

Class Conditional Entropic Prior for MRI Enhanced SPECT Reconstruction

S. Pedemonte¹, M. J. Cardoso¹, A. Bousse², C. Panagiotou¹,
D. Kazantsev¹, S. Arridge¹, B. F. Hutton², S. Ourselin¹

Abstract—Maximum Likelihood Estimation can provide an accurate estimate of activity distribution for Positron Emission Tomography (PET) and Single Photon Emission Computed Tomography (SPECT), however its unconstrained application suffers from *dimensional instability* due to approximation of activity distribution to a grid of point processes. Correlation between the activity distribution and the underlying tissue morphology enables the use of information from an intra-subject anatomical image to improve the activity estimate.

Several approaches have been proposed to include anatomical information in the process of activity estimation. Methods based on information theoretic similarity functionals are particularly appealing as they abstract from any assumption about the nature of the images. However, due to multiplicity of the similarity functional, such methods tend to discard boundary information from the anatomical image.

This paper presents an extension of state of the art methods by introducing a hidden variable denoting tissue composition that conditions an entropic similarity functional. This allows one to include explicit knowledge of the MRI imaging system model, effectively introducing additional information.

The proposed method provides an intrinsic edge-preserving feature, it outperforms conventional methods based on Joint Entropy in terms of *bias/variance* characteristics, and it does not introduce additional parameters.

I. INTRODUCTION

Radio-tracer concentration in PET and SPECT has been shown to correlate with MRI and CT anatomical images in several brain and cardiac imaging studies [1], [2], [3]. Information from the anatomical image can thus be embedded in the reconstruction process in order to improve the estimate of radio-tracer activity. This can be achieved with a *Maximum A Posteriori* (MAP) approach for estimation of the activity, where the prior activity distribution is inferred from the anatomical image. While *Maximum Likelihood* (ML) estimation alone provides accurate estimation of the activity distribution for Emission Tomography, it suffers from *dimensional instability* due to approximation of the activity spatial density to a grid of point sources [4], [5],

This grant has been supported by the Engineering and Physical Sciences Research Council under Grant EP/G026483/1.

UCL/UCLH is partially supported by a UK Department of Health NIHR Biomedical Research Centres funding scheme.

¹Centre for Medical Image Computing, University College London, London, UK. (Email: s.pedemonte@cs.ucl.ac.uk)

²Institute of Nuclear Medicine, University College London Hospitals NHS Trust, 235 Euston Road (T 5), London NW1 2BU, UK

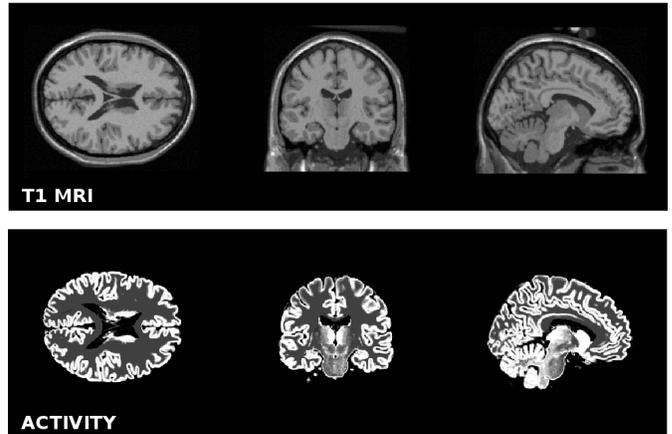


Fig. 1. Activity phantom (bottom) and co-registered T1-weighted MRI image of the same subject (top). Both images are generated synthetically from averaged partial volume segmentation of 27 MRI images of the same subject (BrainWeb [9]).

[6]. Anatomical priors have been shown to be effective in overcoming dimensional instability as they act as regularization terms. Furthermore, the information provided by the anatomical images allows to get around the intrinsic resolution associated with Poisson emission statistics [7], [8]. Multimodality-aided image reconstruction can thus in turn reduce radio-pharmaceutical dose, or equivalently produce more accurate images with unchanged photon counts.

Since its commercial introduction by General Electric in 1999, SPECT/CT has quickly become the standard in clinical practice allowing patient specific attenuation correction. Hybrid PET/MRI and SPECT/MRI imaging systems are currently under development [10], [11]. These will enable new biological and pathological analysis tools for clinical applications and preclinical research [12], [13], by exploiting the correlation between the two modalities. However, loose correlation between observable changes in tissue composition and areas of increased (hot spots) and decreased (cold spots) activity has to be taken into account. The problem has long been studied and several approaches have been proposed, based on different assumptions about the correlation between the two images, the great majority falling within three categories: methods that favor a piecewise uniform reconstruction by segmenting the anatomical image and subsequently applying a smoothing prior within each identified region [14], [15]; methods that explicitly extract

boundary information from the anatomical image and relax the effect of a global smoothing prior across the identified edges [16], [3]; methods based on information theoretic similarity functionals [17], [18], [19], [20], [4].

The last category has proven particularly interesting, as the involved functionals do not require either explicit segmentation nor boundary extraction from the anatomical image, steps that are inherently sensitive to noise because of selection of high frequency components of the image. Information theoretic similarity functionals provide the means to assess the structural similarity between two images in terms of common information content, bypassing the incommensurate relationship - due for example to multi-modality - between intensity of corresponding areas in the two images [21], [22]. The introduction of anatomical prior information via these functionals has been shown to improve the *a posteriori* estimate of activity, by reducing its *bias* and sensitivity to noise [23].

Tang and Rahmim [17] and Somayajula et al. [18] have explored the use of Joint Entropy (JE) as a similarity functionals for multi-modal reconstruction in PET. Somayajula et al. [18] have shown that JE provides best results in terms of *bias/variance* characteristics when compared with other information theoretic similarity functionals. Tang and Rahmim [17] have shown that MAP reconstruction with a prior based on Joint Entropy improves the *bias/variance* characteristics when compared to other methods based on segmentation and extraction of boundaries; furthermore tuning of the algorithm depends only on one parameter - the importance of the prior - which can be optimized on synthetic data. However, as pointed out by Somayajula et al. [18], a spatial reordering of corresponding pairs of voxels in the two images would produce identical Joint Entropy values, thus Joint Entropy minimization presents multiple solutions. Because of multiplicity of the Joint Entropy, the *maximum a posteriori* activity estimate depends on the iterative scheme and the initial object estimate with which the algorithm begins. Usually this initial estimate is a simple uniform field, so the final image is biased towards uniformity, with lack of edge preservation. In order to overcome this drawback Somayajula et al. [18] introduced a scale-space approach, where activity with boundaries similar to the anatomical image are favored by minimizing the JE between the Laplacian of the images. However the introduction of additional parameters complicates the tuning of the reconstruction process.

