Fuzzy hidden Markov chains segmentation for volume determination and quantitation in PET.

Mathieu Hatt, Frédéric Lamare, Nicolas Boussion, Christian Roux, Alexandre Turzo, Catherine Cheze-Lerest, P. Jarritt, K. Carson, Fabien Salzenstein, Christophe Collet, et al.

Fuzzy hidden Markov chains segmentation for volume determination and quantitation in PET

M. Hatt[1], F. Lamare[1], N. Boussion[1], A. Turzo[1,2], C. Collet[3], F. Salzenstein[4], C. Roux[5,1], P. Jarritt[6], K. Carson[6], C. Cheze-Le Rest[1,2], D. Visvikis[1].

[1] INSERM, U650, LaTIM, Brest, F-29200 France.

[2] Academic Department of Nuclear Medicine, CHU Morvan, Brest, F-29609 France

[3] Ecole Nationale Supérieure de Physique de Strasbourg (ENSPS), ULP, Strasbourg, F-67000 France.

[4] Institut d'Électronique du Solide et des Systèmes (InESS), ULP, Strasbourg, F-67000 France.

[5] ENST Bretagne, GET-ENST, Brest, F-29200 France

[6] Medical Physics Agency, Royal Victoria Hospital, Belfast, Northern Ireland.

Corresponding author:

Mathieu HATT,
INSERM U650,
Laboratoire du Traitement de l'Information Médicale (LaTIM)
CHU MORVAN,
Bat 2bis (I3S),
5 avenue Foch,
Brest, 29609
France

Tel: +33 2 98 01 81 11
Fax: +33 2 98 01 81 24

ABSTRACT

Accurate volume of interest (VOI) estimation in PET is crucial in different oncology applications such as response to therapy evaluation and radiotherapy treatment planning. The objective of our study was to evaluate the performance of the proposed algorithm for automatic lesion volume delineation; namely the Fuzzy Hidden Markov Chains (FHMC), with that of current state of the art in clinical practice threshold based techniques. As the classical Hidden Markov Chain (HMC) algorithm, FHMC takes into account noise, voxel's intensity and spatial correlation, in order to classify a voxel as background or functional VOI. However the novelty of the fuzzy model consists of the inclusion of an estimation of imprecision, which should subsequently lead to a better modelling of the "fuzzy" nature of the object on interest boundaries in emission tomography data. The performance of the algorithms has been assessed on both simulated and acquired datasets of the IEC phantom, covering a large range of spherical lesion sizes (from 10 to 37mm), contrast ratios (4:1 and 8:1) and image noise levels. Both lesion activity recovery and VOI determination tasks were assessed in reconstructed images using two different voxel sizes ($8mm^3$ and $64mm^3$). In order to account for both the functional volume location and its size, the concept of % classification errors was introduced in the evaluation of volume segmentation using the simulated datasets. Results reveal that FHMC performs substantially better than the threshold based methodology for functional volume determination or activity concentration recovery considering a contrast ratio of 4:1 and lesion sizes of <28mm. Furthermore differences between classification and volume estimation errors evaluated were smaller for the segmented volumes provided by the FHMC algorithm. Finally, the performance of the automatic algorithms was less susceptible to image noise levels in comparison to the threshold based techniques. The analysis of both simulated and acquired datasets led to similar results and conclusions as far as the performance of segmentation algorithms under evaluation is concerned.

1. Introduction

Positron Emission Tomography (PET) has been long established as a powerful tool in oncology, particularly in the area of diagnosis. However, alternative applications such as the use of PET in radiotherapy planning (Jarritt et al 2006) and response to therapy studies (Krak et al 2005) are rapidly gaining ground. Whereas accurate activity concentration recovery is crucial for correct diagnosis and monitoring response to therapy, applications such as the use of PET in Intensity-Modulated Radiation Therapy (IMRT) treatment planning renders equally vital the accurate shape and volume determination of lesions. Different volume of interest (VOI) determination methodologies have been proposed that can be classified as manual or automatic. On the one hand, manual segmentation methods to delineate boundaries are laborious and highly subjective (Krak et al 2005). On the other hand, automatic segmentation of objects of interest in PET (Reutter et al 1997, Zhu et al 2003, Kim et al 2003, Riddell et al 1999) is not a trivial task because of low spatial resolution and resulting partial volume effects (PVE), low contrast ratios, as well as noise resulting from the statistical nature of radioactive decay or the choice of the reconstruction process.

The most widely used method to semi-automatically determine VOIs in PET is thresholding, either adaptive, using *a priori* Computed Tomography (CT) knowledge (Erdi et al 1997), or fixed threshold (Krak et al 2005) using values derived from phantom studies (from 30 to 75% of maximum local activity concentration value) (Jarritt et al 2006, Krak et al 2005, Erdi et al 1997). Such thresholding techniques however lead to variable VOIs determination as shown in recent clinical studies (Nestle et al 2005). On the other hand, numerous works have addressed automatic lesion detection from PET datasets, including different methodologies such as edge detection (Reutter et al 1997), fuzzy C-Means (Zhu et al 2003), clustering (Kim et al 2003) or watersheds (Riddell et al 1999). The performance of these algorithms is sensitive to variations of noise intensity and/or lesion contrast. In addition, past work has in its majority considered the ability of such automatic methodologies for the detection of lesions but not the accuracy with which the methods are capable for VOI and/or activity concentration determination. Furthermore, all of the afore-mentioned algorithms often involve user-dependent initializations, pre- and post-processing, or additional information like CT or expert

knowledge; rendering their employment more complicated and the outcome dependent on choices made by the user in relation to the pre- and/or post-processing steps necessary. For example in the case of the watershed algorithm a filtering pass as pre-processing step to smooth the image, and a post-processing step to fuse the different regions resulting from the algorithm are necessary.

Hidden Markov Models are automatic segmentation algorithms allowing noise modelling and have proven to be less sensitive to variation of the values in the regions of the images than other segmentation approaches thanks to their statistical modelling. They have only been previously used in PET in the form of Hidden Markov Fields (HMF) (Chen et al 2001). Hidden Markov Chains (HMC) (Benmiloud et al 1995) is a faster model and can offer competitive results (Salzenstein et al 1998). Furthermore, HMC leads to shorter computational times, as quantities of interest can be computed directly on the chain, whereas the HMF algorithm needs iterative Monte-Carlo like estimation procedures (Salzenstein et al 1998) that are time consuming. These algorithms offer an unsupervised estimation of the parameters needed for the image segmentation and limit the user's input to the number of classes to be searched for in the image. Reconstructed images require no further pre- or post-processing treatment (such as for example filtering) prior to the segmentation process. Instead, image noise is considered as additional information (a parameter in the classification decision process) to be taken into account, not to be suppressed or avoided.

The objectives of our study were to (a). develop a new fuzzy HMC (FHMC) model in an attempt to account for the limited spatial resolution in PET and (b) compare the performance of FHMC with these of the thresholding methodologies currently used in clinical practice. Different imaging conditions in terms of statistical quality, as well as lesion size and source-to-background (S/B) ratio were considered in this study. The analysis was carried out on both simulated and acquired images reconstructed using iterative algorithms which form today's state of the art in whole body PET imaging in routine clinical oncology practice (Visvikis et al 2001, Visvikis et al 2004).

2. Materials and methods

2.1 Hard and Fuzzy Hidden Markov Chain models

The HMC model is an unsupervised methodology that takes place in the Bayesian framework. Although we place ourselves in the application of image segmentation this methodology can be used in other applications such as for example speech recognition (Dai 1994). Let T be a finite set corresponding to the voxels of an image. We consider two random processes $\mathrm{Y} = (y_t)_{t \in T}$ and $\mathrm{X} = (x_t)_{t \in T}$. Y represents the observed image, and X represents the "hidden" segmentation map. X takes its values in $\Omega = \{1, ..., K\}$ with K being the number of classes c, and Y takes its values in $\mathbb{R}$. We assume that a Markov process can model the prior distribution of X. The segmentation problem consists in estimating the hidden X from the available noisy observation Y. The relationship between X and Y can be modelled by the joint distribution $P(\mathrm{X},\mathrm{Y})$. This distribution can be obtained thanks to the Bayes formula:

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

$P(\mathrm{Y}|\mathrm{X})$ is the likelihood of the observation Y conditionally with respect to the hidden ground-truth X, and $P(\mathrm{X})$ is the prior knowledge concerning X. The Bayes rule allows us to know the posterior distribution of X with respect to the observation Y. In the Markov Chain framework we have to assume the random variables $\mathrm{Y} = (y_t)_{t \in T}$ are conditionally independent with respect to X and that the distribution of each y_t conditional on X is equal to its distribution conditional on x_t. Many applications of Hidden Markov Models with unsupervised estimation have been successful considering different types of images (radar, sonar, Magnetic Resonance Images (MRI), CT, satellite or astronomical) (Pieczynski 2003, Salzenstein et al 2004, Delignon 1997), but this kind of approach was almost never applied on PET data.

2.1.1 Markov Chain definition

X is a Markov chain if:

$$P(x_t \mid x_1, ..., x_{t-1}) = P(x_t \mid x_{t-1}) \text{ for } 1 < t \leq T \qquad (2)$$

The distribution of X is then defined by the distribution of x_1, called initial probabilities $init(c)$ for each class c ($P(x_1 = c)$) and the transition matrix $trans(c,d)$ (of dimension $K \times K$) containing the

probabilities of transitions from the class c to the class d ; $P(x_{t+1} = d \mid x_t = c)$. As X and Y are one-dimensional elements in the HMC context, a spatial transformation is necessary to process three-dimensional VOIs. For the best preservation of the spatial correlation between voxels we use the Hilbert-Peano space-filling curve. This fractal path can be extended to explore 3D VOIs (Kamata et al 1999). A visual illustration of the Hilbert-Peano path for a 4×4×4 voxels 3D VOI is given in figure 1. Once the chain has been segmented, the inverse path is used to reconstruct the 3D segmentation map.

2.1.2 Adding a fuzzy measure to the model

The general idea behind the implementation of a fuzzy model within the Bayesian framework was previously introduced by Salzenstein (Salzenstein et al 1997). Its implementation in association with HMC developed as part of this work is based on the incorporation of a finite number of fuzzy levels F_i in combination with two homogeneous (or "hard") classes, in comparison to HMC where only a finite number of hard classes are considered. This model allows the coexistence of voxels belonging to one of two hard classes and voxels belonging to a "fuzzy level" depending on its membership to the two hard classes. Therefore, FHMC adds an estimation of imprecision of the hidden data (X, see section 2.1) in contrast to HMC which only models uncertainty of the observed data (Y, see section 2.1). The statistical part of the algorithm models the uncertainty of the classification, with the assumption being that the voxel is clearly identified but the observed data is noisy. On the other hand, the fuzzy part models the imprecision of the voxel's membership, with the assumption being that the voxel may contain both classes. One way to achieve this extension is to simultaneously use Dirac and Lesbegue measures at the class chain level. Hence we consider that X in the fuzzy model takes its values in $[0,1]$ instead of $\Omega = \{1,...,K\}$. Let δ_0 and δ_1 be the Dirac measures at 0 and 1, and ζ the Lesbegue measure on $]0,1[$. We define the new measure $\nu = \delta_0 + \delta_1 + \zeta$ on $[0,1]$. Note that, for example, using two hard classes and two fuzzy levels in the FHMC model is not equivalent to using four hard classes in the HMC model where $\nu = \delta_1 + \delta_2 + ... + \delta_K$. This has been previously stated using Markov Fields based segmentation (Salzenstein et al 1997).

The distribution of X can then be defined using a conjoint density g for (x_t, x_{t+1}) on $[0,1]\times[0,1]$:

$$\begin{aligned}
&\text{let } (a,b)\in[0,1]\times[0,1]\\
&g(a=0,b=0)=\alpha_1 \text{ and } g(a=1,b=1)=\alpha_2\\
&g(a=0,b=1)=\gamma_1 \text{ and } g(a=1,b=0)=\gamma_2\\
&g(a,b)=\beta f_g(a,b) \text{ if } (a,b)\neq\{(0,0),(0,1),(1,0),(1,1)\} \qquad (3)\\
&\text{With } \int_{[0,1]}\int_{[0,1]} g(a,b)d(\nu\otimes\nu)(a,b)=1 \text{ and } \alpha_1+\alpha_2+\gamma_1+\gamma_2+\beta\lambda=1 \qquad (4)
\end{aligned}$$

where, $d(\nu\otimes\nu)(a,b)$ is the notation for integration with respect to the (a,b) variables, each one being with respect to the measure ν on the interval [0,1]. λ is a constant depending on the form of the parameterized function f_g:

$$f_g(a,b)=1-|a-b| \qquad (5)$$

We now define the initial and transition probabilities ($init(c)$ and $trans(c,d)$) using the conjoint density g and an utility density h on $[0,1]$ defined by: $h(a)=\int_0^1 g(a,b)d\nu(b)$:

$init(c)$ using densities g and h :

$$\begin{aligned}
&P\left(x_1\in\{0,1\}\right)=\int_0^1 g(x_1,b)d\nu(b)=h(x_1)\\
&P\left(x_1\in F_i\right)=\int_{\frac{i-1}{N}}^{\frac{i}{N}}\int_0^1 g(a,b)d(\nu\otimes\nu)(a,b)\simeq\frac{1}{N}\int_0^1 g(\varepsilon_i,b)d(\nu)(b)=\frac{1}{N}h(\varepsilon_i)
\end{aligned} \qquad (6)$$

$trans(c,d)$ using the conditional density f deduced from (1) : $f(x_{t+1}\mid x_t)=\frac{g(x_{t+1},x_t)}{h(x_t)}$

$$\begin{aligned}
&P\left(x_{t+1}\in F_j\mid x_t\in F_i\right)=\frac{\int_{F_j}\int_{F_i} g(\varepsilon_j,\varepsilon_i)d(\nu\otimes\nu)(\varepsilon_j,\varepsilon_i)}{\int_{F_i}h(\varepsilon_i)d\nu(\varepsilon_i)}\\
&P\left(x_{t+1}\in F_j\mid x_t\in\{0,1\}\right)=\frac{\int_{F_j} g(\varepsilon_j,x_t)d\nu(\varepsilon_j)}{h(x_t)}\\
&P\left(x_{t+1}\in\{0,1\}\mid x_t\in F_i\right)=\frac{\int_{F_i} g(x_{t+1},\varepsilon_i)d\nu(\varepsilon_i)}{\int_{F_i}h(\varepsilon_i)d\nu(\varepsilon_i)}\\
&P\left(x_{t+1}\in\{0,1\}\mid x_t\in\{0,1\}\right)=\frac{g(x_{t+1},x_t)}{h(x_t)}
\end{aligned} \qquad (7)$$

where $N-1$ is the number of fuzzy levels and $\varepsilon_i=\frac{i}{N}$ is the value associated to a fuzzy level F_i.

