The Use of Mutual Information and Joint Entropy for Anatomical Priors in Emission Tomography.

Johan Nuyts, *Member, IEEE*

***Abstract*—This paper studies the use of mutual information and joint entropy to define anatomical priors for maximum-a-posteriori (MAP) reconstruction in emission tomography. Other groups have used mutual information for this purpose, and reported promising results. Simple simulation studies with the "isolated" prior distribution reveal that mutual information may introduce bias, because of a repelling effect between intensity clusters in the marginal histogram. Deleting the terms involving the marginal histograms leads to the joint entropy prior. A gradient ascent MAP-reconstruction algorithm with this prior is described. Its performance is studied with simulation experiments and illustrated on a two sets of patient data: a whole body PET/CT scan and a PET brain scan, combined with the corresponding MRI-scan using off-line registration. The joint entropy seems to be a useful function for defining anatomical priors that do not require explicit segmentation of the anatomical image.**

I. Introduction

Many groups have studied the use of anatomical side information during emission reconstruction, aiming at improved image quality [1]-[10]. The side information is typically obtained from a spatially aligned MR or CT image. In many of these methods, the anatomical image must be segmented first in order to produce a set of labels or boundary elements. If the labels are appropriate and if the image registration is sufficiently accurate, this can lead to a very effective partial volume correction. The effect on lesion detection is less obvious, the results of several recent studies are discussed in [10]: in many (but not all) studies, the anatomical prior improved lesion detection, if (at least part of) the lesion boundaries were present in the anatomical information. In contrast, the use of organ boundaries, not including tight lesion boundaries, does not seem to be of much help.

For some applications, in particular human brain imaging with MRI, excellent segmentation algorithms are available. For other applications, the segmentation problem remains to be solved. For those cases, anatomical priors requiring little or no segmentation of the anatomy have been proposed [4], [7]. One approach is to define a similarity measure that works directly on the image intensities. Mutual information is such a measure, and it was found to be very effective in intensity based inter-modality image registration [11]. A few groups have used mutual information to define anatomical priors with promising results [6], [7].

In [7], mutual information was used to estimate the similarity between different types of image features, including the intensity, the gradients and local mean. Here, the study is restricted to the mutual information and joint entropy of the voxel intensities only. In addition, it is assumed that the anatomical image has been perfectly aligned.

To evaluate the behaviour of a new prior, it is often instructive to examine solutions allowed by this prior (independent from the likelihood). Applying a gradient ascent algorithm maximizing the prior allows to rapidly verify if a particular image corresponds to a maximum of the prior (the algorithm will change the image if it does not). Similarly, random solutions can be obtained by starting from random images. The results of such analysis indicated that a prior based on mutual information may introduce bias. Using the joint entropy instead of the mutual information was found to solve the problem.

This paper is organized as follows. In section II, mutual information and joint entropy are briefly reviewed, and a simulation experiment is presented showing images maximizing either the mutual entropy or the joint joint entropy. The final subsection proposes a gradient ascent algorithm for MAP reconstruction with these priors. Section III presents three experiments. The first experiment shows anecdotal MAP reconstructions with the joint entropy prior, the relative difference smoothing prior [12] and a combination of both priors. The second experiment presents a preliminary bias-noise analysis using a simple, piecewise constant simulated image. As a third experiment, two clinical images are reconstructed using the joint entropy prior.

II. Theory

A. Mutual information (MI) and joint entropy (JE)

Assume that A and B are two spatially aligned images, each of them having N voxels. The joint entropy (JE) of the two images is then defined as

$$H(A,B) = -\sum_{ab} p_{ab} \ln p_{ab}, \tag{1}$$

where a and b run over all intensities of image A and B, respectively, and p_{ab} equals the fraction of voxels j for which $A_j = a$ and $B_j = b$. It follows that $\sum_{ab} p_{ab} = 1$. The entropies of the individual images are defined similarly. Using $p_a^A = \sum_b p_{ab}$ and $p_b^B = \sum_a p_{ab}$ we have

$$H(A) = -\sum_{a} p_a^A \ln p_a^A. \tag{2}$$

The mutual information (MI) can then be written as

$$I(A,B) = H(A) + H(B) - H(A,B). \tag{3}$$

Nuclear Medicine, K.U.Leuven, B-3000 Leuven, Belgium. (e-mail: Johan.Nuyts@uz.kuleuven.be)

Figure 1. The Shepp-Logan phantom is used as the anatomical prior (right), a uniform image is used as the initial image and the prior distribution is maximized. The MI-prior diverges to an image with anatomical detail and arbitrary intensity (center), the JE-prior has a local maximum for the uniform image (left).