The proposed work is based on the *Maximum A Posteriori* activity estimation with Joint Entropy based prior, as put forth by Somayajula et al. [18] and Tang and Rahmin [17], where Joint Entropy measures the similarity between activity and the anatomical image.

Here the Joint Entropy similarity functional is extended to Conditional Joint Entropy, in order to take into account dependence of the two images on the underlying tissue composition. Introducing a latent variable that takes values in a discrete set denoting tissue type, Joint Entropy conditioned to this variable

expresses the amount of information in common between the two images within each region of tissue. This method allows to model explicitly the probabilistic dependence between tissue fractional content and anatomical image, effectively adding information in the reconstruction process.

II. METHODS

Let the radio-pharmaceutical activity within the region of interest of the patient's body be a continuous function denoted by \tilde{y} . In order to readily discretize the reconstruction algorithm, it is convenient to imagine that the activity is in the first place discrete in space [24]. Let us approximate \tilde{y} by a set of point sources $y = y_b, b = 1, \dots, N_b$ displaced on a regular grid. Note at this point that an iterative algorithm will try to find a discrete activity that explains observations due in reality to a continuous distribution, causing a convergence problem often referred to as *dimensional instability* [4].

As each point source emits radiation at an average rate y_b proportional to the local density of radio-tracer and emission events in a same voxel are not time correlated, the number of emissions in the unit time is a Poisson distribution of expected value y_b . The geometry of the system and attenuation in the patient determine the probability p_{bd} that a photon emitted in b is detected at detector pixel d . From the sum property of the Poisson distribution, the photon count in d has Poisson *pdf* with expected value $\sum_b p_{bd} y_b$. Given activity y , the probability to observe counts z is

$$p(z|y) = \prod_{d=1}^{N_d} \mathcal{P}\left(\sum_b p_{bd} y_b, z_d\right) \quad (1)$$

When counts z are observed, the probability that they were *caused* by activity y is expressed by the Bayes formula, for the general case where spatial distributions of activity are not all considered equally probable *a priori*.

$$\hat{y} = \arg \max_y \frac{p(z|y)p(y)}{p(z)} \quad (2)$$

The prior probability $p(y)$ is defined in section B in terms of similarity of y with the anatomical image, after introduction of a probabilistic model of the MRI imaging system in section A. Section C shows a MAPEM algorithm for maximization of 2.

A. Tissue classification

Let $b \in \{1, 2, \dots, m\}$ index the m voxels of an anatomical image generated by n tissue types $k \in \{1, 2, \dots, n\}$. Assuming each tissue class in the image can be described as having Gaussian distributed intensities with mean μ_k and variance σ_k^2 , the probability to observe intensity x_b in a voxel b that is known to belong to tissue type k is

$$p(x_b|k) = \mathcal{G}\left(\frac{x_b - \mu_k}{\sigma_k^2}\right) \quad (3)$$

Thus, the probability to observe intensity x_b in a randomly selected voxel b is

$$p(x_b) = \sum_{k=1}^{N_k} p(x_b|k) p(k) = \sum_{k=1}^{N_k} p(k) \mathcal{G}\left(\frac{x_b - \mu_k}{\sigma_k^2}\right) \quad (4)$$

where $p(k)$ is a prior probability dependent on the overall mix of tissue types. Under the Bayesian formulation, the probability that the observed intensity x_b is determined by tissue type k is:

$$p(k|x_b) = \frac{p(x_b|k) p(k)}{p(x_b)} \quad (5)$$

Equations (3-5) can be regarded as a Mixture of Gaussians (MOG) model where p_k is normally referred to as the *mixing coefficient* and $p(k|x_b)$ as the *partial membership*.

Using Gaussian functions, $p(k|x_b)$ represents *complete-data* sufficient statistics as it is defined on a linear exponential base [25]. The Expectation Maximization algorithm can thus be adopted to iteratively update an estimate of $p(k)$, μ_k and σ_k from the image intensities [26].

(E)

$$\hat{p}(k|x_b)^{(n+1)} = \frac{\hat{p}(k)^{(n)} \mathcal{G}\left(\frac{x_b - \hat{\mu}_k^{(n)}}{\hat{\sigma}_k^{(n)}}\right)}{\sum_{k=1}^{N_k} \hat{p}(k)^{(n)} \mathcal{G}\left(\frac{x_b - \hat{\mu}_k^{(n)}}{\hat{\sigma}_k^{(n)}}\right)} \quad (6)$$

(M)

$$\hat{p}(k)^{(n+1)} = \frac{1}{N_b} \sum_{b=1}^{N_b} \hat{p}(k|x_b)^{(n+1)} \quad (7)$$

$$\hat{\mu}_k^{(n+1)} = \frac{1}{N_b} \sum_{b=1}^{N_b} \hat{p}(k|x_b)^{(n+1)} x_b \quad (8)$$

$$\hat{\sigma}_k^{(n+1)} = \frac{1}{N_b} \sum_{b=1}^{N_b} \hat{p}(k|x_b)^{(n+1)} (x_b - \hat{\mu}_k^{(n+1)})^2 \quad (9)$$

After convergence to the most likely parameters, equation 6 represents the probability (given the parameters) that k is the underlying tissue that generated intensity x_b in a voxel b .

