.08) was not significant.

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

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

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

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

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

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

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

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

471

472 **CONCLUSIONS**

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

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

481

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

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

486

- 488 ¹ D. De Ruyscher, J. Belderbos, B. Reymen, W. van Elmpt, A. van Baardwijk, R. Wanders, F. Hoebbers, M. Vooijs, M. Ollers,
489 and P. Lambin, “State of the art radiation therapy for lung cancer 2012: a glimpse of the future,” *Clin. Lung Cancer* **14**(2),
490 89–95 (2013).
- 491 ² F.E. Turkheimer, N. Boussion, A.N. Anderson, N. Pavese, P. Piccini, and D. Visvikis, “PET image denoising using a
492 synergistic multiresolution analysis of structural (MRI/CT) and functional datasets,” *J. Nucl. Med. Off. Publ. Soc. Nucl. Med.*
493 **49**(4), 657–666 (2008).
- 494 ³ P. Novosad and A.J. Reader, “MR-guided dynamic PET reconstruction with the kernel method and spectral temporal basis
495 functions,” *Phys. Med. Biol.* **61**(12), 4624–4644 (2016).
- 496 ⁴ C. Coello, F. Willoch, P. Selnes, L. Gjerstad, T. Fladby, and A. Skretting, “Correction of partial volume effect in (18)F-FDG
497 PET brain studies using coregistered MR volumes: voxel based analysis of tracer uptake in the white matter,” *NeuroImage*
498 **72**, 183–192 (2013).
- 499 ⁵ A. Le Pogam, M. Hatt, P. Descourt, N. Boussion, C. Tsoumpas, F.E. Turkheimer, C. Prunier-Aesch, J.-L. Baulieu, D.
500 Guilloteau, and D. Visvikis, “Evaluation of a 3D local multiresolution algorithm for the correction of partial volume effects
501 in positron emission tomography,” *Med. Phys.* **38**(9), 4920–4923 (2011).
- 502 ⁶ A. Bousse, S. Pedemonte, B.A. Thomas, K. Erlandsson, S. Ourselin, S. Arridge, and B.F. Hutton, “Markov random field and
503 Gaussian mixture for segmented MRI-based partial volume correction in PET,” *Phys. Med. Biol.* **57**(20), 6681–6705 (2012).
- 504 ⁷ S. David, D. Visvikis, G. Quellec, C.C. Le Rest, P. Fernandez, M. Allard, C. Roux, and M. Hatt, “Image change detection
505 using paradoxical theory for patient follow-up quantitation and therapy assessment,” *IEEE Trans. Med. Imaging* **31**(9), 1743–
506 1753 (2012).
- 507 ⁸ H. Necib, C. Garcia, A. Wagner, B. Vanderlinden, P. Emonts, A. Hendlisz, P. Flamen, and I. Buvat, “Detection and
508 characterization of tumor changes in 18F-FDG PET patient monitoring using parametric imaging,” *J Nucl Med* **52**(3), 354–
509 61 (2011).
- 510 ⁹ B. Lelandais, S. Ruan, T. Denœux, P. Vera, and I. Gardin, “Fusion of multi-tracer PET images for dose painting,” *Med.*
511 *Image Anal.* **18**(7), 1247–1259 (2014).
- 512 ¹⁰ H. Yu, C. Caldwell, K. Mah, I. Poon, J. Balogh, R. MacKenzie, N. Khaouam, and R. Tirona, “Automated radiation targeting
513 in head-and-neck cancer using region-based texture analysis of PET and CT images,” *Int. J. Radiat. Oncol. Biol. Phys.* **75**(2),
514 618–625 (2009).
- 515 ¹¹ I. El Naqa, D. Yang, A. Apte, D. Khullar, S. Mutic, J. Zheng, J.D. Bradley, P. Grigsby, and J.O. Deasy, “Concurrent
516 multimodality image segmentation by active contours for radiotherapy treatment planning,” *Med Phys* **34**(12), 4738–49
517 (2007).
- 518 ¹² U. Bagci, J.K. Udupa, N. Mendhiratta, B. Foster, Z. Xu, J. Yao, X. Chen, and D.J. Mollura, “Joint segmentation of anatomical
519 and functional images: Applications in quantification of lesions from PET, PET-CT, MRI-PET, and MRI-PET-CT images,”
520 *Med. Image Anal.* **17**(8), 929–945 (2013).
- 521 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
522 tumor in PET-CT images with context information,” *IEEE Trans. Med. Imaging* **32**(9), 1685–1697 (2013).
- 523 ¹⁴ D. Markel, C. Caldwell, H. Alasti, H. Soliman, Y. Ung, J. Lee, and A. Sun, “Automatic Segmentation of Lung Carcinoma
524 Using 3D Texture Features in 18-FDG PET/CT,” *Int. J. Mol. Imaging* **2013**, 980769 (2013).
- 525 ¹⁵ H. Cui, X. Wang, J. Zhou, S. Eberl, Y. Yin, D. Feng, and M. Fulham, “Topology polymorphism graph for lung tumor
526 segmentation in PET-CT images,” *Phys. Med. Biol.* **60**(12), 4893–4914 (2015).
- 527 ¹⁶ D. Markel, H. Zaidi, and I. El Naqa, “Novel multimodality segmentation using level sets and Jensen-Rényi divergence,” *Med.*
528 *Phys.* **40**(12), 121908 (2013).
- 529 ¹⁷ W. Pieczynski, “Modèles de Markov en traitement d’images,” *Trait. Signal* **20**(30), 255–278 (2003).
- 530 ¹⁸ D.W. Montgomery, A. Amira, and H. Zaidi, “Fully automated segmentation of oncological PET volumes using a combined
531 multiscale and statistical model,” *Med Phys* **34**(2), 722–36 (2007).
- 532 ¹⁹ D. Han, J. Bayouth, Q. Song, A. Taurani, M. Sonka, J. Buatti, and X. Wu, “Globally Optimal Tumor Segmentation in PET-
533 CT Images: A Graph-based Co-segmentation Method,” in *Proc. 22Nd Int. Conf. Inf. Process. Med. Imaging* (Springer-Verlag,
534 Berlin, Heidelberg, 2011), pp. 245–256.
- 535 ²⁰ M. Hatt, F. Lamare, N. Boussion, A. Turzo, C. Collet, F. Salzenstein, C. Roux, P. Jarritt, K. Carson, C. Cheze-Le Rest, and
536 D. Visvikis, “Fuzzy hidden Markov chains segmentation for volume determination and quantitation in PET,” *Phys Med Biol*
537 **52**(12), 3467–91 (2007).
- 538 ²¹ A.S. Willsky, “Multiresolution Markov models for signal and image processing,” *Proc. IEEE* **90**(8), 1396–1458 (2002).
- 539 ²² G. Guénard, P. Legendre, and P. Peres-Neto, “Phylogenetic eigenvector maps: a framework to model and predict species
540 traits,” *Methods Ecol. Evol.* **4**(12), 1120–1131 (2013).
- 541 ²³ J. Porway, Q. Wang, and S.C. Zhu, “A Hierarchical and Contextual Model for Aerial Image Parsing,” *Int. J. Comput. Vis.*
542 **88**(2), 254–283 (2010).
- 543 ²⁴ F. Flitti, C. Collet, B. Vollmer, and F. Bonnarel, “Multiband Segmentation of a Spectroscopic Line Data Cube: Application
544 to the HI Data Cube of the Spiral Galaxy NGC 4254,” *EURASIP J Appl Signal Process* **2005**, 2546–2558 (2005).