Figure 2. Joint histograms of the uniform image maximising JE (left) and of the non-uniform image maximising MI (right) of fig 1. The x-as is the "reconstructed" image intensity, the y-axis is the anatomical image intensity.

The (joint) entropy can be considered as a measure of the number of bits required to represent the image(s) with optimal coding of the image intensities. Hence $I(A,B)$ quantifies the storage gain obtained by coding both images together, which should be higher when the images are more similar.

In the rest of the text, image A will be the emission image to be optimized, and image B will be the fixed anatomical image.

B. Images maximizing MI and JE

A gradient ascent algorithm was implemented for the MI and JE priors (details are given in section II.C). For this experiment 20 bins and a Gaussian Parzen window with $\sigma = 0.7$ pixels were used for computing the (joint) histograms. The Shepp-Logan phantom was chosen as the anatomical image. The initial image was uniform except for a small random perturbation (uniform random noise between 0 and 1% of the intensity was added to every voxel). For most applications in PET and SPECT reconstruction, it is desirable that the uniform image were a maximiser of the prior.

Two hundred iterations of the gradient ascent algorithm were applied for each of the priors. Fig 1 shows the results. The image remained essentially unchanged for the JE-prior (the intensity range was reduced from 1 to 0.2%), indicating that the uniform image is a solution. In contrast, with the MI-prior, an image with the same structure as the anatomical image was produced. Its intensity is arbitrary; every new run yielded a different intensity distribution.

Because the anatomical image B is fixed, the MI-prior tries to maximise $H(A)$ and to minimize $H(A,B)$. $H(A,B)$ is smaller if the distribution along each row of the joint histogram is combined in narrower and fewer clusters. Because A is uniform, $H(A,B)$ is already minimal. However, $H(A)$ can still be maximised by making the marginal histogram of A less uniform. This corresponds to distributing the clusters in the joint histogram along the rows (fig 2); structures with different intensity in B get a different intensity in A as well (fig 1).

In image registration problems, $H(A)$ and $H(B)$ are essential to ensure that MI cannot be maximized by reducing the overlap between A and B. However, when used as an anatomical prior in image reconstruction, these entropy terms may introduce bias by shifting the intensities of different anatomical structures. Deleting these terms yields the joint entropy prior. This prior reaches a maximum for any image that can be obtained as a gray scale transformation of the anatomical image, as was verified with simulation experiments (not shown). The rest of this paper focuses on the joint entropy prior. The logarithm of the joint entropy prior is defined as $-\beta H(A,B)$, where β is the weight assigned to the prior.

C. Gradient ascent algorithm for JE

The joint histogram was computed using a Gaussian Parzen window for the intensities of A, and nearest neighbor interpolation for the intensities of B. This ensures that $H(A,B)$ is differentiable with respect to the intensities of A. This yields

$$p_{ab} = \sum_j \Phi(A_j - a, B_j - b) \text{ with } \Phi(x,y) = \text{Gauss}(x,\sigma)\delta_y,$$

where $\text{Gauss}(x,\sigma)$ is a Gaussian with mean x and standard deviation σ and δ_y is the Kronecker delta. For the gradient, one obtains

$$\frac{\partial H(A,B)}{\partial A_j} = \sum_{ab} (\ln p_{ab} + 1)\Phi(A_j - a, B_j - b)\frac{a - A_j}{\sigma^2}. \quad (4)$$