B. Similarity functional

Let X and Y represent random processes that generate the anatomical and functional images respectively. Furthermore, assume that the relationship between the random processes X and Y is dependent on a hidden variable K representing a discrete underlying tissue type k . The probability to observe intensity x and activity y at a random voxel location b can be estimated by assembling a 2D histogram with the occurrence of (x, y) pairs from the images. Adopting a non-parametric estimation method based on Gaussian Parzen Windows, the estimate of the *joint pdf* $p(x, y)$ is differentiable with respect of the samples x_b and y_b [27]. Adopting a Gaussian kernel

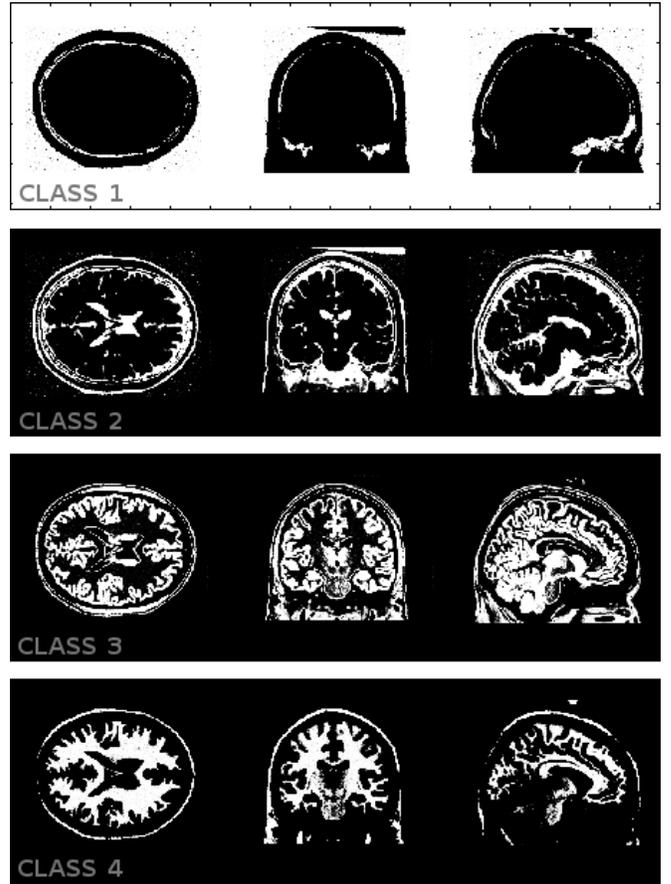


Fig. 2. Partial membership: in each of the four images, the intensity associated with a pixel represents the probability that that pixel *belongs* to the tissue associated with the image. Four tissue types are assumed.

with *bandwidth* (σ_x, σ_y) , the Parzen Windows estimate of the *joint pdf* is

$$\hat{p}(x, y) = \frac{1}{N_b} \sum_{b=1}^{N_b} \mathcal{G}\left(\frac{x - x_b}{\sigma_x^2}\right) \mathcal{G}\left(\frac{y - y_b}{\sigma_y^2}\right) \quad (10)$$

The previous section described a generative model of the MRI imaging system and an algorithm for estimation of the parameters of the model. The generative model provides, given the intensity of a voxel of the MRI image alone, an estimate of the probability that the voxel is an expression of any of the N_k tissue types k . The probability to observe (x, y) in a random voxel b , given that tissue in that voxel is known, can be expressed by means of Bayes formula:

$$p(x, y|k) = \frac{p(k|x, y) p(x, y)}{p(k)} \quad (11)$$

The following simplifying assumption is introduced:

$$p(k|x, y) = p(k|x) \quad (12)$$

Note that this simplification of the joint probability distribution does not assume absolute independence of tissue from activity, but independence conditionally to MRI image intensity. In other words knowledge of activity in a voxel as a first

approximation does not give any information about tissue if the MRI intensity in that voxel is known.

The probability that a voxel belongs to tissue k is estimated with the Gaussian Mixture model in (7). From (11),(12) and (7) the estimate of the conditional probability to extract (x, y) given tissue k is the following:

$$\hat{p}(x, y|k) = \frac{\mathcal{G}\left(\frac{x - \mu_k}{\sigma_k^2}\right)}{\sum_{k=1}^{N_k} p(k) \mathcal{G}\left(\frac{x - \mu_k}{\sigma_k^2}\right)} \hat{p}(x, y) \quad (13)$$

Joint Entropy quantifies the uncertainty associated with two random variables:

$$H(X, Y) = - \iint p(x, y) \log p(x, y) dx dy \quad (14)$$

Conditional Joint Entropy quantifies uncertainty associated with two random variables X and Y given that the value of another random variable K is known. While entropic functionals are defined for random variables taking discrete values, as in this case X and Y are continuous, the generalization of Conditional Joint Entropy to the domain of real numbers is considered here. If K takes values $k = 1, 2, \dots, N_k$ with probability $p(k)$, the Conditional Joint Entropy for continuous variables X and Y is defined as:

$$H(X, Y|K) = \sum_{k=1}^{N_k} p(k) H(X, Y|k) \quad (15)$$

$$H(X, Y|k) = - \iint p(x, y|k) \log p(x, y|k) dx dy \quad (16)$$

Let us approximate the integral with the sum of rectangular parallelepipeds of base $(\Delta x, \Delta y)$:

$$H(X, Y|k) \approx -\Delta x \Delta y \sum_{i,j}^M p(x_i, y_j|k) \log p(x_i, y_j|k) \quad (17)$$

The optimization algorithm described in the next session will require the gradient of $H(X, Y|K)$ with respect to samples y_b ; by the chain rule for differentiation:

$$\frac{\partial H(X, Y|K)}{\partial y_r} = \sum_{k=1}^{N_k} p(k) \frac{\partial H(X, Y|k)}{\partial y_r} \quad (18)$$

$$\frac{\partial H(X, Y|k)}{\partial y_r} = -\Delta x \Delta y \sum_{i,j}^M \left(1 + \log p(x_i, y_j|k)\right) \cdot \frac{\partial p(x_i, y_j|k)}{\partial y_r} \quad (19)$$

Differentiating $\hat{p}(x, y)$ with respect of samples y_r (10):

$$\frac{\partial p(x, y)}{\partial y_r} = \frac{1}{N_b} \mathcal{G}\left(\frac{x - x_r}{\sigma_x^2}\right) \mathcal{G}'\left(\frac{y - y_r}{\sigma_y^2}\right) \quad (20)$$

And from (13)

$$\frac{\partial p(x, y|k)}{\partial y_r} = \frac{\mathcal{G}\left(\frac{x - \mu_k}{\sigma_k^2}\right)}{\sum_{k=1}^{N_k} p(k) \mathcal{G}\left(\frac{x - \mu_k}{\sigma_k^2}\right)} \frac{\partial p(x, y)}{\partial y_r} \quad (21)$$

C. MAP reconstruction

The Conditional Joint Entropy similarity derived in the previous section is embedded in a Bayesian reconstruction framework by means of the Gibbs prior. Activity is then estimated by *maximum a posteriori* expectation maximization (MAPEM) with the one-step-late (OSL) approach proposed by Green for inclusion of the prior term [28].