The fuzzy model is a generalization of the hard model. The use of the Dirac measures allows one to retrieve the standard hard model when the fuzzy component is null. As the theoretical framework described above has not been developed for a specific kind of image, but as a general

segmentation algorithm, the *a priori* and the noise (also called observation) models are not directly derived from PET image characteristics. However this segmentation approach may be appropriate in segmenting PET images since they are both noisy and of low resolution. The "noise" aspect when considering Hidden Markov Models in general is the way the values of each class to be found in the image are distributed around a mean value. The noise model used, whose respective mean and variance are to be determined by the estimation steps, can therefore be adapted to image specific characteristics. On the other hand, the fuzzy measure allows a more realistic modelling of the objects' borders transitions between foreground and background, allowing in such a way to indirectly account for the effects of blurring (partial volume effects) associated with low resolution images, such as those in PET.

2.1.3 Segmentation and parameters estimation

In order to perform segmentation on the chain level, we need to use a criterion to classify each element as background or functional VOI. For this purpose we use the Marginal Posterior Mode (MPM) (Marroquin et al 1987). This approach aims to minimize the expectation $E\{L(x_t, \hat{x}_t) \mid Y\}$ where L is a loss (or cost) function:

$$L(x_t, \hat{x}_t) = |x_t - \hat{x}_t| \tag{8}$$

With x_t the real class and $\hat{x}_t = \hat{s}(y_t)$ the one affected by the segmentation process $\hat{s}$. This criterion is adequate for the segmentation problem as it penalizes a configuration with respect to the number of misclassified elements. In order to compute a solution, the MPM segmentation needs the parameters defining the *a priori* model (initial and transition probabilities of the chain) as well as the noisy observation data model (mean and variance of each class). The assumption that the noise for each class of the observed data can fit a Gaussian distribution was made as a first step. The mean and variance of each fuzzy level F_i is derived from the ones estimated in the two hard classes as follows:

$$\begin{aligned} \mu_{F_i} &= \mu_0(1-\varepsilon_i) + \varepsilon_i\mu_1 \\ {\sigma_{F_i}}^2 &= {\sigma_0}^2(1-\varepsilon_i)^2 + \varepsilon_i^2{\sigma_1}^2 \end{aligned} \tag{9}$$

Both *a priori* and noise models parameters are unknown in the real case and therefore they must be estimated. In order to achieve such estimation, we use the stochastic iterative procedure called Stochastic Expectation Maximization (SEM) (Celeux & Diebolt 1986), a stochastic version of the EM algorithm (Dempster 1977). This is achieved in a similar fashion to that used in the classical HMC context by simulating posterior realizations of X (see Appendix for detailed posterior realization of X and the SEM procedure) and computing empirical values of the parameters of interest using the simulated chain. The stochastic nature of this procedure makes it less sensitive to the initial guess of the parameters using the K-Means (McQueen 1967) than deterministic procedures like the EM algorithm. Both the MPM segmentation and SEM parameters estimation use a practical recursive computation of the values of interest called Forward-Backward procedure that is performed directly on the chain (Benmiloud et al 1995). The implementation of the FHMC segmentation algorithm in a step by step fashion can be found in the Appendix. Note that the overall algorithm is entirely unsupervised (except for the number of classes and fuzzy levels to use) and it is able to adjust to a large spectrum of image structures, noise or contrast. For example, no *a priori* is made on the shape of the objects to extract or the source-to-background ratio in the image.

2.2 Thresholding

Various thresholding methodologies have been proposed in the past for both functional volume segmentation and/or activity concentration recovery (Krak et al 2005, Erdi et al 1997, Nestle et al 2005). Thresholding using 42% and 50% of the maximum value in the lesion was chosen for VOI determination and quantitation purposes respectively, based on previous publications (Krak et al 2005, Erdi et al 1997). The methodology was implemented through region growing using the voxel of maximum intensity in the object of interest as a seed. Using a 3-D neighbourhood (26 neighbours) the region is iteratively increased by adding neighbouring voxels if their intensity is superior or equal to the selected threshold value. The results derived using these methods will be denoted from here onwards as T42 and T50 for the thresholds of 42% and 50% respectively.

2.3 Validation studies

2.3.1 Simulated and acquired datasets

Simulated datasets using the IEC image quality phantom (Jordan 1990), containing six different spherical lesions of 10, 13, 17, 22, 28 and 37 mm in diameter (figure 2), were generated using Geant4 Application for Tomographic Emission (GATE) and a validated model of the Philips Allegro PET scanner (Lamare et al 2006). Images, considering only the detected true coincidences, were subsequently reconstructed using the OPL-EM iterative algorithm (Reader et al 2002) with 7 iterations (Lamare et al 2006). Two different voxel sizes were considered in the reconstructed images; namely $2\times2\times2$ mm^3 and $4\times4\times4$ mm^3. The 8 mm^3 voxel size configuration leads to better sampled objects of interest but with higher noise due to the number of counts being divided by eight in each voxel in comparison to the 64 mm^3 voxel sizes. A uniform activity was simulated throughout the phantom cylinder and the lesions. Different parameters were however considered to cover a large spectrum of configurations allowing assessment of the influence of different parameters susceptible to affect the functional VOI determination or quantitation accuracy. The statistical quality of the images was varied by considering 20, 40 and 60 millions of true coincidences. Two different signal to background (S/B) ratios were also considered; 4:1 and 8:1 (with around 6kBq/cm^3 in the background, and 24 or 48kBq/cm^3 in the spheres respectively). Visual illustration of the reconstructed images corresponding to different simulated configurations is given in figure 3(a)-(d).

In addition to the simulated datasets, acquisitions of the IEC phantom were carried out in list mode format using a Philips GEMINI PET/CT scanner. The only difference with the simulated datasets was the exclusion of the 28 mm diameter sphere in the study because in the phantom used it was replaced by a plastic sphere of unknown diameter. The same S/B ratios of 4:1 and 8:1 used in the simulations were also employed in this part of the study, by introducing 7.4kBq/cm^3 in the background and 29.6 or 59.2kBq/cm^3 respectively in the spheres. Different count statistical qualities were obtained by reconstructing 1 min, 2min or 5min list-mode time frames using the 3D RAMLA algorithm, with specific parameters previously optimised (Visvikis et al 2004). The same voxel sizes as for the simulated datasets (8 mm^3 and 64 mm^3) were used in the reconstruction of each of the different statistical quality datasets considered. Visual illustration of the acquired images is given in figure 3(e)-

(h). Each sphere in both simulated and acquired images was isolated in a box of the same size (16×16×10 for the 4mm case, and 32×32×20 for the 2mm case) prior to the segmentation process.

2.3.2 Computed volume versus classification error measurement

The majority of previous works dealing with VOI determination in PET measure the performance of a given methodology by computing the VOI obtained on the segmentation map and comparing it with the true known volume of the object of interest. This type of approach has the potential of leading to biased performance measurements since a segmentation result may contain two different types of error. On the one hand, one may have voxels of the background that are classified as belonging to the object of interest, denoted from here on as positive classification errors (PCE), while on the other hand, one may end up with voxels of the object that are classified as belonging to the background, denoted from here on as negative classification errors (NCE). These classification errors essentially occur on the boundaries of the objects of interest because of "spill in" (increasing probabilities of a NCE) and "spill out" (increasing probabilities of a PCE). If the segmentation results in PCEs and NCEs of equal amounts, the computed VOI would be very close to the true known volume whereas the shape and position of the object would be incorrect. The shape and position information is as important as the total volume of the object in order to accurately derive a radiotherapy treatment planning or the activity concentration of interest in a response to therapy study based on the derived functional volume. For example, let us assume that the segmentation process results in 20% NCEs and 15% PCEs. This leads to a classification error of 35% whereas the error in the overall computed volume is only -5%. Hence, the use of classification error is a more pertinent measurement of the accuracy with which a given algorithm performs the task of functional volume delineation since it takes into account not only the segmented volume in comparison to the actual volume of interest but also its position and shape.

In the simulation study the total number of PCEs and NCEs is considered with respect to the number of voxels defining the sphere (VoS) in the digital phantom (the ground truth) in order to obtain a percentage classification error (CE):

$$CE = \frac{(PCE + NCE)}{VoS} \times 100 \qquad (10)$$

The size of classification errors can be bigger than 100% in the case where a large number of background voxels in the selected area of interest are misclassified as belonging to the sphere. In practical terms, maximum classification errors calculated during this work where limited to 200%, since any such values represent complete failure of the segmentation process. In addition, the interest of classification errors is when they occur at the borders of the objects and not in other regions of the background. One should also keep in mind that a combined representation of PCE and NCE into CE leads to a loss of information as far as the direction of the bias is concerned. It does however still represent more pertinent information than overall volume estimation errors, which reflect neither accurate magnitude nor direction of the bias for a segmented volume.

On the other hand in the case of the images reconstructed from the acquired datasets only overall computed volumes were considered in order to avoid any biases as a result of misalignment and rescaling inaccuracies, as well as reconstruction artefacts in the higher and lower slices of the associated CT datasets. As the goal is not to detect the lesion in the whole image but to estimate its volume, shape and position with the best accuracy possible, we assume that the lesion has been previously identified by the clinician and automatically or manually placed in a 3-D "box" well encompassing the object. Subsequently, the images of the selected area were segmented in two classes (*functional VOI* and *background*) using each of the three methods under evaluation (thresholding, FHMC and HMC). In the FHMC case, different numbers of fuzzy levels were considered in the segmentation process (namely 2 and 3). Following the segmentation by FHMC, volumes of interest can be defined using the hard classes and any number of the fuzzy levels considered.

2.3.3 Quantitation Accuracy

In terms of quantitation the objective of our study was to determine the accuracy of the average activity concentration recovered from a volume derived using a given segmentation algorithm. The "ground truth" for comparison purposes was established using the exact size, shape and location of each lesion (using the known digital phantom employed in the generation of the simulated datasets).

As a result, these recovered activity concentration values represented an under-estimation of the true activity due to PVE. A comparison on a lesion by lesion basis was subsequently carried out with the measured activity concentration from the segmented volumes obtained by the three algorithms considered. T50 should lead to some improvements in the lesion activity recovery with respect to T42 as a result of including less voxels in the volume used to compute the activity and therefore less voxels associated with PVE. Similarly FHMC 0/2 (see section 3. Results for the definition of FHMC x/y) should lead to concentration recovery improvements with respect to FHMC 1/2, since voxels belonging to the fuzzy levels are found at the edges of the lesions and their intensity is most significantly reduced by PVE. Therefore the inclusion of these voxels should only result in even stronger under-evaluation of the true lesion activity concentrations.

3. Results

Different segmentation maps obtained using each of the methods under evaluation are presented in figure 4 for a slice centred on the 28 mm sphere of the simulated images to visually illustrate the variations of the segmentation maps obtained. Figure 5(a) shows the impact of the number of fuzzy levels included in the FHMC segmentation. The various FHMC maps are denoted as FHMC x/y with x being the number of fuzzy levels included in the segmentation map, and y being the total number of fuzzy levels used in the segmentation process. The error bars in these figures represent different results obtained for each of the 3 different levels of statistical quality considered (the top of the error bar is the result concerning the worst statistical quality, the medium one concerns the medium quality, and the lowest one corresponds to the best quality considered). As Figure 5(a) shows, for the range of simulated spheres considered, no improvement was obtained in the % classification errors by having more than 2 fuzzy levels in the FHMC segmentation process and keeping in the overall segmented volume more than the voxels identified in the first fuzzy level. It should be emphasized at this point, that this conclusion was reached considering the results on the whole of the range of simulated sphere diameters and keeping in mind that our objective is determining a single best configuration of the algorithm parameters across a wide range of imaging conditions and not different parameters for individual lesion sizes, image statistics or contrast ratios. In addition, it is

clearly showed in figure 5 that HMC leads to worse segmentation results in comparison to FHMC for all different configurations considered. Therefore for all subsequent volume determination analyses, the results associated with the FHMC 1/2 versus T42 are presented. As shown in figure 5(b), no benefits are observed through the inclusion in the segmentation map of any voxels belonging to the fuzzy domain. This confirms what was anticipated in section 2.3.3. Therefore from here onwards all the quantitation results presented for FHMC have been calculated using only the hard class voxels resulting from the segmentation process (FHMC 0/2).

The % classification errors for reconstructed images of the simulated datasets as a function of lesion size and contrast are presented in figure 6(a) for 64 mm^3 and (b) for 8 mm^3, for the FHMC and the threshold based method (T42). A breakdown, in terms of PCEs and NCEs, of the % classification errors in figure 6(a) is given in figures 7(a)-(c) for the FHMC, HMC and T42 segmentation methods respectively. Finally, in order to facilitate a comparison of the segmentation results between the simulated and the acquired datasets, the % computed volume error is given in figures 8(a)-(b) for the same configurations as in figures 6(a)-(b).