The final gradient ascent algorithm was then defined to be

$$A_j^{\text{new}} = A_j - \frac{\partial H(A,B)/\partial A_j}{\sum_{ab} \ln p_{ab} \Phi(A_j - a, B_j - b)/\sigma^2}. \quad (5)$$

The denominator was obtained using a very crude but conservative approximation of the second derivative. This algorithm is heuristic, but was found to converge in all experiments. This expression was combined with the gradient of the likelihood as in [12] to produce the MAP reconstruction algorithm, which maximizes

$$\text{posterior}(A|Y,B) = L(A,Y) - \beta H(A,B) \quad (6)$$

where A is the emission image to be reconstructed, L is the likelihood, Y is the measured emission sinogram, and B is the anatomical image.

III. MAP RECONSTRUCTION EXPERIMENTS

Because $-H(A,B)$ is a similarity measure, the joint entropy prior can be used to transfer information from the anatomical image into the emission image during reconstruction. However, because the joint entropy is computed from the (joint) grey level histogram, it operates on individual pixel intensities, ignoring the expected correlations between neighboring pixel values. For that reason, it seems useful to combine the JE-prior with a smoothing prior. In the following, the relative difference prior [12] was used as the smoothing prior. This prior is similar to the well-known quadratic prior, except that its smoothing strength is a function of the relative difference between neighboring pixels, and not of the absolute difference.

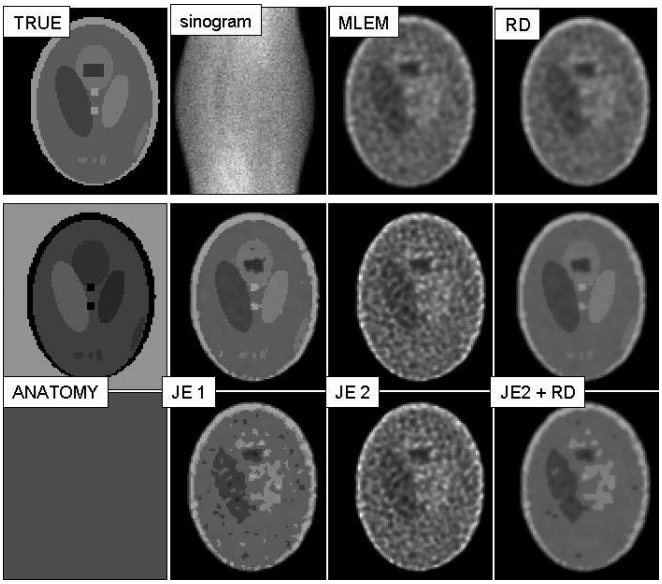

Figure 3. The first row shows the true tracer distribution, the simulated noisy sinogram, and the reconstructions using post-smoothed ML and MAP with the relative difference prior (RD). The second and third row show experiments with an anatomical prior: from left to right: the anatomical image, MAP-reconstruction using the JE-prior with a high weight (JE1), MAP-reconstruction using the JE-prior with a small weight (JE2), MAP-reconstruction using the combination of the JE2 and RD priors.

A. Simulation experiment 1.

A phantom image similar to the Shepp-Logan phantom was used as the true tracer distribution (100 × 100 pixels), and a sinogram (120 projection angles over 180 degrees) was computed using parallel beam projection with finite resolution (2 pixels full width at half-maximum) and Poisson noise was added. Other effects such as attenuation and scatter were not simulated.

Figure 3 shows the image and the noisy sinogram. Also shown are the "standard" maximum-likelihood(ML) reconstruction and a MAP reconstruction using the relative difference (RD) prior. Two different anatomical images were used to evaluate the prior. First an image with nearly matched anatomy was used, the only mismatch was a rectangular area of reduced activity that does not correspond to any anatomical boundary (fig 3). Three MAP-reconstructions with anatomical prior were computed: the first one used the JE-prior with a high weight (β in eq (6)), the second one used the JE-prior with a low weight, and the third one used the combination of the RD-prior and the JE-prior with the low weight. This experiment was repeated, using a uniform image as the anatomical image. The anatomical images and the corresponding reconstructions are shown in fig 3.