545 25 J.M. Laferté, P. Pérez, and F. Heitz, “Discrete Markov image modeling and inference on the quadtree,” *IEEE Trans. Image*
546 *Process. Publ. IEEE Signal Process. Soc.* **9**(3), 390–404 (2000).

547 26 F. Li, X. Jia, D. Fraser, and A. Lambert, “Super Resolution for Remote Sensing Images Based on a Universal Hidden Markov
548 Tree Model,” *IEEE Trans. Geosci. Remote Sens.* **48**(3), 1270–1278 (2010).

549 27 C. Li and Q. Hao, “Functional MR image statistical restoration for neural activity detection using Hidden Markov Tree
550 model,” *Int. J. Comput. Biol. Drug Des.* **6**(3), 190–209 (2013).

551 28 J. Wu and A.C.S. Chung, “A novel framework for segmentation of deep brain structures based on Markov dependence tree,”
552 *NeuroImage* **46**(4), 1027–1036 (2009).

553 29 M. Wels, Y. Zheng, M. Huber, J. Hornegger, and D. Comaniciu, “A discriminative model-constrained EM approach to 3D
554 MRI brain tissue classification and intensity non-uniformity correction,” *Phys. Med. Biol.* **56**(11), 3269–3300 (2011).

555 30 H. Hanzouli, J. Lapuyade-Lahorgue, E. Monfrini, G. Delso, W. Pieczynski, D. Visvikis, and M. Hatt, “PET/CT image
556 denoising and segmentation based on a multi observation and a multi scale Markov tree model,” in *2013 IEEE Nucl. Sci.*
557 *Symp. Med. Imaging Conf. NSSMIC*(2013), pp. 1–4.