Let $z = z_d$ represent the number of photons collected at each detector bin $d \in D$; the *maximum a posteriori* estimate of the activity is the argument that maximizes the Poisson likelihood and the prior activity distribution

$$\hat{y} = \arg \max_y \frac{p(z|y)p(y)}{p(z)} \quad (22)$$

The equivalence of Markov Random Field and Gibbs distribution expresses the prior probability of an activity configuration in terms of its entropy, as entropy represents an energy functional on the maximal complete subgraph of y .

$$p(y) = \frac{1}{Z} e^{-\beta V(y)} \quad (23)$$

where Z is a normalizing factor called the partition function. The optimization in 22 may be performed with a variety of methods, including gradient ascent and conjugate gradient, however the Expectation Maximization (EM) of Shepp Vardi [24] is adopted in the following as it intrinsically guarantees positivity of \hat{y} and has faster convergence rate than gradient ascent.

Exponential priors can be embedded in the EM algorithm with the One Step Late (OSL) approximation that was introduced by Green [28]. As discussed in [28], explicit formulation of the M step exists only if $V(y)$ is quadratic, however with the OSL approximation the gradient of $V(y)$ is computed at the previous EM step, allowing for any $V(y)$. Convergence is not guaranteed, but the method behaves well in practical cases.

$$\hat{y}_b^{(n+1)} = \hat{y}_b^{(n)} \frac{1}{\sum_{d=1}^{N_d} p_{bd} + \beta \frac{\partial V(y)}{\partial y_b} \Big|_{y_b^{(n)}}} \frac{\sum_{d=1}^{N_d} p_{bd} z_d}{\sum_{b'=1}^{N_b} p_{b'd} \hat{y}_{b'}^{(n)}} \quad (24)$$

Minimizing the Joint Entropy between MRI and activity, conditional to tissue:

$$p(y) = \frac{1}{Z} e^{-\beta \sum_{k=1}^{N_k} p(k) H(X, Y|k)} \quad (25)$$

In equation (24), the gradient of the energy function is:

$$\beta \frac{\partial V(y)}{\partial y_b} = \beta \sum_{k=1}^{N_k} p(k) \frac{\partial H(X, Y|k)}{\partial y_b} \quad (26)$$

Which is expressed by (19)(20)(21).

Note that the difference between this approach and the conventional Joint Entropy approach (see for example [18]) is the following:

$$\left. \frac{\partial V(y)}{\partial y_b} \right|_{CCJE} = \frac{\sum_{k=1}^{N_k} p(k) \mathcal{G}\left(\frac{x - \mu_k}{\sigma_k^2}\right)}{\sum_{k=1}^{N_k} p(k) \mathcal{G}\left(\frac{x - \mu_k}{\sigma_k^2}\right)} \left. \frac{\partial V(y)}{\partial y_b} \right|_{JE} \quad (27)$$

Where one can recognize from (3)(4)(5) the *partial membership* of x , i.e. the probability that a voxel with intensity x is a member of tissue k . This gives an intuitive interpretation of the method: Joint Entropy between MRI and activity is optimized concurrently in k directions, each weighted by the probability that the voxel belongs to a class of tissue k .

III. VALIDATION

Synthetic brain data from BrainWeb [9] was adopted in order to validate the proposed reconstruction algorithm and compare it with other methods. The MRI and functional imaging processes were decoupled by adopting a manually segmented anatomical image as ground truth of tissue composition. BrainWeb database provides a partial volume segmentation of the brain obtained from manual segmentation of 27 T1-weighted MRI images of the same subject: to each voxel is assigned a percentage of each type of tissue by averaging the 27 labels. This segmentation with partial volume information is considered the ground truth model of the brain tissue, which is used to simulate independently MRI imaging and radio-tracer activity. The MRI image was generated with the BrainWeb simulator, which uses first-principles modeling based on the Bloch equations to simulate the signal production, and realistically accounts for noise of the imaging system. The parameters of the simulator were set for T1-weighted imaging with noise standard deviation set at 3% of the brightest tissue and perfect uniformity of the magnetic field (in accordance with the simplistic MOG model). Activity of $^{99m}Tc - HMPAO$ was simulated by associating typical activity levels to different tissue types, proportionally to partial voxel occupation. Specifically the activity in gray matter was set to a value 4 times higher than in all other tissues. The total number of counts was set to 2.5 Million.

The SPECT imaging system was simulated by means of a rotation-based projector with realistic Collimator-Detector Response (CDR) [29] and applying Poisson noise to the projections. The parameters of the imaging system (Point Spread Function, size of the detector plane, distance of the detector from the axis of rotation, number of positions of

Detector width*	W	540 mm
Detector height	H	400 mm
Distance from axis	R	133 mm
Number of positions	N_P	120
Rotation step	θ	3 deg
Total counts	N_T	$2.5e^6$
PSF FWHM @ 0 mm	$FWHM0$	4.29 mm
PSF FWHM @ 20 mm	$FWHM20$	5.11 mm
PSF FWHM @ 40 mm	$FWHM40$	5.70 mm
PSF FWHM @ 80 mm	$FWHM80$	7.32 mm
PSF FWHM @ 160 mm	$FWHM160$	9.98 mm

TABLE I
PARAMETERS OF SPECT IMAGING SYSTEM, BASED ON GE INFINIA WITH LEHR COLLIMATOR. * DETECTOR WIDTH IS PARALLEL TO THE AXIS OF ROTATION OF THE GANTRY.

the Gamma Camera) were set to emulate a SPECT imaging system based on GE Infinia with Low Energy High Resolution (LEHR) collimator (Table III). Attenuation was not accounted for in the SPECT simulation and in the reconstruction process.

The MRI and activity images were defined on a cubic grid of $(128 \times 128 \times 128)$ voxels; figure 1 shows three slices of the synthetic MRI and activity images along the transverse, coronal and sagittal planes at the centre of each axis.