It appears that the JE-prior indeed improves the image quality by favoring smoothness within anatomical regions. Interestingly, the combination of the JE-prior and the RD-prior seems to perform better than each of the priors alone. This was also the case in the second experiment, where the anatomical image was entirely uniform.

Figure 4. Shepp phantom used for the bias-noise analysis. Left: emission image with the three contours; center: the emission image; right: the anatomical image.

B. Simulation experiment 2: bias-noise analysis

A preliminary bias-noise analysis was carried out using a simple simulation experiment. Fig 4 shows the emission and anatomical images. The same image dimensions, sinogram size and projector were used as in the first experiment. The image was projected to obtain the noise-free sinogram, and 200 Poisson noise realisations of the sinogram were computed. All the sinograms were reconstructed with various MAP algorithms.

The mean activity in the three regions indicated in fig 4 was studied. Bias and noise were estimated as follows

$$R_i(x) = \sum_{j \in \mathrm{ROI}_i} x_j \tag{7}$$

$$\mathrm{bias} = (R_i(\lambda_M) - R_i(\lambda_T))/R_i(\lambda_T) \tag{8}$$

$$\mathrm{noise} = \frac{1}{R_i(\lambda_T)} \sqrt{\frac{1}{N-1} \sum_{n=1}^{N} (R_i(\lambda_{S_n}) - R_i(\lambda_M))^2} \tag{9}$$

where λ_T is the true image, λ_{S_n} is the reconstruction of the nth noisy sinogram, $N = 200$ is the number of noisy sinograms, λ_M is the mean image computed from the 200 reconstructions, and ROI_i is one of the three regions-of-interest shown in fig 4. ROI_1 and ROI_2 correspond to structures with matching boundaries in the emission and anatomical images. ROI_3 did not have matching boundaries in the anatomical image. The activity in ROI_1 and ROI_3 was twice the background activity, and in RIO_2 three times the background activity.

MAP-reconstructions with five different priors were evaluated: the RD-prior, the JE-prior, and 3 combinations of JE and RD, where the relative weight of the RD-prior was multiplied with 1, 3 and 10. For the JE-prior, the weight of the JE-prior was gradually increased during iterations, to avoid undesired local minima. It was observed that without this strategy the bias increases, indicating that if the prior is immediately applied at its full strength, the MAP-reconstruction indeed tends to be trapped in a local minimum.

The resulting bias-noise curves are shown in figure 5. As expected, smoothing with the RD-prior reduces the noise but introduces negative bias (because the regions are more active than the background). For regions 1 and 2, the JE-prior achieves lower bias than the RD-prior, but the noise remains relatively high. The combinations of the RD-prior and the JE-prior result in improved bias-noise performance when

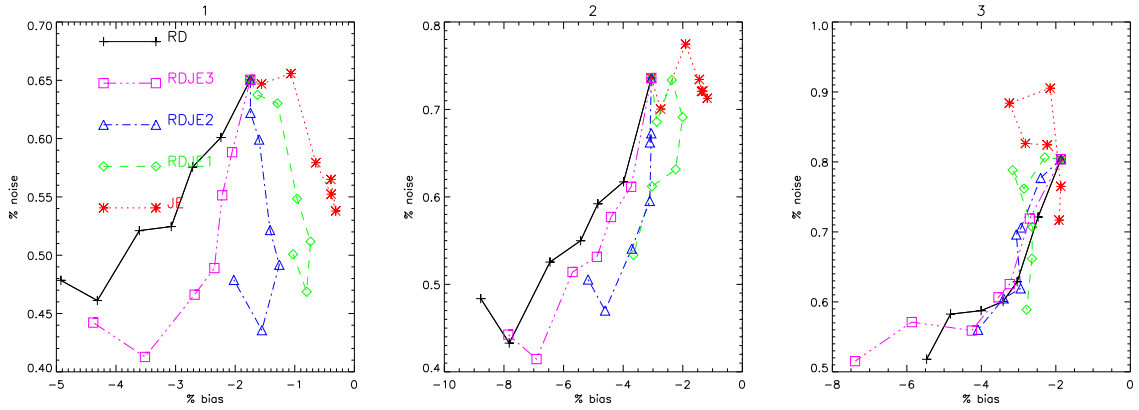

Figure 5. Bias-noise curves for the regions 1 (left), 2 (center) and 3 (right) respectively, obtained with different reconstructions: RD = relative difference prior, JE = joint entropy prior, RDJE1, RDJE2, RDJE3: combination of RD and JE, with a low (RDJE1), medium (RDJE2) and high (RDJE3) weight for the RD-prior.