558 31 Y. Delignon, A. Marzouki, and W. Pieczynski, “Estimation of generalized mixtures and its application in image
559 segmentation,” *IEEE Trans Image Process* **6**(10), 1364–75 (1997).

560 32 M. Hatt, C. Cheze le Rest, A. Turzo, C. Roux, and D. Visvikis, “A fuzzy locally adaptive Bayesian segmentation approach
561 for volume determination in PET,” *IEEE Trans Med Imaging* **28**(6), 881–93 (2009).

562 33 T.K. Pogány and S. Nadarajah, “On the characteristic function of the generalized normal distribution,” *Comptes Rendus*
563 *Math.* **348**(3–4), 203–206 (2010).

564 34 G. Celeux and J. Diebolt, “A stochastic approximation type EM algorithm for the mixture problem,” (1991).

565 35 J.-B. Durand and P. Gonçalves, “Statistical Inference for Hidden Markov Tree Models and Application to Wavelet Trees,”
566 (2001).

567 36 L. Baum, “{An inequality and associated maximization technique in statistical estimation for probabilistic functions of
568 Markov processes},” *Inequalities* **3**, 1–8 (1972).

569 37 P.A. Devijver, “Baum’s forward-backward algorithm revisited,” *Pattern Recognit. Lett.* **3**(6), 369–373 (1985).

570 38 M. Abdoli, R.A.J.O. Dierckx, and H. Zaidi, “Contourlet-based active contour model for PET image segmentation,” *Med.*
571 *Phys.* **40**(8), 082507 (2013).

572 39 P. Dutilleul, “An Implementation of the ‘algorithme à trous’ to Compute the Wavelet Transform,” in *Wavelets*, edited by
573 P.J.-M. Combes, P.A. Grossmann and P.P. Tchamitchian (Springer Berlin Heidelberg, 1989), pp. 298–304.

574 40 M.N. Do and M. Vetterli, “The contourlet transform: an efficient directional multiresolution image representation,” *IEEE*
575 *Trans. Image Process.* **14**(12), 2091–2106 (2005).

576 41 A.L. da Cunha, J. Zhou, and M.N. Do, “The Nonsampled Contourlet Transform: Theory, Design, and Applications,”
577 *IEEE Trans. Image Process.* **15**(10), 3089–3101 (2006).

578 42 M. Hatt, J. Lee, C.R. Schmidlein, I. El Naqa, C. Caldwell, E. De Bernardi, W. Lu, S. Das, X. Geets, V. Gregoire, R. Jeraj,
579 M.P. MacManus, O.R. Mawlawi, U. Nestle, A.B. Pugachev, H. Schöder, T. Shepherd, E. Spezi, D. Visvikis, H. Zaidi, and
580 A.S. Kirov, “Classification and evaluation strategies of auto-segmentation approaches for PET: Report of AAPM Task Group
581 No. 211,” *Med. Phys.* (2017).

582 43 M. Hatt, C. Cheze le Rest, P. Descourt, A. Dekker, D. De Ruyscher, M. Oellers, P. Lambin, O. Pradier, and D. Visvikis,
583 “Accurate automatic delineation of heterogeneous functional volumes in positron emission tomography for oncology
584 applications,” *Int J Radiat Oncol Biol Phys* **77**(1), 301–8 (2010).

585 44 A. Le Maitre, W. Segars, S. Marache, A. Reilhac, M. Hatt, S. Tomei, C. Lartizien, and D. Visvikis, “Incorporating Patient-
586 Specific Variability in the Simulation of Realistic Whole-Body 18F-FDG Distributions for Oncology Applications,” *Proc.*
587 *IEEE* **9**(12), 2026–2038 (2009).

588 45 P. Papadimitroulas, G. Loudos, A. Le Maitre, M. Hatt, F. Tixier, N. Efthimiou, G.C. Nikiforidis, D. Visvikis, and G.C.
589 Kagadis, “Investigation of realistic PET simulations incorporating tumor patient’s specificity using anthropomorphic models:
590 creation of an oncology database,” *Med. Phys.* **40**(11), 112506 (2013).

591 46 R.L. Siddon, “Fast calculation of the exact radiological path for a three-dimensional CT array,” *Med. Phys.* **12**(2), 252–255
592 (1985).

593 47 K. Van Slambrouck and J. Nuyts, “Reconstruction Scheme for Accelerated Maximum Likelihood Reconstruction: The
594 Patchwork Structure,” *IEEE Trans. Nucl. Sci.* **61**(1), 173–181 (2014).