Considering the simulated datasets, the introduction of FHMC led to superior results in comparison to the current "gold standard" in functional volume delineation of T42. FHMC segmentations led to <25% classification errors in computed volumes for lesions sizes >13mm irrespective of contrast ratio, level of noise or lesion size. Errors of more than 200% for FHMC were only observed for the 10mm sphere. Results for the T42 were more dependent on the lesion size, relative to FHMC results, varying from 10% to more than 200% (even for spheres up to 22mm in diameter for a contrast of 4:1 and 64mm^3 voxel size). However, the use of T42 was found to work well for lesion sizes of >17mm and a lesion to background ratio of 8:1 with % classification errors of 20-30%. On the other hand, for a lesion to background ratio of 4:1, the T42 threshold led to over 100% overestimation in the functional volume for lesions <28mm in diameter. As the errors bars in the different figures reveal, there was a larger dependence on the statistical quality of the reconstructed images observed with T42 in comparison to FHMC for the majority of the lesion sizes and contrast configurations considered. In particular this was true for all of the lesions for a contrast ratio of 4:1 and for lesions <22mm for a contrast ratio of 8:1. For example, for the 17mm sphere and a contrast

ratio of 8:1, T42 resulted in classification errors of 20 to 35% whereas FHMC classification errors from 15 to 17% were observed (fig. 6). On the other hand in the case of the 28mm sphere and a contrast ratio of 4:1, T42 errors were ranging from 85 to 110% whereas FHMC resulted in errors of 17-18%. The reduction in the reconstruction voxel size (from 64 mm^3 to 8 mm^3) led to small differences in the functional volumes determined using the FHMC segmentation algorithm, and although it led to improvements in the T42 based segmented volumes, the % classification errors remained at 80-200%. The trend observed with the standard voxel sizes on the variation of the segmentation results as a function of statistical quality was similar for the reduced voxel size images. For example in the case of the 22m sphere and a contrast ratio of 4:1 errors of 77-100% and 26-27% were observed for T42 and FHMC respectively. In general, the largest errors were observed for the smaller lesions of 10 and 13mm, where none of the segmentation algorithms considered performed well under any of the configurations tested, with errors largely >200%. As shown in figure 7(a) FHMC classification errors are essentially NCEs for the two biggest spheres and PCEs for the small ones. In contrast, as shown in figures 7(b)-(c), T42 and HMC methods result essentially in PCEs, apart from T42 in association with lesions >28mm in diameter and a lesion to background ratio of 8:1.

In terms of overall volume estimation errors on simulated datasets (see figures 8(a)-(b)) FHMC results on errors of up to 10% and between 10% and 20% for contrast ratio of 8:1 and 4:1 respectively, for lesions >13mm. T42 led to volume determination errors of <10% for lesions >17mm in diameter and a lesion to background ratio of 8:1, while errors of over 100% where observed for lesions <28mm with a lesion to background ratio of 4:1. However, while the lowest overall volume error of T42 was around 10%, the corresponding classification error was >20%. In the case of an 8 mm^3 reconstructed voxel size (figure 8(b)) small improvements were seen using the T42 for lesions ≥ 13mm and >22mm for a lesion to background ratio of 8:1 and 4:1 respectively. Finally, no noticeable differences were seen in the FHMC based segmentation results, apart from an improvement to <15% in the volume estimation error for the 13mm lesion with a contrast size of 8:1.

Figures 9(a)-(b) show the results in terms of % error in the recovered activity as a function of lesion size and contrast ratio considering the segmented volumes using 64 mm^3 and 8 mm^3 reconstructed voxel sizes. As it can be seen from this figure, FHMC and T50 led to the best results in

comparison to the "ground truth" throughout the different lesions sizes and contrasts evaluated, although T50 introduces larger errors in comparison to the "ground truth" for lesion sizes of <22mm and a contrast of 4:1. The use of the 8 mm^3 voxels does not alter the conclusions as far as the relationship between the results for the two methods evaluated is concerned, although in absolute terms all algorithms perform worse in comparison to the results obtained for 64 mm^3 voxels.

Considering the acquired datasets, figures 10(a)-(b) contains the results for the % overall lesion volume estimation for the 64 mm^3 and 8 mm^3 voxels, while figures 11(a)-(b) show the corresponding results for the activity quantitation errors. In terms of the volume estimation the general trends were similar to those observed for the simulated datasets, with the FHMC performing better than the T42 throughout the range of lesion sizes and contrasts evaluated. In absolute terms, the FHMC results were better particularly in the case of 8 mm^3 voxels where errors of <20% and 10% were seen for lesions >10mm and >22mm respectively. T42 errors were similar to FHMC for the 8:1 ratio and spheres >13mm but ranged from 20 to >100% for the 4:1 ratio configuration. A larger dependence to the statistical quality of the reconstructed images can be observed with the acquired datasets, demonstrating the more robust performance of the FHMC algorithm in comparison to the T42 methodology which was seen to be more affected by the images' statistical quality. Using again the example of the 22mm sphere (figure 10(a)), T42 errors were from 30 to 95% while FHMC errors were less than 5%. Although the variation of the FHMC results was higher for smaller spheres (10 and 13 mm), it was still smaller than in the case of the T42 results. For example, FHMC applied to the 13mm sphere with a 4:1 contrast ratio (figure 10(b)) resulted in errors between 5 and 30% whereas T42 errors ranged from 50 to 150%. Similar results between the FHMC and the T50 algorithms were globally seen in terms of the % accuracy of the recovered activity concentration, confirming the trends observed with the simulated datasets. Finally, similarly with the volume estimation, better results were seen with the 8 mm^3 reconstructed voxel's size for both the T50 and the FHMC leading to activity concentration estimation errors of between +10% and -10% for lesions >17mm in diameter.

4. Discussion

Although PET imaging applications are currently, in their majority, diagnostic and largely based on visual interpretation, there is increasing interest in applications such as the use of PET for radiotherapy treatment planning, as well as response to therapy and outcome prediction studies where accurate functional volume and concentration of activity estimation respectively are indispensable. Current state of the art methodologies for functional volume determination involve the use of adaptive thresholding based on anatomical information or phantom studies. The performance of these techniques is greatly dependent on lesion contrast and image noise characteristics and as this work has demonstrated can lead to variable performance. On the other hand, already proposed automatic segmentation methodologies have been mostly evaluated for use in lesion detection rather than lesion volume determination. In addition, their performance is highly dependent, similarly to the thresholding algorithms, on image contrast and noise characteristics.

Hidden Markov Chains is an automatic segmentation algorithm that allows noise modelling in the images but has also previously been evaluated for lesion detection rather than functional volume estimation. In the presented work a new algorithm (Fuzzy HMC) has been introduced and evaluated allowing the incorporation within Hidden Markov Chains of a finite number of fuzzy levels in combination with the "hard" classes considered in HMC, adding this way an estimation of imprecision that can account for the effects of limited spatial resolution in emission tomography images. During the evaluation of the FHMC, the inclusion of more than 2 fuzzy levels was found to not substantially alter the segmentation results, while only the inclusion of the voxels belonging to the first fuzzy level led to the most accurate results in terms of functional volume calculations throughout the range of configurations considered. Although it would be possible to consider the use of HMC with four hard classes and an additional rule to cluster the resulting segmentation map, the fuzzy nature of the borders leads to computation issues in transition probabilities that HMC is not able to deal with. Note that the significant addition of the fuzzy measure and mathematical changes in the model allows FHMC to take into account such a configuration, mainly due to the fact that one given voxel can contain both classes.

Finally, in this paper we have introduced the concept of classification errors rather than volume estimation errors in the evaluation of segmentation algorithms for volume determination tasks.

An evaluation based on classification errors is more robust since it does not simply consider the segmented volume but also its location with respect to the "ground truth" known in simulated datasets. Therefore, while the absolute segmented volume may be correct its location may be wrong, a fact that is as significant as the correct estimation of the overall functional volume particularly for applications such as the use of PET volumes in radiotherapy treatment planning.

In comparison to the recommended T42 for the accurate functional volume determination in PET (Krak et al 2005), the FHMC algorithm gave clearly superior results for lesions <28mm, in particular considering a lesion contrast of 4:1, where the T42 methodology completely failed to recover the functional volume. In addition, FHMC was more robust considering the different image statistical quality levels evaluated, while the results of T42 were greatly influenced by the level of noise present in the images. Differences between classification and volume estimation errors across the different configurations evaluated were smaller for the segmented volumes provided by the FHMC algorithm. In addition, the classification error results allows us to establish that the accuracy obtained on the estimated volumes using the FHMC algorithm is not by chance due to a similar levels of negative and positive classification errors. A smaller reconstructed voxel size at the same statistical quality level led to worse overall segmentation results, without altering the conclusions as far as the relative performance of the different algorithms evaluated is concerned.

The performance of the segmentation algorithms under evaluation in the reconstructed images of the acquired datasets was similar to that obtained with the simulated datasets in terms of volume estimation errors. The only difference observed in comparison with the simulated dataset results was an improvement in the overall segmentation results for 8 mm^3 reconstructed images in comparison to the 64 mm^3, which can be attributed to an associated adjustment of the optimised reconstruction parameters as a function of the voxel size in the 3D RAMLA algorithm used to reconstruct the acquired datasets.

As far as concentration recovery results are concerned, the performance of the different segmentation algorithms was compared in the simulated datasets to the recovered activity concentration considering the exact size and location of the simulated lesions. These results were influenced by the effects of partial volume leading to increasing underestimation of the activity

concentration with decreasing lesion size. Segmentation algorithms concentrate on accurate edge modelling in the object of interest and do not as such account for changes in the values of the voxels as a result of PVE. FHMC and the current "state of the art" threshold of 50% of the maximum lesion value (Krak et al 2005) led to similar results independently of the configurations evaluated, with absolute differences of 10-15% (due to an extra underestimation for the T50 results). Similar trends to those observed with the simulated datasets were obtained from the segmentation of the acquired images.

The presented results demonstrate the interest of FHMC over thresholding algorithms as the flexibility of the fuzzy levels choice may allow the use of the same segmentation map for different tasks, across a large range of lesion contrasts and sizes. FHMC through the addition of the fuzzy levels associated with each hard class is able to more accurately model the object of interest edges in reconstructed PET images. In addition, FHMC is clearly less susceptible to alterations in statistical image quality and lesion contrasts than other methodologies. This was observed on both images of simulated and acquired datasets. Having said that, none of the evaluated algorithms was successful in accurate volume estimation for lesion sizes of <17mm, considering typical PET image statistical qualities and reconstructed voxels of either $8mm^3$ or $64mm^3$. The main reason behind the failure of FHMC concerning the segmentation of such small lesions is the small number of voxels associated with the object of interest in combination to image noise levels, and the Hilbert-Peano path used to transform the image into a chain. The spatial correlation of such small objects may be lost once the image is transformed into a chain. A local model may be able to overcome such an issue (Hatt M et al 2007).

The results for FHMC may be further improved. Firstly, the direct estimation of the noise in the reconstructed images may lead to better results in comparison to the assumed Gaussian model used in this work to fit the distribution for each of the classes. Secondly, other *a priori* models may be used for Markovian modelling, like couple (Pieczynski et al 2004) or triplet (Lanchantin et al 2004) Markov Chains or Fields. These may be of interest considering a better modelling of the transitions between boundaries classes, as well as the non stationary nature of the hidden *a priori* model. In addition, the

fuzzy model may be extended to more than two hard classes to better model inhomogeneous or non spherical objects of interest.

5. Conclusion

A modified version of Hard Markov Chains segmentation algorithm has been developed by introducing a fuzzy measure (FHMC). Our results with both simulated and acquired datasets have shown that FHMC is more effective than these of the reference thresholding methodologies for both VOI determination and quantification in PET imaging. As part of the evaluation process, we have also introduced and assessed the interest of classification errors as a robust measurement of the performance of segmentation algorithms for VOI determination in contrast to a simple volume estimation which may introduce biases in terms of the segmented lesion location. Future developments will concentrate on the use of more than two "hard" classes in FHMC, which may more accurately account for the presence of inhomogeneous or non-spherical functional volumes, as well as an investigation into more adequate noise and *a priori* models.

APPENDIX. The FHMC algorithm step by step

For the calculation of the expressions a quantization of the interval $[0,1]$ into intervals $\left\{0,\frac{1}{N},\frac{2}{N}...,\frac{N-1}{N},1\right\}$ is used. For example with 2 fuzzy levels (or intervals) F_1, F_2, we have $N=3$ and there are $N-1=2$ fuzzy levels with $\varepsilon_i=\frac{i}{N}$: $\varepsilon_1=\frac{1}{3}$ and $\varepsilon_2=\frac{2}{3}$. Note : the symbol $\widetilde{....}$ denotes a density instead of a probability.

1. Transformation of the 2D or 3D image in a 1D chain using the Hilbert-Peano path (Kamata et al 1999) **(save the path to be used in step 5 of the procedure).**

From this point on, every step is performed on the image transformed into a chain.

2. Parameters initialization:

A priori model parameters:

$\left.\begin{array}{l}\alpha_1=\alpha_2=0.25\\ \gamma_1=\gamma_2=0\end{array}\right\}$ $init(c)$ and $trans(c,d)$ are computed according to (3), (4) and the following:

$$\lambda=\frac{2}{N}\left(\sum_{i=1}^{N-1}\left(1-\frac{i}{N}\right)+\sum_{i=1}^{N-1}\left(1-\left|\frac{i}{N}-1\right|\right)\right)+\frac{1}{N^2}\sum_{i=1}^{N-1}\sum_{j=1}^{N-1}\left(1-\left|\frac{i}{N}-\frac{j}{N}\right|\right)\quad \beta=\frac{1-(\alpha_1+\alpha_2+\gamma_1+\gamma_2)}{\lambda}$$

Initial and transition probabilities initializations can then be computed as follows:

$$init(0)=\alpha_1+\gamma_1+\frac{\beta}{N}\sum_{i=1}^{N-1}\left(1-\frac{i}{N}\right)$$

$$init(1)=\alpha_2+\gamma_2+\frac{\beta}{N}\sum_{i=1}^{N-1}\left(1-\left|1-\frac{i}{N}\right|\right)$$

$$init(\varepsilon_i)=\frac{\beta}{N}\left(\left(1-\varepsilon_i\right)+\left(1-\left|\varepsilon_i-1\right|\right)+\frac{1}{N}\sum_{j=1}^{N-1}\left(1-\left|\varepsilon_i-\frac{j}{N}\right|\right)\right)$$

$$trans(c,d)=C\frac{g(c,d)}{h(d)}\ \text{ with }\begin{cases}C=1 \text{ if } d\in\{0,1\}\\ C=\frac{1}{N} \text{ if } d\in\ ]0,1[\end{cases}$$

Noise model parameters:

$(\{\mu_0,\mu_1\},\{\sigma_0,\sigma_1\})=KMeans(Y,2)$ with Y the image and 2 for the two hard classes to look for. Then we determine parameters of each fuzzy level with (9).