The Mixture of Gaussians Expectation Maximization algorithm (MOG-EM) for tissue classification was implemented according to equations (6)(7)(8)(9) in section II-A. The number of tissue types was assumed to be $N_k = 4$; the parameters μ_k were initialized to evenly spaced values in the range of intensity of the MRI image; σ_k was initialized to $1/N_k$ of the image intensity range; the mixing coefficients were initialized to $p(k) = 1/N_k \forall k \in N_k$. Then 30 iterations of the MOG-EM algorithm were performed and the resulting parameters were verified by visual assessment of the histogram of the image intensity, which presented peaks at each μ_k , with spread given by σ_k^2 and area $p(k)$. This inspection was performed because unconstrained MOG-EM is sensible to initial values of the parameters; however, as briefly discussed in section II-A, the algorithm could be made more robust and fully automatic with the use of statistical atlases of the brain, which is common practice in algorithms for image segmentation [30], [31].

Ideal sinogram data were generated from the activity phantom using the rotation-based projector; multiple instances of the sinogram were then generated by applying Poisson noise to the ideal projection.

Activity was estimated from each sinogram instance by running 100 iterations of the OSL MAP-EM algorithm of (24) with the CCJE prior of (26) and (19).

The following formulas for the gradient of the energy function were used in replacement of (26) and (19), for reconstruction

with conventional Joint Entropy:

$$\beta \frac{\partial V(y)}{\partial y_b} = \beta \frac{\partial H(X, Y)}{\partial y_r} \quad (28)$$

where

$$\frac{\partial H(X, Y)}{\partial y_r} = -\Delta x \Delta y \sum_{i,j}^M \left(1 + \log p(x_i, y_j)\right) \cdot \frac{\partial p(x_i, y_j)}{\partial y_r} \quad (29)$$

The prior based on Joint Entropy has a number of parameters: β , which controls the importance of the prior; the bandwidth of the Gaussian kernel for Parzen Windows estimation of the *joint pdf* σ_x and σ_y (10); the size of the discretization grid for the joint entropy Δx and Δy (17). The *bias/variance* curves were computed varying Δx and Δy , with an arbitrary value of $\sigma_x = \sigma_y$ and the activity estimate appeared to be largely independent upon Δx and Δy when the discretization grid for the *joint pdf* has more than about 200 points in x and y , for any value of β . A discretization grid of size 400×400 was chosen. For what concerns the choice of σ_x and σ_y , multiple reconstructions were performed again (with the grid 400×400) and the quality of the reconstruction in terms of *bias/variance* appeared to increase when σ_x and σ_y decrease, and then to abruptly decrease when they are down to the order of Δx and Δy . The following values were adopted $\sigma_x = 10 * \Delta x$ and $\sigma_y = 10 * \Delta y$. This procedure gave insight of the effect of the parameters and allowed us to consider β as the only parameter of the algorithm. For the Class Conditional Joint Entropy prior, the same parameters Δx , Δy , σ_x , σ_y of the Joint Entropy prior were used.

Figure 5 reports *bias/variance* at each iteration of OSL MAP-EM, for varying values of the hyper-parameter β . Figures 3 and 4 represent slices of the activity estimate after 100 iterations, with the value of β optimally chosen for each algorithm, as explained in the next section.

A. Bias/Variance analysis

With access to multiple instances of the sinogram data, a *bias/variance* characterization of the reconstruction algorithms was performed.

Specifically the activity was estimated from 15 realizations of the sinogram, indexed by $r \in N_r$ ($N_r = 15$).

Let us define the *bias* image (at reconstruction step n) as the voxel-wise mean difference from the true activity:

$$B_b^{(n)} \triangleq \frac{1}{N_r} \sum_{r=1}^{N_r} \left(\hat{y}_b^{[r](n)} - y_b^{true} \right) \quad (30)$$

The *variance* image is the voxel-wise variance of the reconstructed activity:

$$\sigma_b^2(n) \triangleq \frac{1}{N_r} \sum_{r=1}^{N_r} \left(\hat{y}_b^{[r](n)} - \bar{y}_b^{(n)} \right)^2 \quad (31)$$

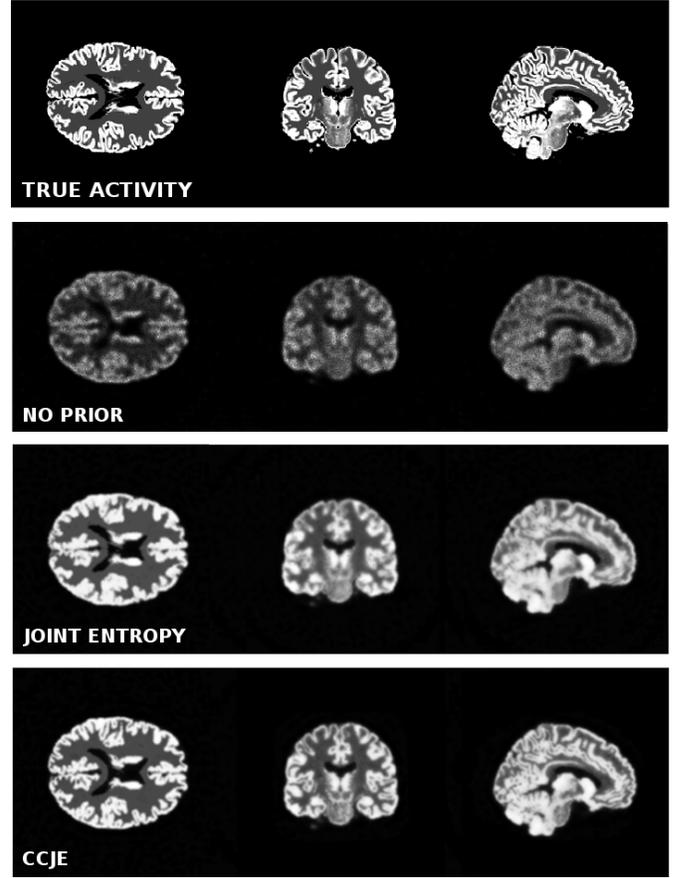


Fig. 3. Reconstructed activity distribution. From top to bottom: True activity; activity estimated with maximum likelihood (MLEM) reconstruction; activity estimated with maximum a posteriori (MAPEM) reconstruction with Joint Entropy prior; activity estimated with maximum a posteriori (MAPEM) reconstruction with Class Conditional Joint Entropy prior.