595 48 P. Thevenaz, T. Blu, and M. Unser, “Interpolation revisited,” *IEEE Trans Med Imaging* **19**(7), 739–58 (2000).

596 49 A. van Baardwijk, G. Bosmans, L. Boersma, J. Buijsen, S. Wanders, M. Hochstenbag, R.-J. van Suylen, A. Dekker, C.
597 Dehing-Oberije, R. Houben, S.M. Bentzen, M. van Kroonenburgh, P. Lambin, and D. De Ruyscher, “PET-CT-based auto-
598 contouring in non-small-cell lung cancer correlates with pathology and reduces interobserver variability in the delineation of
599 the primary tumor and involved nodal volumes,” *Int. J. Radiat. Oncol. Biol. Phys.* **68**(3), 771–778 (2007).

600 50 S.K. Warfield, K.H. Zou, and W.M. Wells, “Simultaneous truth and performance level estimation (STAPLE): an algorithm
601 for the validation of image segmentation,” *IEEE Trans. Med. Imaging* **23**(7), 903–921 (2004).

602 51 S. Krinidis and V. Chatzis, “A robust fuzzy local information C-Means clustering algorithm,” *IEEE Trans. Image Process.*
603 *Publ. IEEE Signal Process. Soc.* **19**(5), 1328–1337 (2010).

604 52 L.R. Dice, “Measures of the Amount of Ecologic Association Between Species,” *Ecology* **26**(3), 297 (1945).

605 53 T.F. Heston, "Standardizing predictive values in diagnostic imaging research," *J. Magn. Reson. Imaging JMRI* **33**(2), 505;
606 author reply 506-507 (2011).

607 54 R. Parikh, A. Mathai, S. Parikh, G. Chandra Sekhar, and R. Thomas, "Understanding and using sensitivity, specificity and
608 predictive values," *Indian J. Ophthalmol.* **56**(1), 45–50 (2008).

609 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
610 manual contouring: which is the better method for target delineation using 18F-FDG PET/CT in non-small cell lung cancer?,"
611 *J. Nucl. Med. Off. Publ. Soc. Nucl. Med.* **51**(10), 1517–1523 (2010).

612 56 M.P. Mac Manus and R.J. Hicks, "The role of positron emission tomography/computed tomography in radiation therapy
613 planning for patients with lung cancer," *Semin. Nucl. Med.* **42**(5), 308–319 (2012).

614 57 J. van Loon, A. van Baardwijk, L. Boersma, M. Ollers, P. Lambin, and D. De Ruyscher, "Therapeutic implications of
615 molecular imaging with PET in the combined modality treatment of lung cancer," *Cancer Treat. Rev.* **37**(5), 331–343 (2011).

616 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,
617 and D. Visvikis, "Reproducibility of 18F-FDG and 3'-Deoxy-3'-18F-Fluorothymidine PET Tumor Volume Measurements,"
618 *J. Nucl. Med.* **51**(9), 1368–1376 (2010).

619 59 M. Hatt, C. Cheze-le Rest, A. van Baardwijk, P. Lambin, O. Pradier, and D. Visvikis, "Impact of tumor size and tracer uptake
620 heterogeneity in (18)F-FDG PET and CT non-small cell lung cancer tumor delineation," *J Nucl Med* **52**(11), 1690–7 (2011).

621 60 J. Buijsen, J. van den Bogaard, H. van der Weide, S. Engelsman, R. van Stiphout, M. Janssen, G. Beets, R. Beets-Tan, P.
622 Lambin, and G. Lammering, "FDG-PET-CT reduces the interobserver variability in rectal tumor delineation," *Radiother*
623 *Oncol* **102**(3), 371–6 (2012).

624 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.
625 Lambin, and G. Lammering, "FDG-PET provides the best correlation with the tumor specimen compared to MRI and CT in
626 rectal cancer," *Radiother Oncol* **98**(2), 270–6 (2011).

627 62 F. Lamare, M.J. Ledesma Carbayo, T. Cresson, G. Kontaxakis, A. Santos, C.C. Le Rest, A.J. Reader, and D. Visvikis, "List-
628 mode-based reconstruction for respiratory motion correction in PET using non-rigid body transformations," *Phys Med Biol*
629 **52**(17), 5187–204 (2007).

630 63 J. Orban de Xivry, G. Janssens, G. Bosmans, M. De Craene, A. Dekker, J. Buijsen, A. van Baardwijk, D. De Ruyscher, B.
631 Macq, and P. Lambin, "Tumour delineation and cumulative dose computation in radiotherapy based on deformable
632 registration of respiratory correlated CT images of lung cancer patients," *Radiother Oncol* **85**(2), 232–8 (2007).