3. SEM procedure for parameters estimation

At each iteration q until no significant modification of the estimated parameters (convergence):

a. $\widetilde{fwd}$ and $\widetilde{bwd}$ densities computation for each class c, $c\in\{0,1,\varepsilon_i\}$, $i=1,...,N-1$ is performed recursively as follows:

For $t=1$: $\widetilde{fwd}_1(c)=h(c)G_c(y_1)$

For $t>1$: $\widetilde{fwd}_t(c)=G_c(y_t)\left(\widetilde{fwd}_{t-1}(0)\widetilde{trans}(c,0)+\widetilde{fwd}_{t-1}(1)\widetilde{trans}(c,1)+\frac{1}{N}\sum_{i=1}^{N-1}\widetilde{fwd}_{t-1}(\varepsilon_i)\widetilde{trans}(c,\varepsilon_i)\right)$

For $t=T$: $\widetilde{bwd}_T(c)=1$

For $t<T$: $\widetilde{bwd}_t(c)=G_0(y_{t+1})\widetilde{bwd}_{t+1}(0)\widetilde{trans}(0,c)$

$$+G_1(y_{t+1})\widetilde{bwd}_{t+1}(1)\widetilde{trans}(1,c)+\frac{1}{N}\sum_{i=1}^{N-1}G_{\varepsilon_i}(y_{t+1})\widetilde{trans}(\varepsilon_i,c)\widetilde{bwd}_{t+1}(\varepsilon_i)$$

These computations must be normalized. G_c is given by: $G_c(y)=\frac{1}{\sigma_c\sqrt{2\pi}}\exp\left(-\frac{(y-\mu_c)^2}{2\sigma_c^{\,2}}\right)$

b. Stochastic re-estimation of parameters:

To obtain one *a posteriori* realization of X, simulate a fuzzy Markov chain using the following :
Posterior distributions of X are defined by:

$$\widetilde{init}(c)=\widetilde{fwdbwd}_1(c) \text{ and } \widetilde{trans}(c,d)^{t+1}=\frac{f(d\mid c)G_d(y_{t+1})\widetilde{bwd}_{t+1}(d)}{\int_0^1 f(d\mid c)G_d(y_{t+1})\widetilde{bwd}_{t+1}(d)dv(d)}$$

$$init(c)=\begin{cases}\widetilde{init}(\mathrm{c}) \text{ if } c\in\{0,1\}\\ \frac{1}{N}\widetilde{init}(\mathrm{c}) \text{ if } c\in\,]0,1[\end{cases} \text{ and } trans(c,d)^t=\begin{cases}\widetilde{trans}(\mathrm{c,d})^{\mathrm{t}} \text{ if } d\in\{0,1\}\\ \frac{1}{N}\widetilde{trans}(\mathrm{c,d}) \text{ if } d\in\,]0,1[\end{cases}$$

It has to be noted that $trans(c,d)^t$ depends on t since a different transition matrix is computed for each element of the posterior realization, as we are dealing with a non stationary Markov chain.
The estimated values of the parameters at the iteration q are computed on the simulated *a posteriori* chain $\{x_t\mid t=1...T\}$ as follows:

$$init(c)^{[q]}=\delta(x_1^{[q]},c)$$

For the *a priori* model:

$$trans(c,d)^{[q]}=\frac{\sum_{t=2}^{T}\delta(x_{t-1}^{[q]},c)\delta(x_t^{[q]},d)}{\sum_{t=2}^{T}\delta(x_{t-1}^{[q]},c)}$$

For the noise model:

$$\mu_c^{[q]}=\frac{\sum_{t=1}^{T}y_t\delta(x_t^{[q]},c)}{\sum_{t=1}^{T}\delta(x_t^{[q]},c)} \quad \sigma_c^{2[q]}=\frac{\sum_{t=1}^{T}\delta(x_t^{[q]},c)(y_t-\mu_c^{[q]})^2}{\sum_{t=1}^{T}\delta(x_t^{[q]},c)}$$

for $c=0$ and $c=1$. For fuzzy levels ($c=\varepsilon_i$) noise parameters, use equation (9)

with $\delta(m,n)=\begin{cases}1 \text{ if } m=n\\ 0 \text{ if } m\neq n\end{cases}$

4. MPM segmentation of the chain using estimated parameters:

For each x_t, determine the class (hard class or fuzzy level) minimizing the error classification probability by minimizing the following expression:

$$\widetilde{fwdbwd}_t(0)L(0,\hat{s}(y_t)) + \widetilde{fwdbwd}_t(1)L(1,\hat{s}(y_t)) + \int_0^1 \widetilde{fwdbwd}_t(\varepsilon_i)L(\varepsilon_i,\hat{s}(y_t))d\varepsilon_i$$

for every $\hat{s}(y_t)$, and where $\widetilde{fwdbwd}$ denotes the product of the forward and backward densities. The cost function L is given by (8).

5. Reverse transformation of the 1D segmented chain into the 2D or 3D segmentation map using the path saved at step 1.

Acknowledgments

This work is financially supported by a Region of Brittany research grant under the “Renouvelement des competences” program 1202-2004.

References

Benmiloud B, Pieczynski W 1995 Estimation des paramètres dans les chaînes de Markov cachées et segmentation d'images *Traitement du signal* Vol. 12 No. 5 pp. 433-454.

Celeux G, Diebolt J 1986 L'algorithme SEM : un algorithme d'apprentissage probabiliste pour la reconnaissance de mélanges de densités *Revue de statistique appliquée* Vol. 34 No. 2.

Chen JL, Gunn SR, Nixon MS 2001 Markov Random Field Model for segmentation of PET images, *Lecture Notes on Computer. Science* Vol. 2082 pp. 468-474.

Dai J 1994 Hybrid approach to speech recognition using hidden Markov models and Markov chains *Proc. Inst. Elect. Eng: Vision Image Signal Processing* Vol. 141(5) pp. 273-279.

Dempster A P, Laird N M and Rubin D B 1977 Maximum likelihood from incomplete data via the EM algorithm *J. R. Stat. Soc. B* Vol. 39 pp. 1-38.

Delignon Y, Marzouki A, Pieczynski W 1997 Estimation of Generalized Mixtures and Its Application in Image Segmentation *IEEE Transactions on Image Processing* Vol. 6(10).

Erdi YE, Mawlawi O, Larson SM, Imbriaco M, Yeung H, Finn R, Humm JL 1997 Segmentation of Lung Lesion Volume by Adaptative Positron Emission Tomography Image Thresholding *Cancer* Vol. 80(S12) pp. 2505 - 2509.

Jarritt H, Carson K, Hounsel AR, Visvikis D 2006 The role of PET/CT scanning in radiotherapy planning *British Journal of Radiology* Vol. 79**,** S27-S35.

Jordan K 1990 IEC emission phantom Appendix Performance evaluation of positron emission tomographs *Medical and Public Health Research Programme of the European Community*.

Hatt M, Roux C, Visvikis D 2007 3D Fuzzy Adaptive Unsupervised Bayesian Segmentation for Volume Determination in PET *4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, Arlington, USA, April 2007, accepted.

Kamata S, Eason RO, and Bandou Y 1999 A New Algorithm for N-dimensional Hilbert Scanning *IEEE Transactions on Image Processing* Vol. 8 pp 964-973.

Kim J, Feng DD, Cai TW, Eberl S 2002 Automatic 3D Temporal Kinetics Segmentation of Dynamic Emission Tomography Image Using Adaptative Region Growing Cluster Analysis *IEEE NSS-MIC Conference Record* Vol. 3 pp. 1580-1583.

Krak NC, Boellaard R, et al 2005 Effects of ROI definition and reconstruction method on quantitative outcome and applicability in a response monitoring trial *Eur. J. Nucl. Med. Mol. Im.* Vol. 32 pp. 294-301.

Lamare F, Turzo A, Bizais Y, Cheze Le Rest C, Visvikis D 2006 Validation of a Monte Carlo simulation of the Philips Allegro/GEMINI PET systems using GATE *Physics in Medicine and Biology* Vol. 51 pp. 943-962.

Lanchantin P, Pieczynski W 2004 Unsupervised non stationary image segmentation using triplet Markov chains *Advanced Concepts for Intelligent Vision Systems (ACVIS 04)* Belgium.

Marroquin J, Mitter S, Poggio T 1987 Probabilistic solution of ill-posed problems in computational vision Journal of the American Statistical Association 82 pp. 76-89.

McQueen J 1967 Some methods for classification and analysis of multivariate observations *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* 1:281-297.

Nestle U, Kremp S, Schaefer-Schuler A, Sebastian-Welch C, Hellwig D, Rübe C, Kirsch CM 2005 Comparison of Different Methods for Delineation of 18F-FDG PET-Positive Tissue for Target Volume Definition in Radiotherapy of Patients with Non-Small Cell Lung Cancer, *Journal of Nuclear Medicine* Vol. 46(8) pp. 1342-8.

Pieczynski W 2003 Modèles de Markov en traitement d'images *Traitement du Signal* Vol. 20(3) pp. 255-277.

Pieczynski W, Derode S 2004 Signal and Image Segmentation Using Pairwise Markov Chains *IEEE Transactions on Signal Processing* Vol. 52(9).

Reader AJ, Ally S, Bakatselos F, Manavaki R, Walledge RJ, Jeavons AP et al. 2002 One-pass list-mode EM algorithm for high-resolution 3D PET image reconstruction into large array *IEEE Transactions on Nuclear Science* Vol. 49(3) pp. 693-699.

Reutter BW, Klein GJ, Huesman RH 1997 Automated 3D Segmentation of Respiratory-Gated PET Transmission Images *IEEE Transactions on Nuclear Science* Vol. 44(6) pp. 2473-2476.

Riddell C, Brigger P, Carson RE, Bacharach SL 1999 The Watershed Algorithm: A Method to Segment Noisy PET Transmission Images *IEEE Transactions on Nuclear Science* Vol. 46(3) pp. 731-719.

Salzenstein F, Pieczynski W 1997, Parameter Estimation in hidden fuzzy Markov random fields and image segmentation, CVGIP : Graphical Models and Image Processing Vol. 59(4) pp. 205-220.

Salzenstein F, Pieczynski W 1998 Sur le choix de méthode de segmentation statistique d'images *Traitement du signal* Vol. 15 No. 2 pp.119-128.

Salzenstein F, Collet C, Petremand M 2004, Champs de Markov flous pour images multispectrales, *Traitement du signal* Vol. 21(1) pp. 37-54.

Visvikis D, Cheze Le Rest C, Costa DC et al, 2001, Influence of OSEM and segmented attenuation correction in the calculation of standardised uptake values for FDG PET. *Eur J Nucl Med Mol Imag,* 28, pp. 1326-1335

Visvikis D, Turzo A, Gouret S, Damine P, Lamare F, Bizais Y, Cheze Le Rest C 2004 Characterisation of SUV accuracy in FDG PET using 3D RAMLA and the Philips Allegro PET scanner *J. Nucl. Med.* Vol. 45(5):103P.

Zhu W, Jiang T 2003 Automation Segmentation of PET Image for Brain Tumors *IEEE Nuclear Science Symposium Conference Record – NSS-MIC* Vol. 4 pp. 2627- 2629.

Figure Captions

Figure 1: The 3D Hilbert-Peano space filling curve for a 4× 4× 4 voxels VOI.

Figure 2: (a) A graphical representation of the IEC phantom, and the central slice of the digital IEC phantom used in the generation of the simulated datasets (b) with 2× 2× 2 mm^3 and (c) 4× 4× 4 mm^3.

Figure 3: Different images used in the segmentation study; (a)-(d) simulated : (a) ratio 4:1, 20 millions of coincidences, 64 mm^3, (b) ratio 8:1, 40 millions, 64 mm^3, (c) ratio 4:1, 20 millions of coincidences, 8 mm^3, (d) ratio 8:1, 40 millions, 8 mm^3 (e)-(h) acquired : (e) ratio 4:1, 2min acquisition time, 64 mm^3, (f) ratio 8:1, 5min, 64 mm^3, (g) ratio 4:1, 5min, 8 mm^3, (h), ratio 8:1, 5min, 8 mm^3.

Figure 4: Examples of segmentation maps for the 28mm sphere (one slice) (a) PET ROI, (b) Digital "ground truth", (c) HMC map, (d) T42 map, (e) FHMC with 2 fuzzy levels (light and dark grey voxels) segmentation map, (f) Map used for VOI (hard class + 1st fuzzy level, FHMC 1/2), (g) Map for quantitation (only hard class voxels, FHMC 0/2), (h) T50 map. Note that in this particular case, FHMC 1/2 for VOI and T50 result in the same map but this is of course not always the case (especially having considered the complete 3D volume).

Figure 5: (a) Classification errors for the lesion VOI determination and (b) lesion activity recovery errors in the simulated images for the FHMC vs HMC segmentation. Different number of fuzzy levels (2 or 3) were used in the segmentation process and different number of these were subsequently selected to (a) form the segmented volumes or (b) determine lesion average activity concentrations for the different imaging conditions considered.

Figure 6: Classification errors in lesion VOI determination from the simulated images: (a) 64 mm^3 voxels and (b) 8 mm^3 voxels for the FHMC 1/2 vs T42 segmentation under variable imaging conditions.

Figure 7: Repartition of PCEs and NCEs from the (a) FHMC 1/2, (b) HMC and (c) T42 segmentation results for the different simulated imaging configurations considered.

Figure 8: Lesion VOI estimation errors from the simulated images: (a) 64 mm^3 voxels and (b) 8 mm^3 voxels for the FHMC 1/2 vs T42 segmentation under variable imaging conditions.

Figure 9: Lesion average activity concentration estimation errors from the simulated images: (a) 64 mm^3 voxels and (b) 8 mm^3 voxels for FHMC 0/2 vs T50 segmentation under variable imaging conditions.