Fig. 4. Detail of reconstructed activity distribution: true activity (left), activity reconstructed with the prior based on Joint Entropy (middle), activity reconstructed with the prior based on Class Conditional Joint Entropy (right).

$$\bar{y}_b^{(n)} = \frac{1}{N_r} \sum_{r=1}^{N_r} \hat{y}_b^{[r](n)} \quad (32)$$

The ensemble *bias* and *variance* are defined as the average of the two measures over the voxel space:

$$BIAS^{(n)} = \sqrt{\frac{1}{N_b} \sum_{b=1}^{N_b} \left(B_b^{(n)} \right)^2} \quad (33)$$

$$VAR^{(n)} = \frac{1}{N_b} \sum_{b=1}^{N_b} \sigma_b^2(n) \quad (34)$$

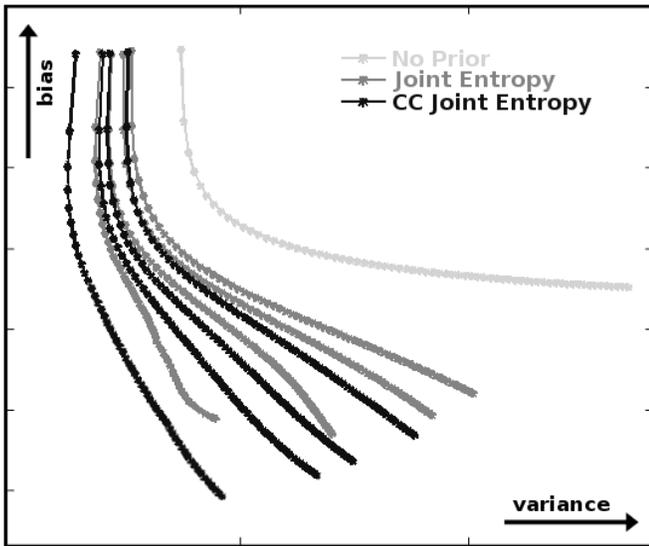


Fig. 5. *bias/variance* curves obtained with multiple reconstructions over 15 noise instances. Curves are plotted for varying values of the hyper-parameter associated with the priors. Considering the best parameters (leftmost curves) Class Conditional Joint Entropy achieves low *bias* (i.e. the activity estimation is similar to the phantom) while it controls the *variance* of the activity estimation (dimensional instability).

Figure 5 reports the ensemble *bias* and *variance* at each iteration of the reconstruction algorithms for unconstrained MLEM, Joint Entropy MAP-EM and CCJE MAP-EM. In presence of a prior, after a number of iterations the curves tend to converge, while with unconstrained MLEM the noise keeps increasing because of *dimensional instability*.

In order to compare the *bias/variance* curves of the different algorithms, the optimal values of the hyper-parameter for each reconstruction method was found by creating a set of curves with varying hyper-parameters (figure 5). The optimal value of β for each algorithm was chosen as that value that determines convergence of the *bias/variance* curve to a point closer to the axis origin (low *bias* and low *variance*).

The reconstruction with Class Conditional Joint Entropy produces images with lower *bias* when compared with the conventional Joint Entropy method, while the noise is approximately unaffected. Consistently with the *bias/variance* curves, the images of the activity reconstructed with the two algorithms (figures 3 and 4) show approximately the same level of random variations and the image obtained by CCJE is more similar to the true activity at visual inspection.

IV. CONCLUSION

When compared with the conventional Joint Entropy prior, the proposed method provides lower *bias* of the activity estimate. At visual inspection the images result more similar to the true activity and present discontinuities more consistent with the anatomical image.

Because of the simplification in (12), the estimation of tissue composition and the reconstruction are independent tasks as tissue composition is completely determined by the anatomical image. For this reason the method integrates well with any

algorithm for probabilistic tissue classification, and may take advantage of further assumptions about tissue composition, such as spatial correlation and prior knowledge of tissue distribution from a population average. Furthermore it can take into account a more realistic model of the MRI image formation, including local Markovian assumptions and bias field correction.

Further work has to be done in order to validate the robustness of the algorithm in case of mis-registration and to assess its effect on lesions that are not correlated to the anatomical image.

APPENDIX A

EDGE PRESERVING FEATURE: INTUITIVE INTERPRETATION

One drawback of classic Joint Entropy maximization is that the reconstructed activity can be continuous across boundaries of the anatomical image. The reason for this is found in the fact that multiple solutions to minimum JE are possible, as has been pointed out by Somayajula *et al.* [18]. Several methods have been developed to explicitly force edges in the anatomical image to appear in the activity; in the context of information theoretic functionals Somayajula *et al.* have proposed to add a term that maximizes the JE between the Laplacians of the two images.

Experiments (see image 4) have shown that the proposed method provides implicit edge preservation; the following presents an intuitive explanation of the implicit edge preserving feature.

Consider n samples $\bar{x} = x_1, x_2, \dots, x_n$ of a random variable X , with unknown probability distribution $p(x)$; The following experiment is set up: the samples x_1, x_2, \dots, x_n are updated by gradient ascent in order to maximize the entropy $H(X)$ associated with X ; $p(x)$ is estimated from the samples by Parzen Windows with Gaussian kernels. Suppose, for simplicity, that X is discrete and binary, with outcomes x_0 and x_1 . The histogram of \bar{x} and the Parzen Window estimate of the *pdf* $p(x)$ are pictured in 6 (first and second from the top) for a random initialization of \bar{x} .

$$H(X) \approx \Delta x \sum_{i=1}^M p(x_i) \log p(x_i) \quad (35)$$

$$\hat{p}(x) = \frac{1}{N} \sum_{b=1}^N \mathcal{G} \left(\frac{x - x_b}{\sigma_x^2} \right) \quad (36)$$

$$\frac{\partial p(x)}{\partial x_r} = \frac{1}{N} \mathcal{G}' \left(\frac{x - x_r}{\sigma_x^2} \right) \quad (37)$$

$$x_r^{(n+1)} = x_r^{(n)} - \frac{\partial H(x)}{\partial x_r} \quad (38)$$

$$\frac{\partial H(x)}{\partial x_r} = \Delta x \sum_{i=1}^M (1 + \log p(x_i)) \frac{\partial p(x_i)}{\partial x_r} \quad (39)$$