633 64 X. Geets, J.A. Lee, A. Bol, M. Lonnew, and V. Gregoire, "A gradient-based method for segmenting FDG-PET images:
634 methodology and validation," *Eur J Nucl Med Mol Imaging* **34**(9), 1427–38 (2007).

635 65 E. Monfrini, J. Lecomte, F. Desbouvries, and W. Pieczynski, "Image and signal restoration using pairwise Markov trees," in
636 *2003 IEEE Workshop Stat. Signal Process.*(2003), pp. 174–177.

637 66 W. Pieczynski, "Arbres de Markov Triplet et fusion de Dempster-Shafer," *Comptes Rendus Math.* **336**(10), 869–872 (2003).

638 67 P. Lanchantin and W. Pieczynski, "Evidential Markov chains and trees with applications to non stationary processes
639 segmentation," *Trait. Signal* **22**(1), 15–26 (2005).

640

641

642 Appendix

643

644 Annex

645

646 Let S_0, S_1, \dots, S_N be successive scales, with $S_0 = \{r\}$ containing just the root, S_1 containing four children of
 647 r , and so on, as showed in Figure 1. Let $S = S_0 \cup S_1 \cup \dots \cup S_N$ be the set of all nodes. For $s \in S_n$, with $0 \leq n < N$,
 648 we will denote with s^+ the four children of s (which are in S_{n+1}), and for $s \in S_n$, with $0 < n \leq N$ we will denote
 649 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
 650 the paper is defined by

651

$$652 \quad p(x, y) = [p(x_r) \prod_{n=0}^{N-1} \prod_{s_n \in S_n} \prod_{s_{n+1} \in s_n^+} p(x_{s_{n+1}} | x_{s_n})] \prod_{s_N \in S_N} p(y_{s_N} | x_{s_N}), \quad (1)$$

653

654 $p(y_{s_N} | x_{s_N})$ are generalized Gaussian distributions modeling PET/CT images $y_{s_N} = (y_{s_N}^{PET}, y_{s_N}^{CT})$. Then the problem is
 655 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-
 656 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
 657 $p(x_s = tumor | y) > p(x_s = notumor | y)$, and “no tumor” if not. Setting ω_1 as “tumor” and ω_2 as “no tumor” the co-
 658 segmentation $y = y_{s_N} = (y_{s_N}^{PET}, y_{s_N}^{CT}) \rightarrow \hat{x}_{s_N}$ is thus given with

659

$$660 \quad \text{For each node } s \in S_N : \hat{x}_s = \begin{cases} \omega_1 & \text{if } p(x_s = \omega_1 | y) > p(x_s = \omega_2 | y) \\ \omega_2 & \text{if } p(x_s = \omega_1 | y) < p(x_s = \omega_2 | y) \end{cases} \quad (2)$$

661

662 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.
 663 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
 664 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:

665

$$666 \quad \beta^n(x_s) = \prod_{u \in s^+} [\sum_{x_u \in \Omega} \beta^{n+1}(x_u) p(x_u | x_s)], \quad (3)$$

667

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

669 The core point is that the distribution of the whole tree X_S conditional on both observed images

670 $Y_{S_N} = y = (y_{s_N}^{PET}, y_{s_N}^{CT})$, denoted as p^y , also is a Markov tree distribution, with $p^y(x_r)$ and transitions $p^y(x_s | x_{s^-})$
 671 given by

672

$$673 \quad p^y(x_r) = \frac{p(x_r) \beta^1(x_s)}{\sum_{x_r \in \Omega} p(x_r) \beta^1(x_s)}, \quad p^y(x_s | x_{s^-}) = \frac{\beta^{n+1}(x_s) p(x_s | x_{s^-})}{\sum_{x_s \in \Omega} \beta^{n+1}(x_s) p(x_s | x_{s^-})} \quad (4)$$

674 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
 675 following Forward – Down procedure.

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

679
680
$$p^y(x_{s_1}) = p^y(x_r)$$
 given with (4); for $n = 1, \dots, N - 1$:
$$p^y(x_{s_{n+1}}) = \sum_{x_{s_n} \in \Omega} p^y(x_{s_n}) p^y(x_{s_{n+1}} | x_{s_n})$$
 (5)

681
682 Finally the co-segmentation algorithm runs as follows:
683

- 684 (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);
685 (2) For each $s \in S$ and $x_s \in \Omega$ compute $p^y(x_r)$, $p^y(x_s | x_{s^-})$ with (4);
686 (3) For each $s \in S_N$ compute $p(x_s | y_{s_N})$ with the Forward – Down procedure (5);
687 (4) Perform co-segmentation using (2).