Figure 10: Lesion VOI estimation errors from the acquired images: (a) 64 mm^3 voxels and (b) 8 mm^3 voxels for the FHMC 1/2 vs T42 segmentation under variable imaging conditions.

Figure 11: Lesion average activity concentration estimation errors from the acquired images: (a) 64 mm^3 voxels and (b) 8 mm^3 voxels for the FHMC 0/2 vs T50 segmentation under variable imaging conditions.

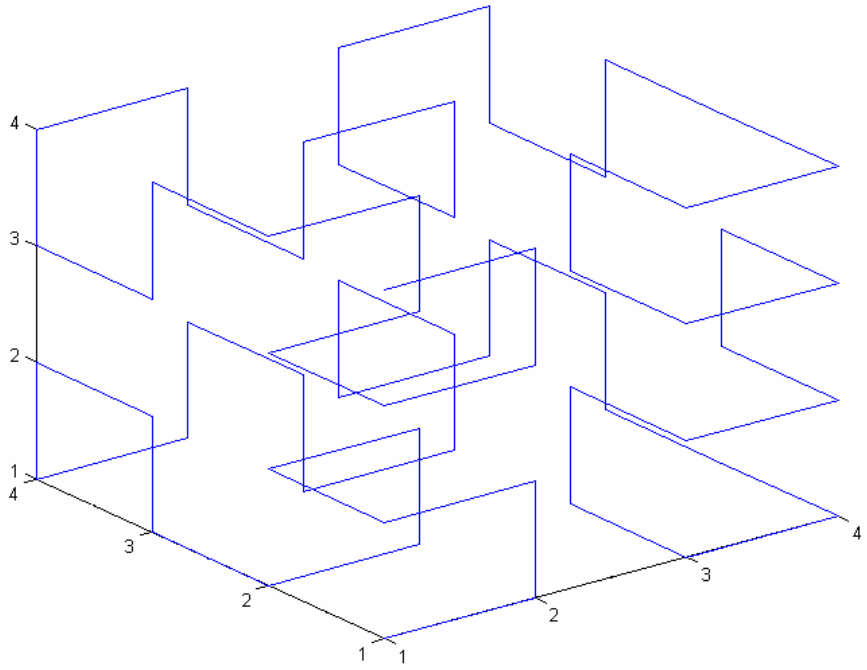
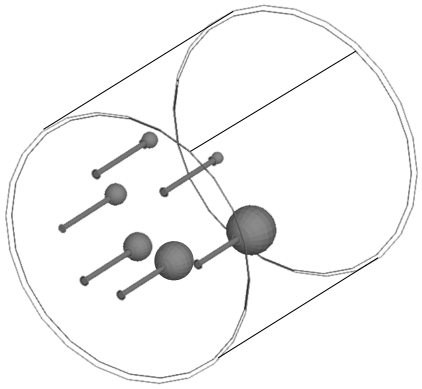
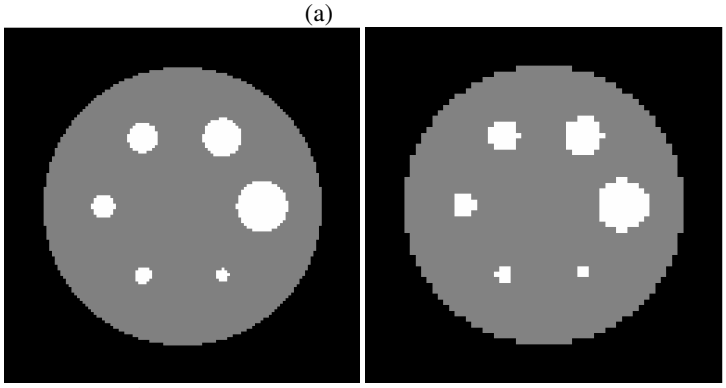


Figure 1

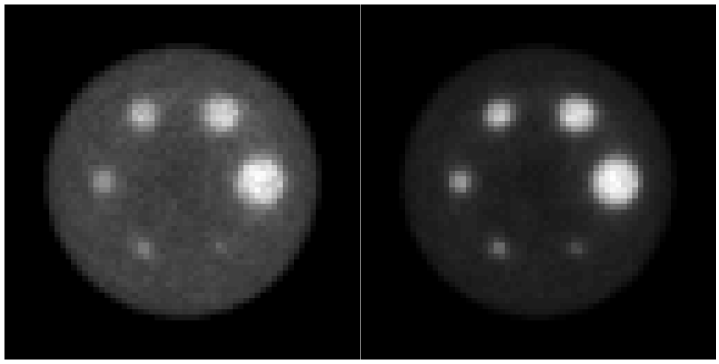


(a)

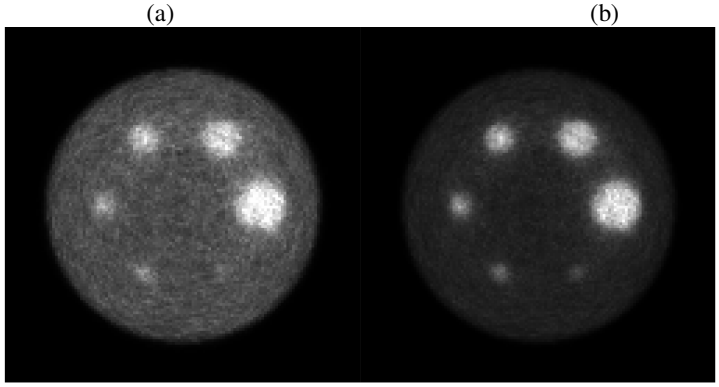


(b) (c)

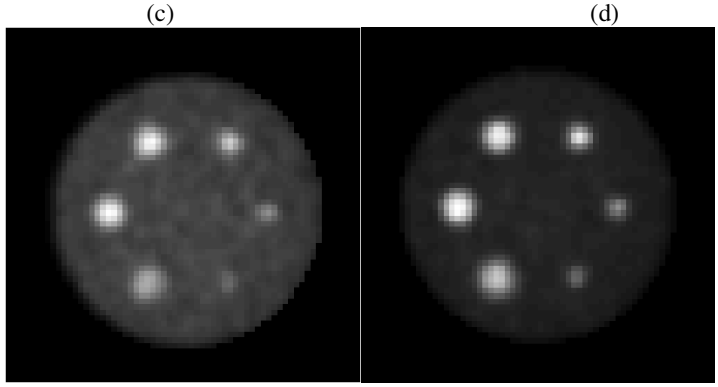
Figure 2



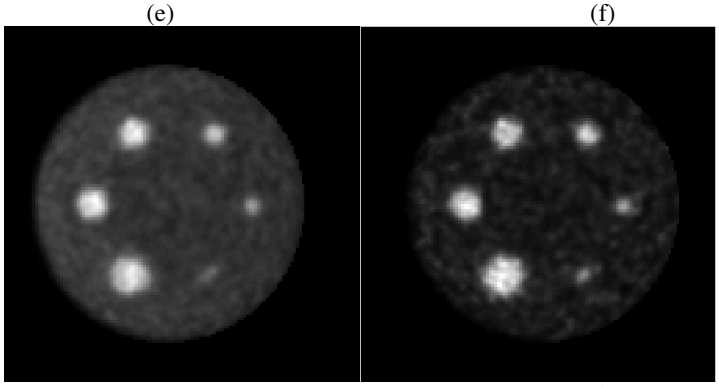
(a) (b)



(c) (d)



(e) (f)



(g) (h)

Figure 3

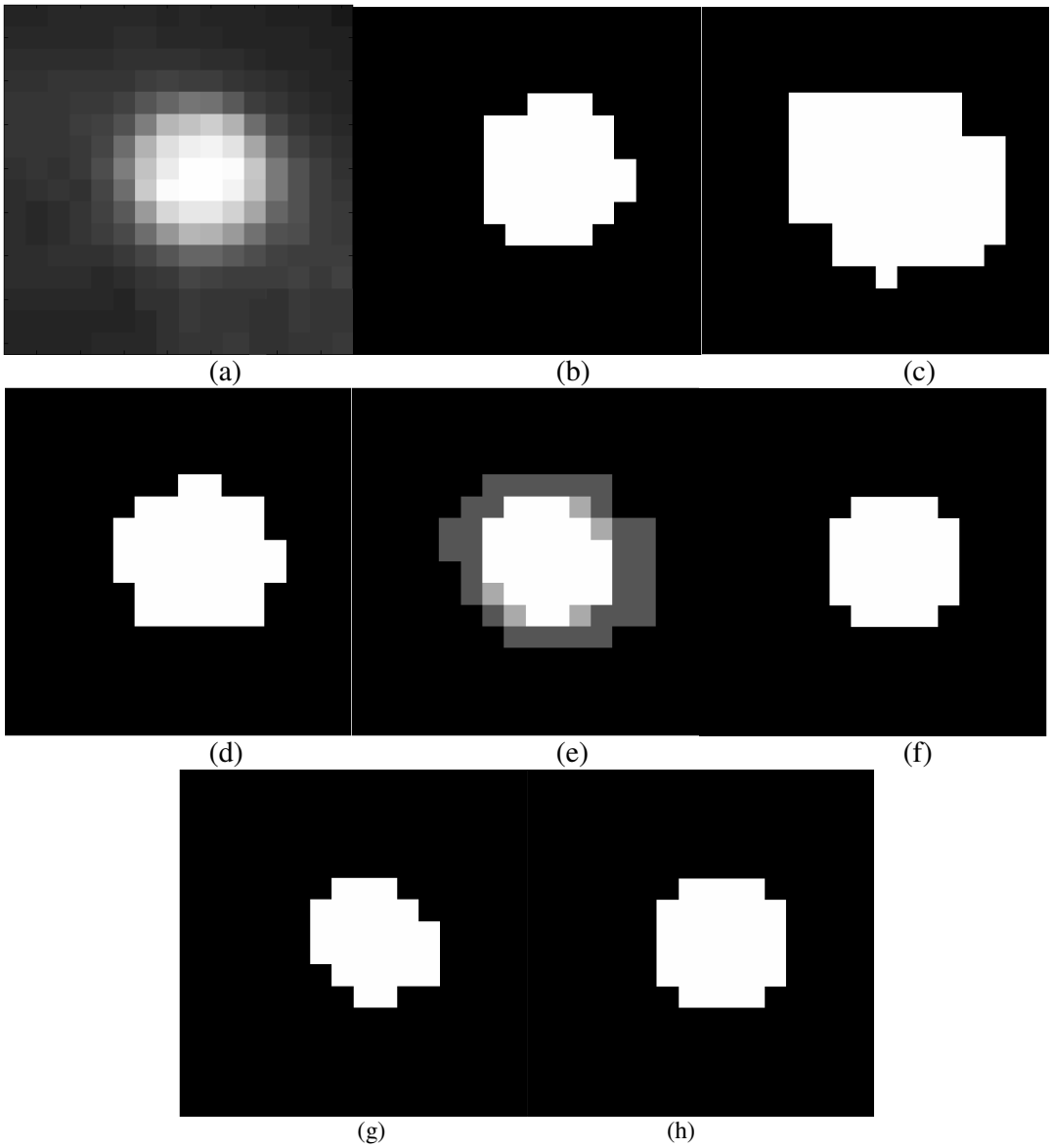
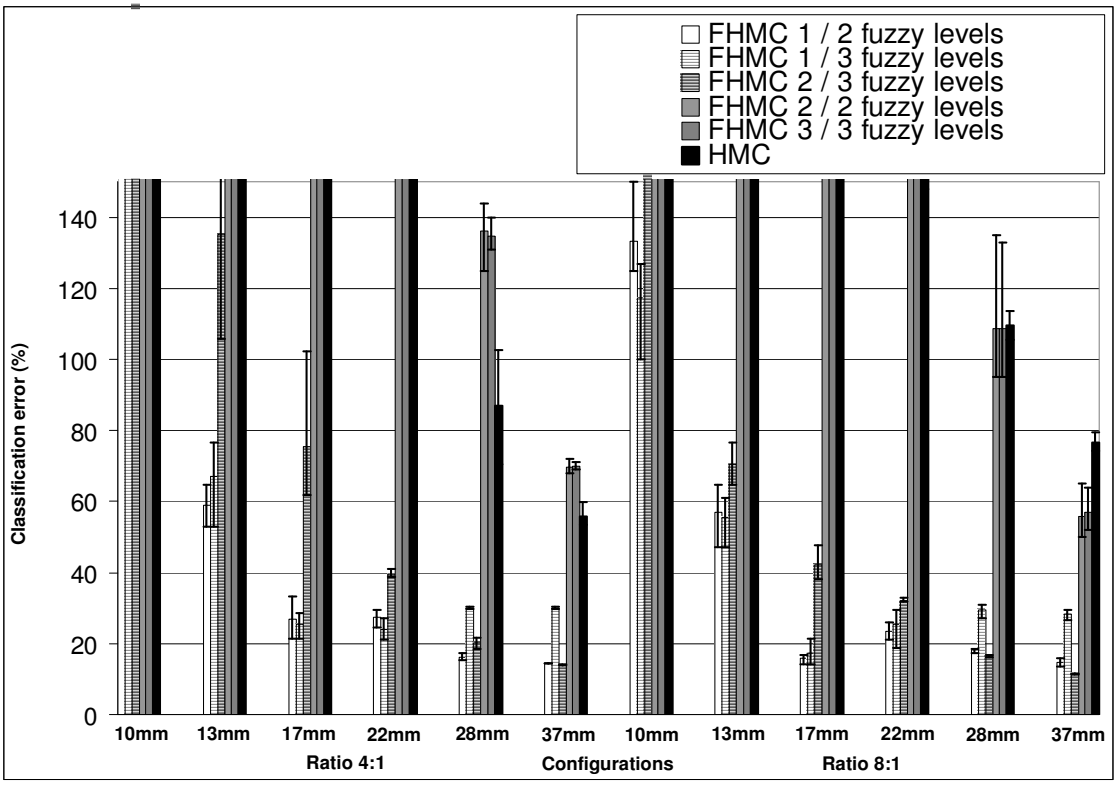
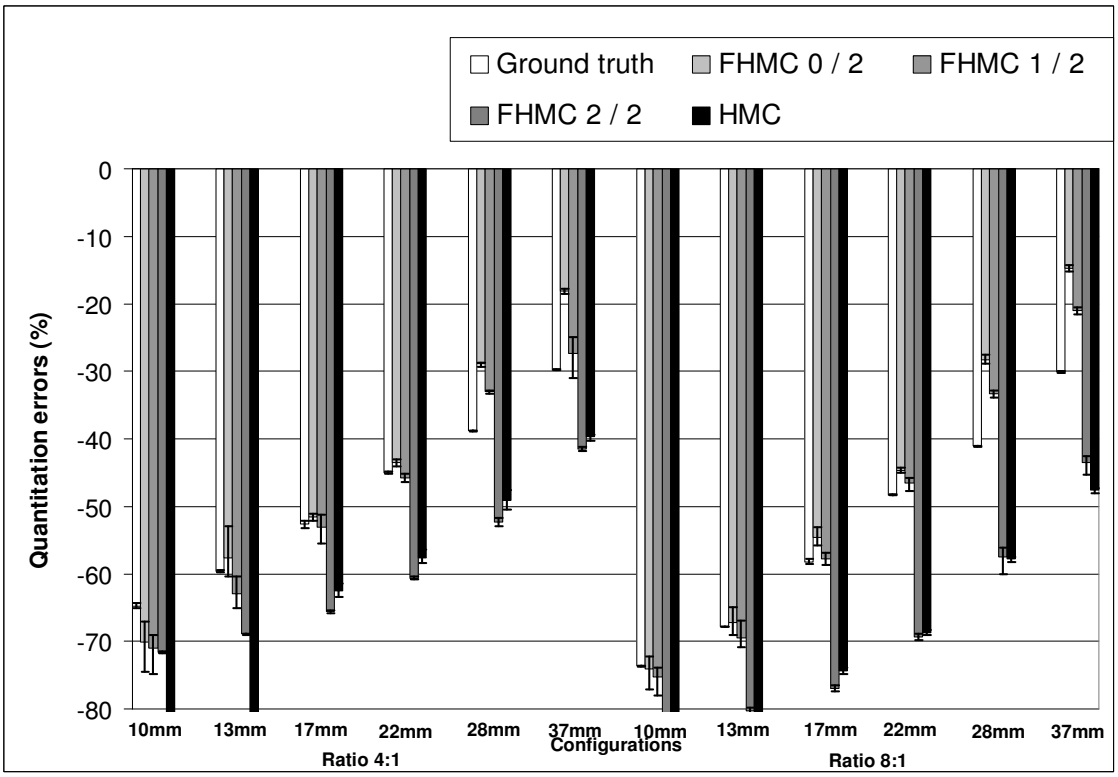