The derivative of $H(X)$ with respect of \hat{x}_r drives the change of \hat{x}_r at each step of the ascent algorithm in order to maximize $H(X)$. Figure 6 (third and fourth from the top) shows the two

terms of (39) before the summation. The second term has an odd symmetry and the first term is almost symmetric around x_0 except for a tail due to x_1 , the same holds for x_1 . Figure 6 (bottom) shows the resulting gradient of the Entropy with respect of x_0 and x_1 : the two Gaussians attract each other, but the further away the less they attract as the tails become smaller and smaller; when x_0 and x_1 coincide, the gradient is zero. One could say that maximization of the entropy *clusters* the probability density function as samples that are far stay far and samples that are close attract to each other.

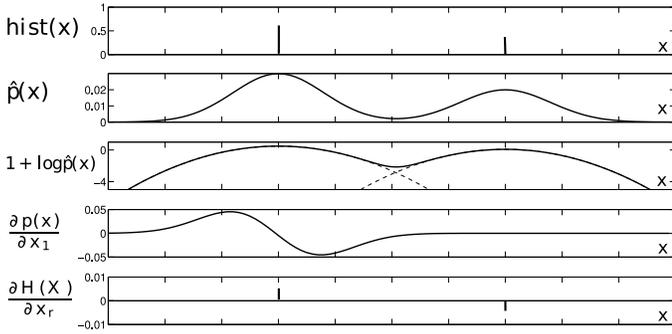


Fig. 6. Example of Entropy optimization.

In the 2-dimensional case, random variables X and Y represent the intensity of two images, respectively functional and anatomical, JE is defined in (16)(17). The same experiment as in the 1-dimensional case is set up, where the samples of Y are now controlled in order to maximize the JE. For the two images in figure A the histogram has 3 deltas and the Parzen Window estimate of the *joint pdf* $p(x, y)$ is reported in figure A (top). The Gaussians, as in the 1-D example, will attract each other; but as \hat{x} does not change, the Gaussians in figure move only along the y axis. After convergence the *joint pdf* and the image \hat{y} look like in figure A (middle). B and C have attracted to one another until the two areas in the image have assumed the same value, while A has been attracted by C (and vice versa).

The proposed method optimizes concurrently multiple Conditional Joint Entropy terms; for each, the gradient of JE is multiplied by a function that depends on x only (21). This function expresses the probability that the activity being optimized belongs to each of the N_k classes. The anatomical image in the example is represented by two classes. Figure A (bottom) reports the two probability functions f_1 and f_2 with an arbitrary value of σ_k . Now the optimization of JE conditional to tissue $k = 2$ affects less A and attracts B and C while the optimization of JE conditional to $k = 1$ affects less C . B and C still attract to one another, but A is not attracted by C .

REFERENCES

[1] J.E. Bowsher, L.W. Hedlund, T.G. Turkington, G. Akabani, a. Badae, W.C. Kurylo, C.T. Wheeler, G.P. Cofer, M.W. Dewhirst, and G.a. Johnson. Utilizing MRI Information to Estimate F18-FDG Distributions in Rat Flank Tumors. *IEEE Nuclear Science Symposium Conference Record*, 00(C):2488–2492, 2004.

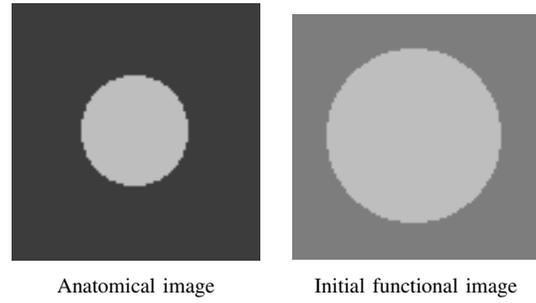


Fig. 7. Example of JE optimization, initial Anatomical and functional images

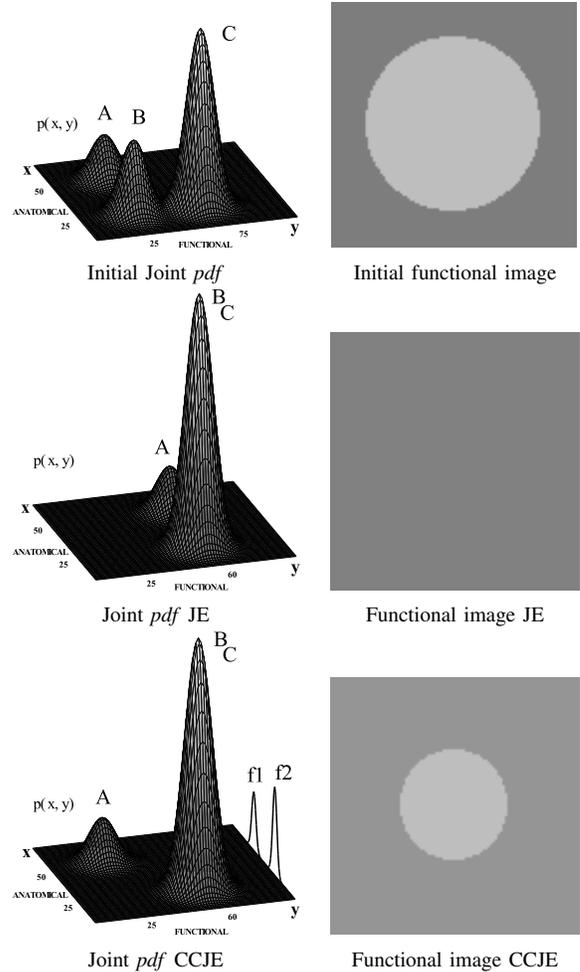


Fig. 8. Implicit edge preserving feature of Conditional Joint Entropy anatomical prior

[2] T.G. Turkington, R.J. Jaszczak, C.A. Pelizzari, C.C. Harris, J.R. MacFall, J.M. Hoffman, and R.E. Coleman. Accuracy of registration of PET, SPECT and MR images of a brain phantom. *Journal of Nuclear Medicine*, 34(9):1587, 1993.