Figure 4

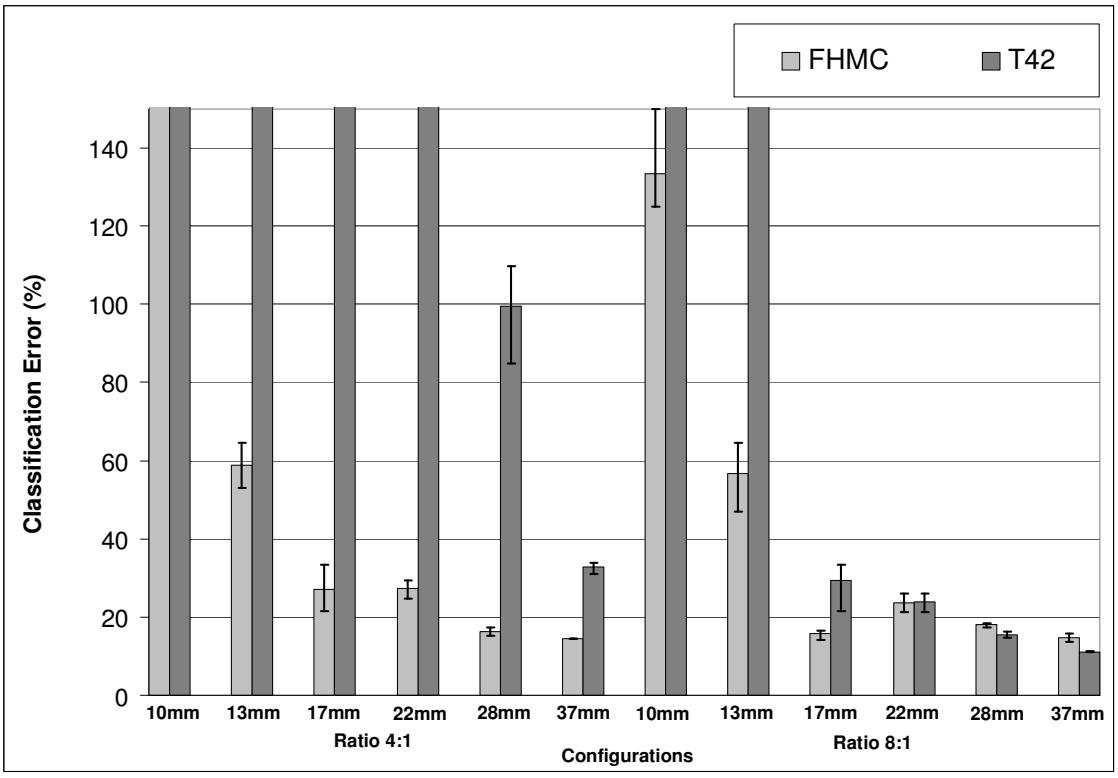

FHMC 1 / 2 fuzzy levels
FHMC 1 / 3 fuzzy levels
FHMC 2 / 3 fuzzy levels
FHMC 2 / 2 fuzzy levels
FHMC 3 / 3 fuzzy levels
HMC
Classification error (%)
0
20
40
60
80
100
120
140
10mm
13mm
17mm
22mm
28mm
37mm
10mm
13mm
17mm
22mm
28mm
37mm
Ratio 4:1
Configurations
Ratio 8:1

(a)

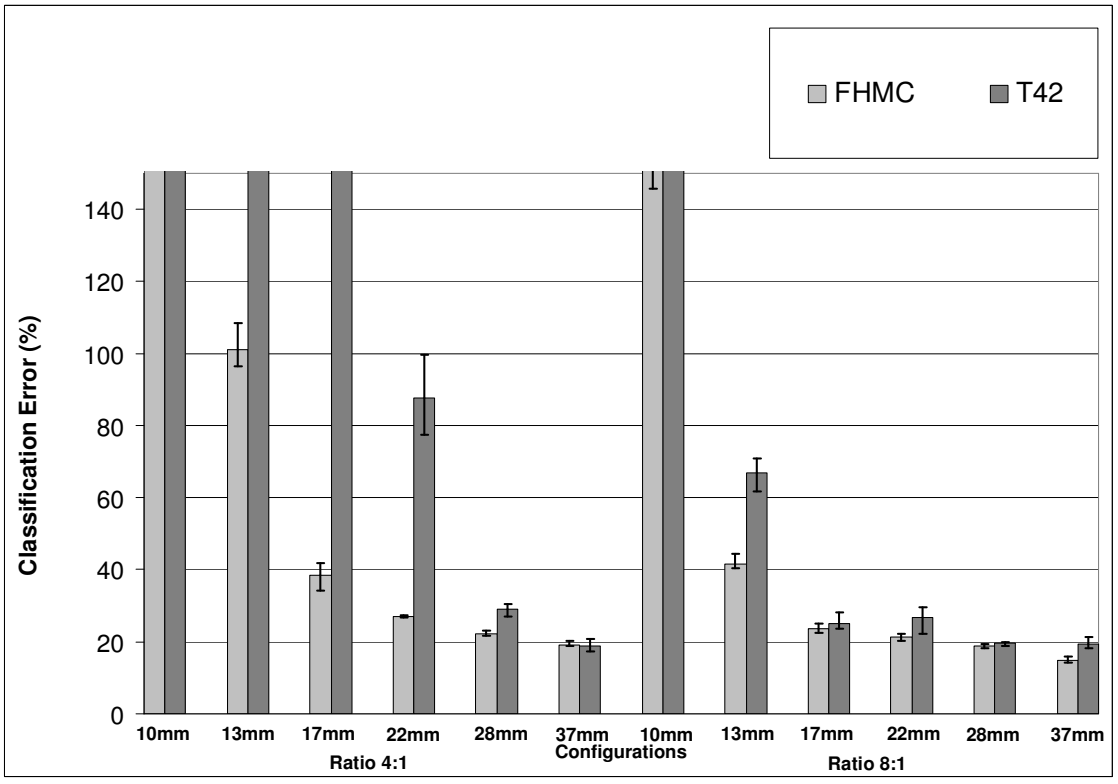

Ground truth
FHMC 0 / 2
FHMC 1 / 2
FHMC 2 / 2
HMC
Quantitation errors (%)
0
-10
-20
-30
-40
-50
-60
-70
-80
10mm
13mm
17mm
22mm
28mm
37mm
10mm
13mm
17mm
22mm
28mm
37mm
Ratio 4:1
Configurations
Ratio 8:1

(b)

Figure 5

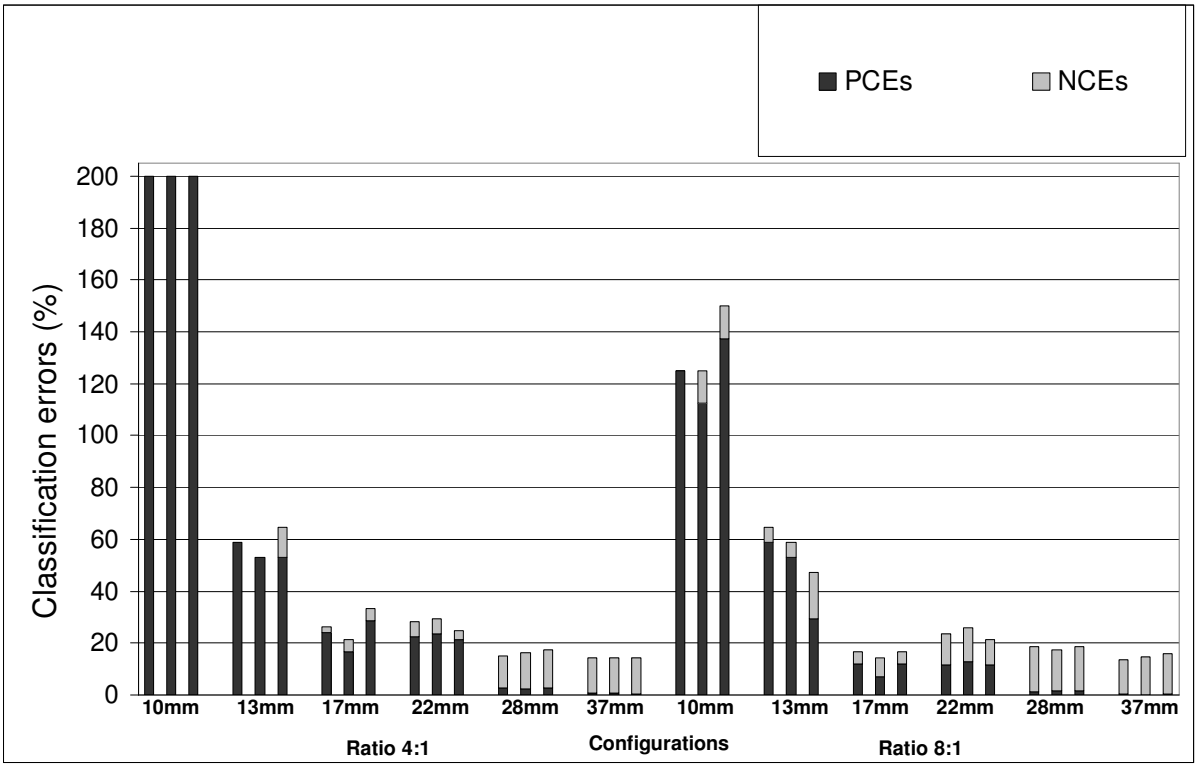

FHMC
T42
Classification Error (%)
0
20
40
60
80
100
120
140
10mm
13mm
17mm
22mm
28mm
37mm
10mm
13mm
17mm
22mm
28mm
37mm
Ratio 4:1
Configurations
Ratio 8:1

(a)

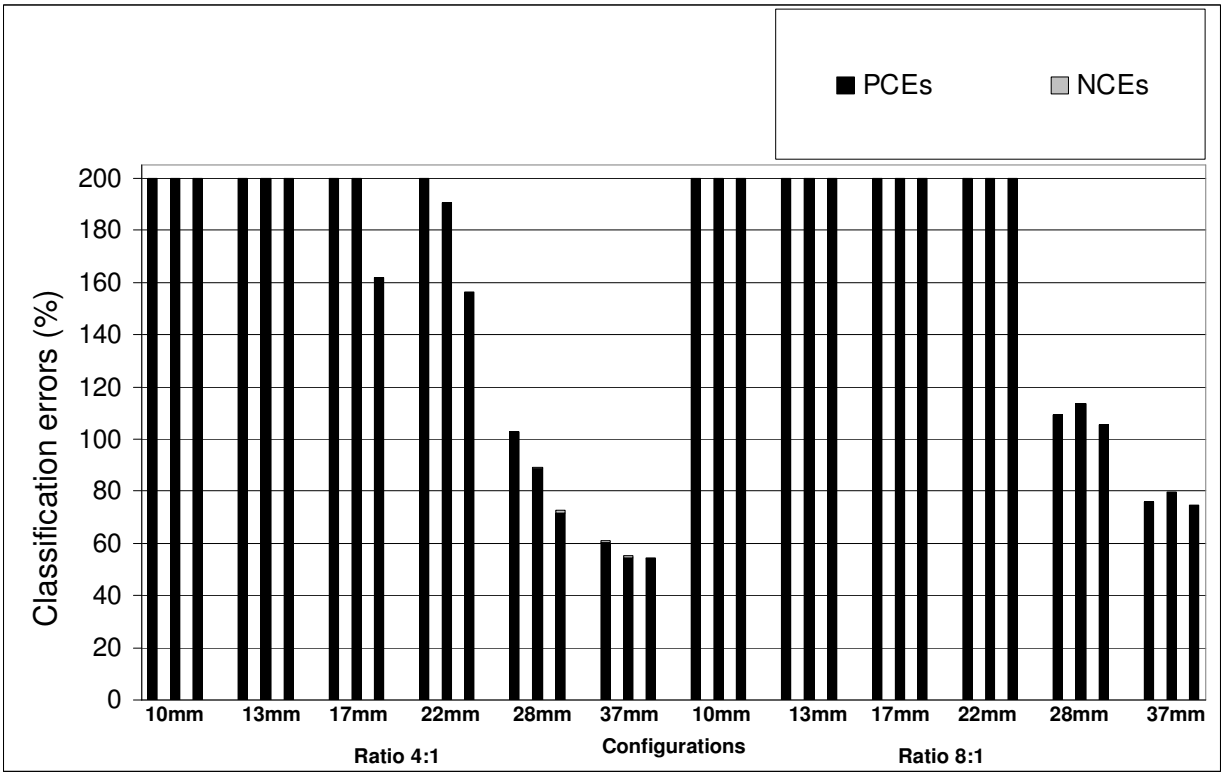

FHMC
T42
Classification Error (%)
0
20
40
60
80
100
120
140
10mm
13mm
17mm
22mm
28mm
37mm
10mm
13mm
17mm
22mm
28mm
37mm
Ratio 4:1
Configurations
Ratio 8:1

(b)

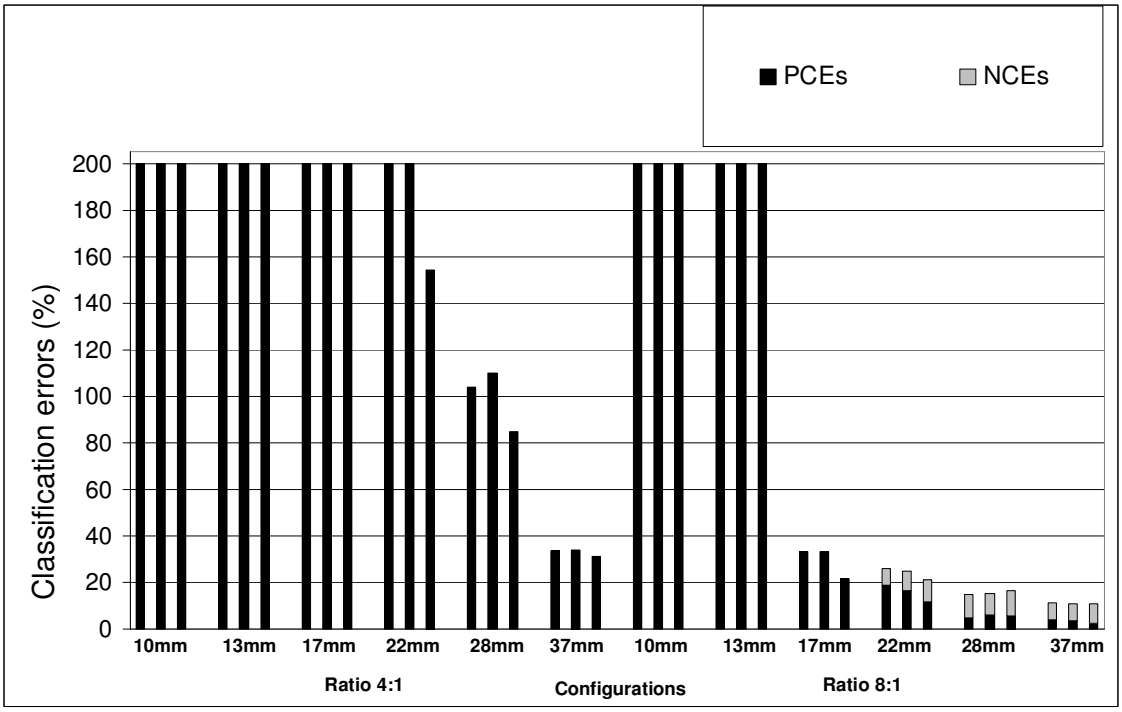
Figure 6


PCEs
NCEs
Classification errors (%)
200
180
160
140
120
100
80
60
40
20
0
10mm
13mm
17mm
22mm
28mm
37mm
10mm
13mm
17mm
22mm
28mm
37mm
Ratio 4:1
Configurations
Ratio 8:1

(a)

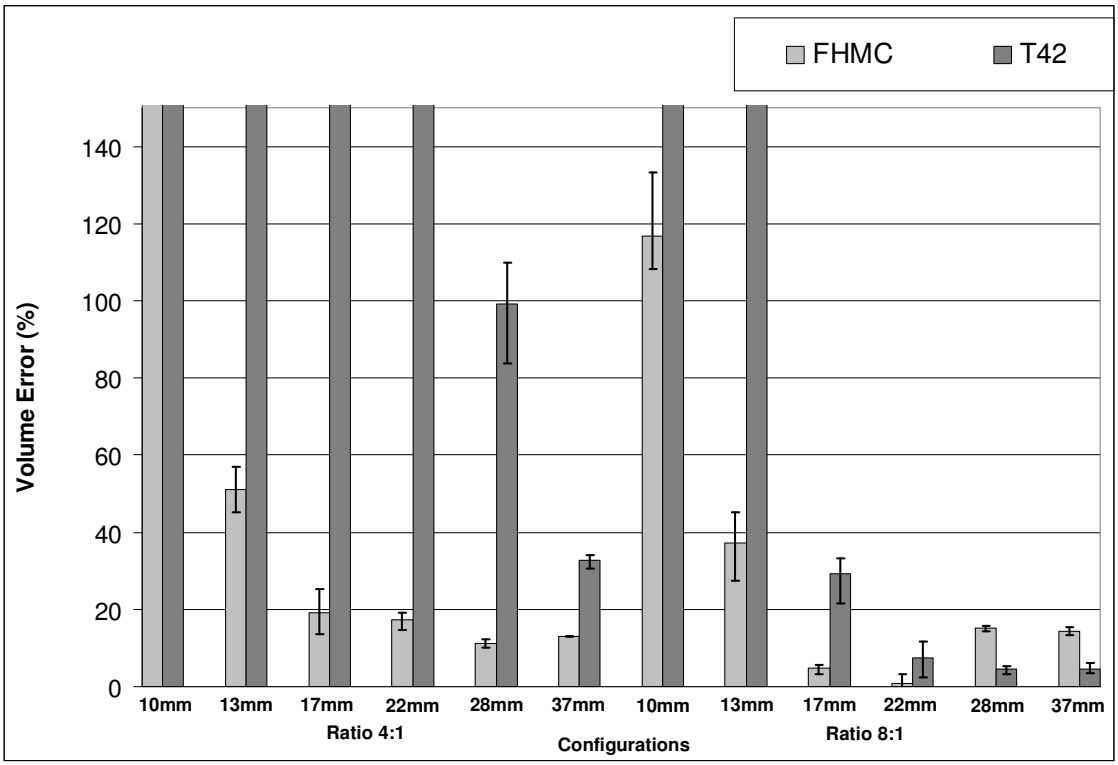

PCEs
NCEs
Classification errors (%)
200
180
160
140
120
100
80
60
40
20
0
10mm
13mm
17mm
22mm
28mm
37mm
10mm
13mm
17mm
22mm
28mm
37mm
Ratio 4:1
Configurations
Ratio 8:1

(b)

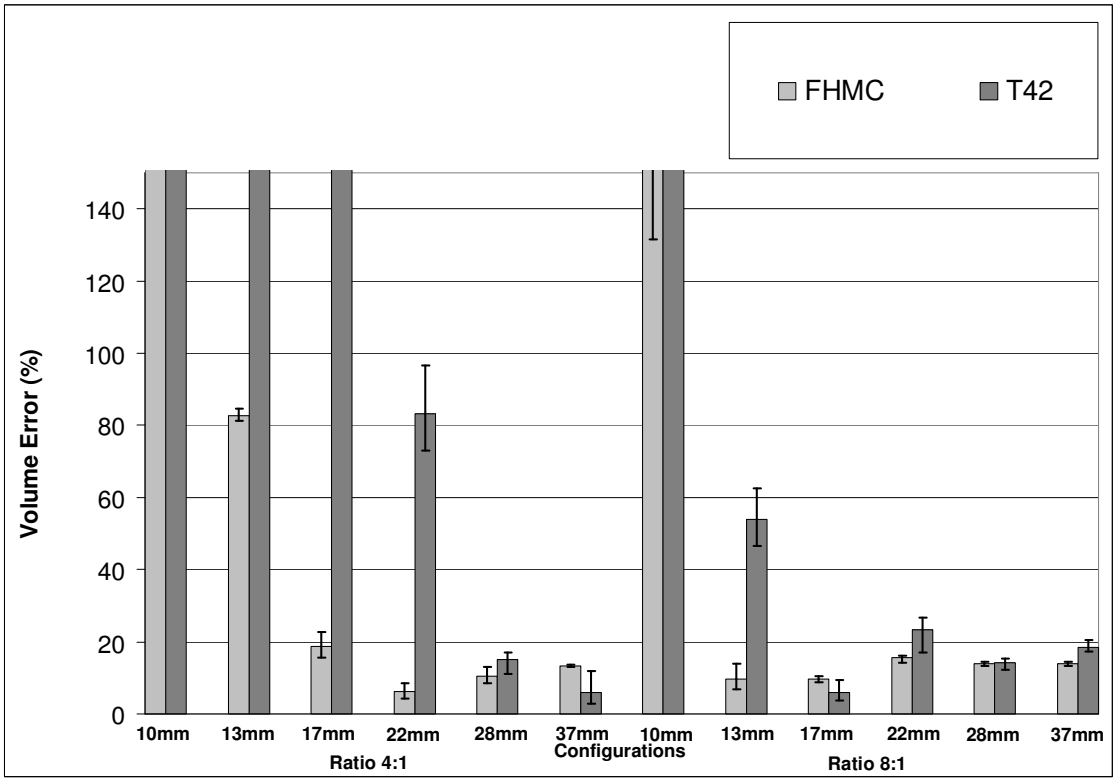

PCEs
NCEs
Classification errors (%)
200
180
160
140
120
100
80
60
40
20
0
10mm
13mm
17mm
22mm
28mm
37mm
10mm
13mm
17mm
22mm
28mm
37mm
Ratio 4:1
Configurations
Ratio 8:1

(c)

Figure 7

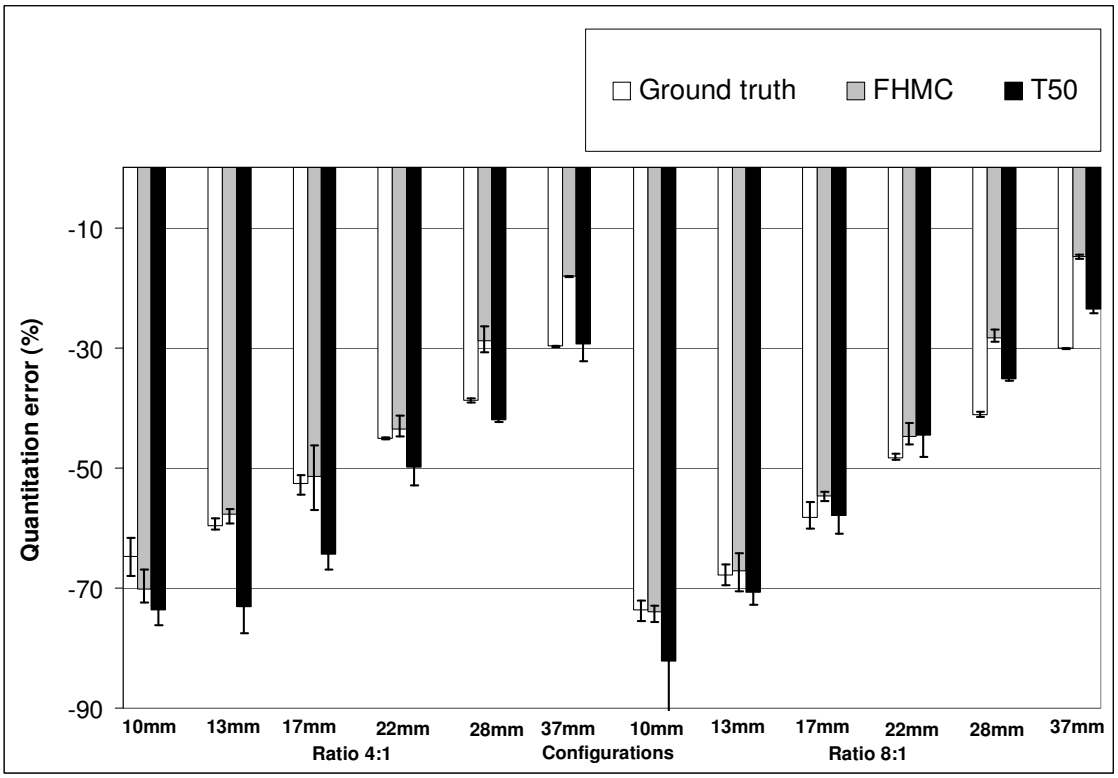

FHMC
T42
Volume Error (%)
0
20
40
60
80
100
120
140
10mm
13mm
17mm
22mm
28mm
37mm
10mm
13mm
17mm
22mm
28mm
37mm
Ratio 4:1
Configurations
Ratio 8:1

(a)