[3] Gene Gindi, Mindy Lee, Anand Rangarajan, and IG Zubal. Bayesian reconstruction of functional images using anatomical information as priors. *IEEE Transactions on Medical Imaging*, 12(4):670–680, 1993.

[4] P.P. Mondal and K. Rajan. Image reconstruction by conditional entropy maximisation for PET system. *IEEE Proceedings Vision, Image and Signal Processing*, 151(5):345–352, 2004.

[5] D.L. Snyder, M.I. Miller, L.J. Thomas, and D.G. Politte. Noise and edge artifacts in maximum-likelihood reconstructions for emission tomography. *IEEE Transactions on Medical Imaging*, 6(3):228–38, January 1987.

- [6] D.L. Snyder and M.I. Miller. The use of sieves to stabilize images produced with the EM algorithm for emission tomography. *IEEE Transactions on Nuclear Science*, 274(1-2):127–132, September 1985.
- [7] M.P. Lichy, P. Aschoff, C. Pfannenber, and S. Heinz-Peter. Case Reports: Tumor Detection by Diffusion-Weighted MRI and ADC-Mapping with Correlation to PET/CT Results. *medical.siemens.com*, 2:47–51, 2009.
- [8] C. Comtat, P.E. Kinahan, J.A. Fessler, T. Beyer, D.W. Townsend, M. Defrise, and C. Michel. Clinically feasible reconstruction of 3D whole-body PET/CT data using blurred anatomical labels. *Physics in Medicine and Biology*, 47:1, 2002.
- [9] BrainWeb. Brainweb: Simulated brain database. <http://www.bic.mni.mcgill.ca/brain-web/>.
- [10] G.K. Von Schulthess and H.W. Schlemmer. A look ahead: PET/MR versus PET/CT. *European Journal of Nuclear Medicine and Molecular Imaging*, 36 Suppl 1(December 2008):S3–9, March 2009.
- [11] B.J. Pichler, M.S. Judenhofer, and H.F. Wehrl. PET/MRI hybrid imaging: devices and initial results. *European Radiology*, 18(6):1077–86, June 2008.
- [12] S.R. Cherry. In vivo molecular and genomic imaging: new challenges for imaging physics. *Physics in Medicine and Biology*, 49(3):R13–R48, February 2004.
- [13] Z.H. Cho, Y.D. Son, H.K. Kim, K.N. Kim, S.H. Oh, J.Y. Han, I.K. Hong, and Y.B. Kim. A fusion PET-MRI system with a high-resolution research tomograph-PET and ultra-high field 7.0 T-MRI for the molecular-genetic imaging of the brain. *Proteomics*, 8(6):1302–1323, March 2008.
- [14] K. Baete, J. Nuyts, K. Van Laere, W. Van Paesschen, S. Ceysens, L. De Ceuninck, O. Gheysens, A. Kelles, J. Van den Eynden, P. Suetens, and P. Dupont. Evaluation of anatomy based reconstruction for partial volume correction in brain FDG-PET. *Neuroimage*, 23(1):305–17, September 2004.
- [15] S. Sastry and R.E. Carson. Multimodality Bayesian algorithm for image reconstruction in positron emission tomography: a tissue composition model. *IEEE Transactions on Medical Imaging*, 16(6):750–61, December 1997.
- [16] R. Leahy and X. Yan. Incorporation of Anatomical MR Data for Improved Functional Imaging with PET. In *Information Processing in Medical Imaging*, pages 105–120. Springer, 1991.
- [17] J. Tang and A. Rahmim. Bayesian PET image reconstruction. *Physics in Medicine and Biology*, 54(2):7063–7075, May 2009.
- [18] S. Somayajula, A. Rangarajan, and R.M. Leahy. PET image reconstruction using anatomical information through mutual information based priors: a scale space approach. *International Symposium on Biomedical Imaging*, pages 165–168, April 2007.
- [19] A. Rangarajan, I.T. Hsiao, and G. Gindi. A Bayesian joint mixture framework for the integration of anatomical information in functional image reconstruction. *Journal of Mathematical Imaging and Vision*, 12(3):199–217, 2000.
- [20] D. Van De Sompel and M. Brady. Robust Joint Entropy Regularization of Limited View Transmission Tomography Using Gaussian Approximations to the Joint Histogram. *Information Processing in Medical Imaging*, 21:638–50, January 2009.
- [21] C. Studholme, D.L.G. Hill, and D.J. Hawkes. Automated three-dimensional registration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxel similarity measures. *Medical Physics*, 24(1):25, January 1997.
- [22] W.R. Crum, T. Hartkens, and D.L.G. Hill. Non-rigid image registration: theory and practice. *British Journal of Radiology*, 77(Special Issue 2):S140, December 2004.
- [23] A. Atre, K. Vunckx, K. Baete, A. Reilhac, and J. Nuyts. Evaluation of different MRI-based anatomical priors for PET brain imaging. In *IEEE Nuclear Science Symposium Conference Record*, pages 1–7, Orlando, October 2009.
- [24] L.A. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Transactions on Medical Imaging*, 1(2):113–122, 1982.
- [25] A.P. Dempster, N.M. Laird, D.B. Rubin, and Others. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.
- [26] R.A. Redner. Mixture densities, maximum likelihood and the EM algorithm. *SIAM review*, 26(2):195–239, 1984.
- [27] B.W. Silverman. *Density Estimation for statistics and data analysis*. Chapman & Hall, New York, 1986.
- [28] P.J. Green. Bayesian Reconstructions From Emission Tomography Data Using a Modified EM Algorithm. *IEEE Transactions on Medical Imaging*, 9(1):84–93, 1990.
- [29] S. Pedemonte, A. Bousse, K. Erlandsson, M. Modat, S.R. Arridge, B.F. Hutton, and S. Ourselin. GPU Accelerated Rotation-Based Emission Tomography Reconstruction. In *IEEE Nuclear Science Symposium Conference Record*, Nov 2010.
- [30] K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens. Automated model-based tissue classification of MR images of the brain. *IEEE Transactions on Medical Imaging*, 18(10):897–908, Oct 1999.
- [31] J. Ashburner and K.J. Friston. Unified segmentation. *Neuroimage*, 26(3):839–51, July 2005.