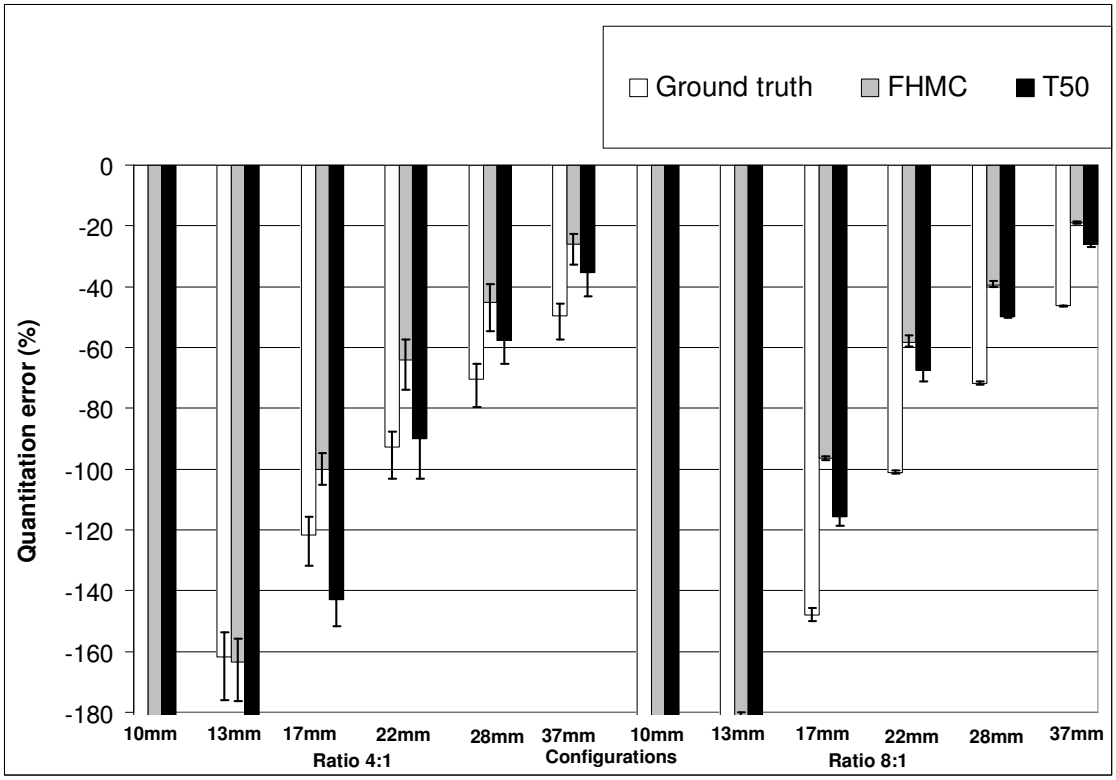

FHMC
T42
Volume Error (%)
0
20
40
60
80
100
120
140
10mm
13mm
17mm
22mm
28mm
37mm
10mm
13mm
17mm
22mm
28mm
37mm
Ratio 4:1
Configurations
Ratio 8:1

(b)

Figure 8

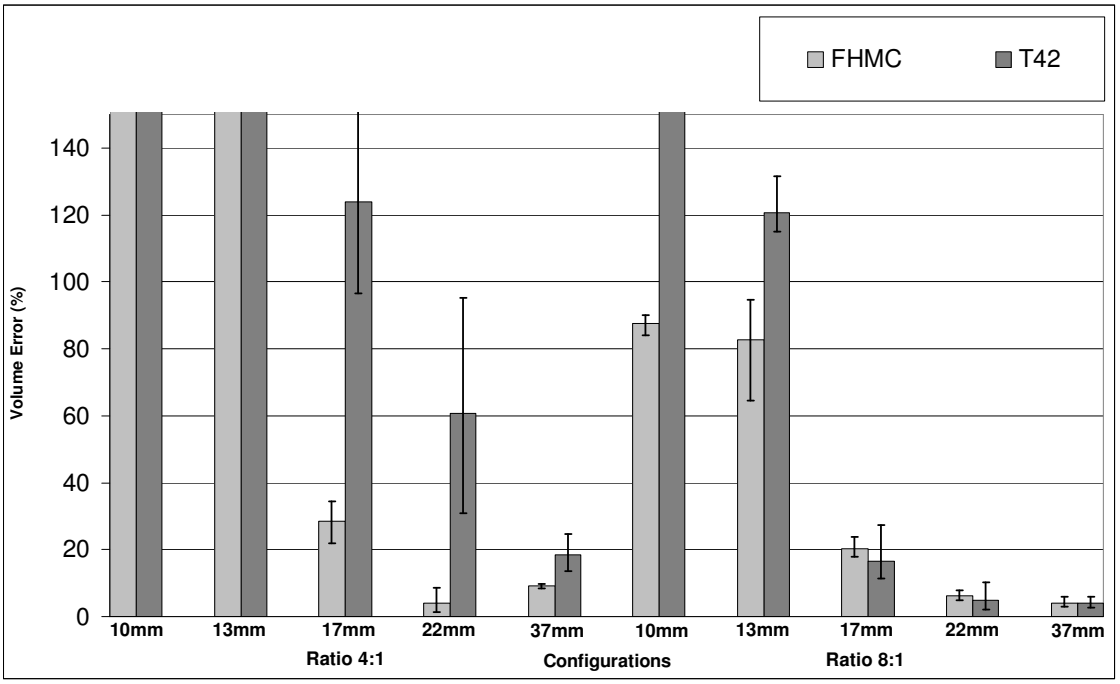

Ground truth
FHMC
T50
-10
-30
-50
-70
-90
Quantitation error (%)
10mm
13mm
17mm
22mm
28mm
37mm
10mm
13mm
17mm
22mm
28mm
37mm
Ratio 4:1
Configurations
Ratio 8:1

(a)

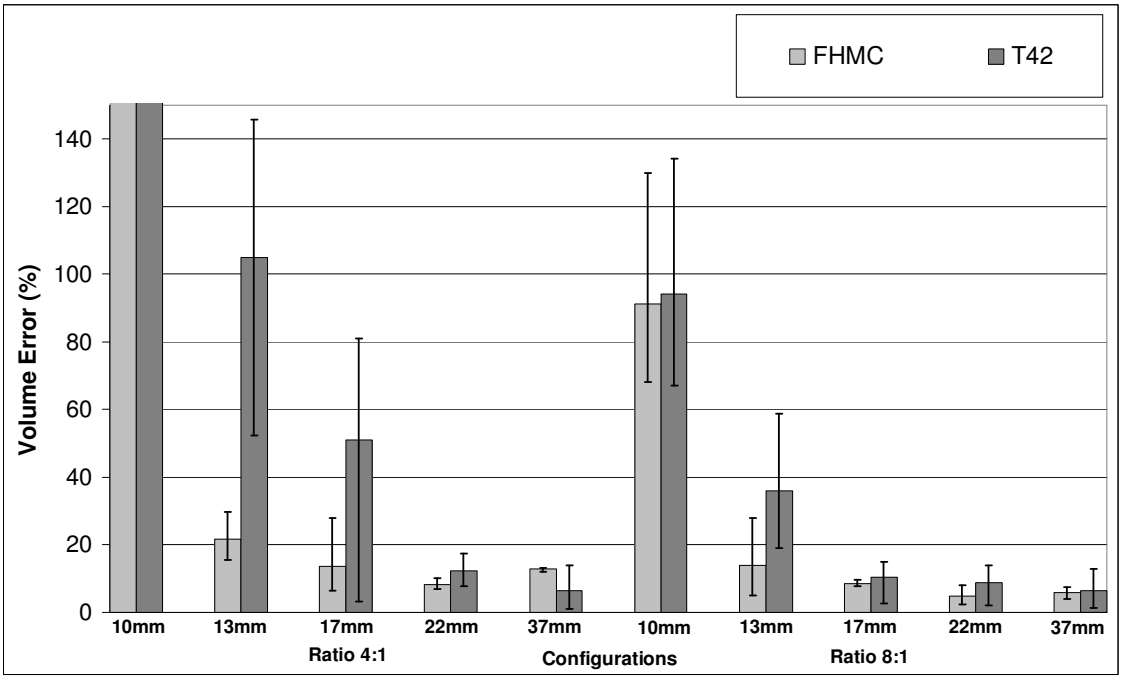

Ground truth
FHMC
T50
0
-20
-40
-60
-80
-100
-120
-140
-160
-180
Quantitation error (%)
10mm
13mm
17mm
22mm
28mm
37mm
10mm
13mm
17mm
22mm
28mm
37mm
Ratio 4:1
Configurations
Ratio 8:1

(b)

Figure 9

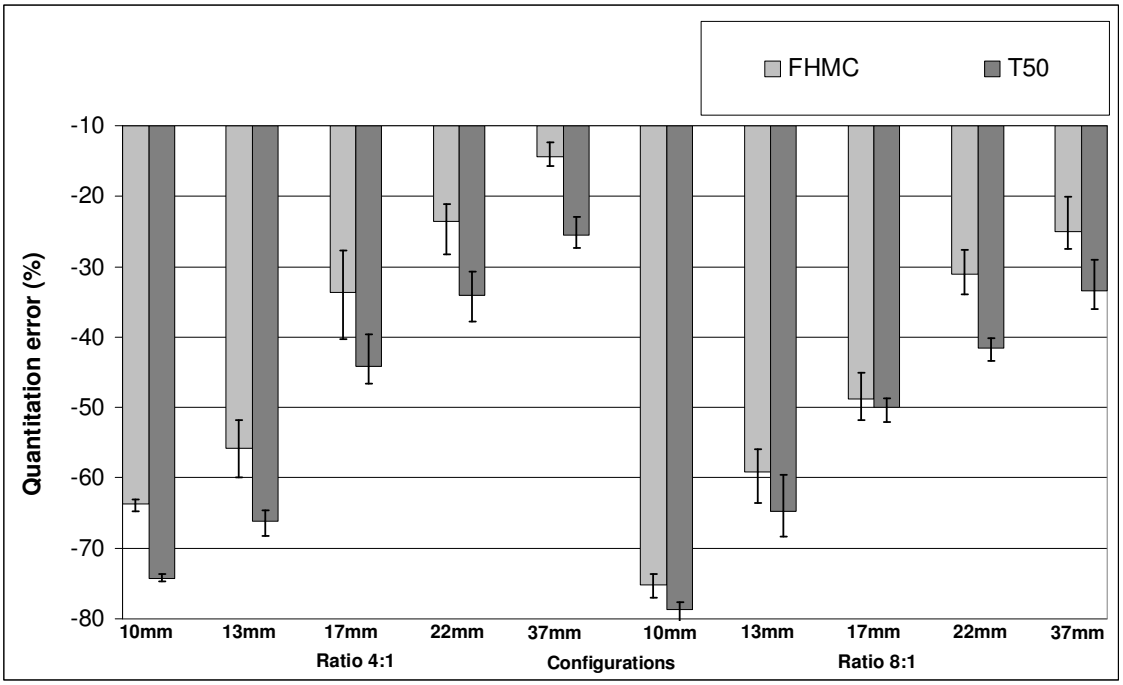

FHMC
T42
Volume Error (%)
0
20
40
60
80
100
120
140
10mm
13mm
17mm
22mm
37mm
10mm
13mm
17mm
22mm
37mm
Ratio 4:1
Configurations
Ratio 8:1

(a)

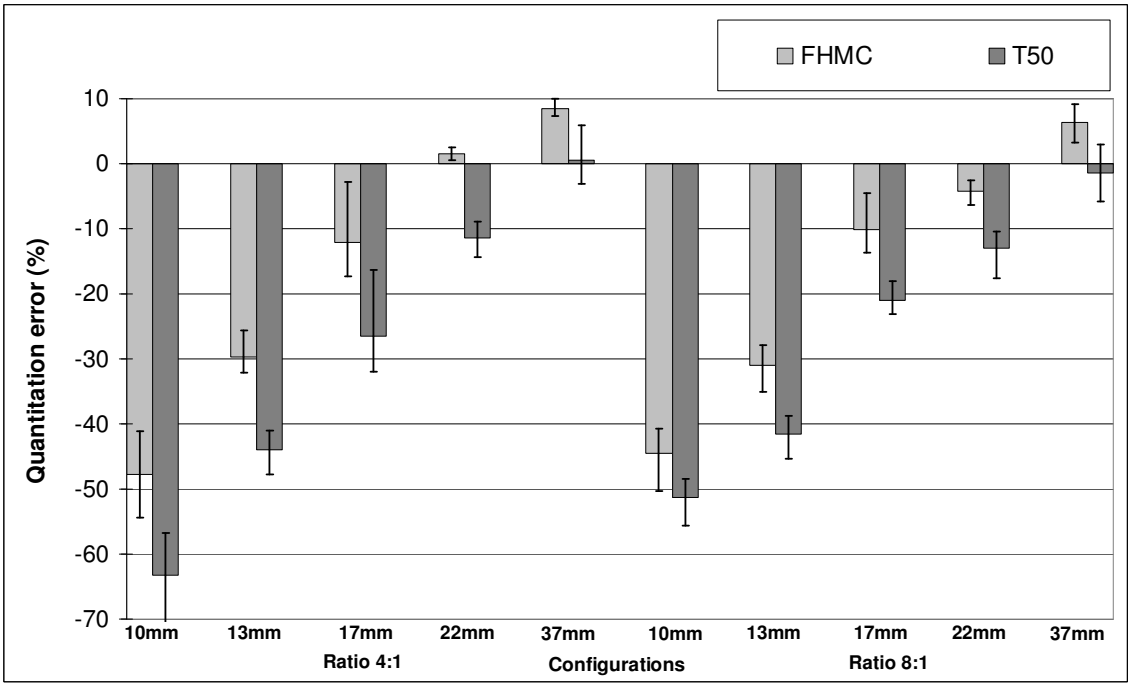

FHMC
T42
Volume Error (%)
0
20
40
60
80
100
120
140
10mm
13mm
17mm
22mm
37mm
10mm
13mm
17mm
22mm
37mm
Ratio 4:1
Configurations
Ratio 8:1

(b)

Figure 10


FHMC
T50
Quantitation error (%)
-10
-20
-30
-40
-50
-60
-70
-80
10mm
13mm
17mm
22mm
37mm
10mm
13mm
17mm
22mm
37mm
Ratio 4:1
Configurations
Ratio 8:1

(a)


FHMC
T50
Quantitation error (%)
10
0
-10
-20
-30
-40
-50
-60
-70
10mm
13mm
17mm
22mm
37mm
10mm
13mm
17mm
22mm
37mm
Ratio 4:1
Configurations
Ratio 8:1

(b)

Figure 11