compared to the RD-prior alone. For region 3, which does not have an anatomical counterpart, the performance of all priors is similar.

C. Patient images

Fig 6 shows a whole body PET reconstruction using the CT as the anatomical reference. Before binning the histograms, the CT was windowed to ensure that the histogram bins would contain useful tissue information. The study was acquired on a Siemens Biograph 16 PET/CT scanner, the MAP PET reconstruction was done using a fully 3D projector/backprojector, assuming a shift invariant resolution of 4 mm. When the JE-prior was used alone, it tended to produce some isolated spots due to sinogram noise. The result shown in fig 6 was obtained by combining the joint entropy prior and the relative difference prior.

Fig. 7 shows a PET reconstruction of a (low noise) brain scan using only the JE-prior based on an MRI image. The PET scan was acquired on a CTI/Siemens ECAT HR+, and the ML and MAP PET reconstruction was done using fully 3D (back)projection, assuming a shift invariant resolution of 5 mm.

IV. Discussion

The simulation experiments illustrated in figures 1 and 2 revealed that mutual information may not be the best function for an anatomical prior. The highest mutual information is obtained when the joint histogram is optimally clustered (minimal joint entropy), while at the same time the entropy of the emission image is as high as possible. The latter is obtained when each of the "anatomical structures" (here defined as a set of all pixels with a particular grey value in the anatomical image) is assigned a different tracer uptake in the emission image. As a result, the uptake values in different anatomical regions have a repelling effect on each other. This effect is expected to produce bias, unless the tracer uptake in each identifiable anatomical structure would indeed be sufficiently different from that of all others. In contrast, the joint entropy is maximized when for each anatomical structure a uniform tracer uptake is found in the emission image. As a result, the two images (left and center) in fig 1 have the same joint entropy.

As mentioned above, the joint entropy considers as a single "anatomical structure" every group of all pixels having a particular intensity in the anatomical image (or more precisely, all pixels that have been assigned to the same bin in the (joint) histogram). Consequently, it ignores the expected correlations between neighboring pixels. By combining the JE-prior with a smoothing prior such as the RD-prior, this additional prior knowledge can be supplied. Because we assume that no segmentation is available, the RD-prior is unrestricted and will tend to smooth over anatomical boundaries. This tendency will be counteracted by the JE-prior, which encourages uniformity within the anatomical structures. This is illustrated in fig 3, where sharp boundaries are seen for the joint prior. Using a smoothing prior with improved tolerance for edges may further improve the compatibility of both priors.

The bias-noise experiment showed that for a very simple, piecewise constant tracer uptake, the use of the anatomical priors resulted in reduced bias for regions with corresponding boundaries in the anatomical image. For the hot region 3, which had no corresponding structure in the anatomical image, the performance of the anatomical priors was similar to that of the smoothing prior.

Unfortunately, the JE-prior introduces local maxima. In particular when a high weight is assigned to the prior, the reconstruction may converge to an undesired local maximum of the posterior. Gradually increasing the weight of the prior until it reaches its final value in the last iterations reduced the problem.

The effect of the JE-prior was illustrated in two patient images. No quantitative analysis could be performed, because in these images the ground truth is unknown. Visual inspection indicated that the noise in the PET images was considerably reduced, the significant tracer uptake patterns were preserved, while the agreement with the anatomical image was improved. In brain images, the correspondance between tracer uptake and anatomy from the MRI image is higher than in typical PET/CT images. In this case, the JE-prior improved the contrast in the

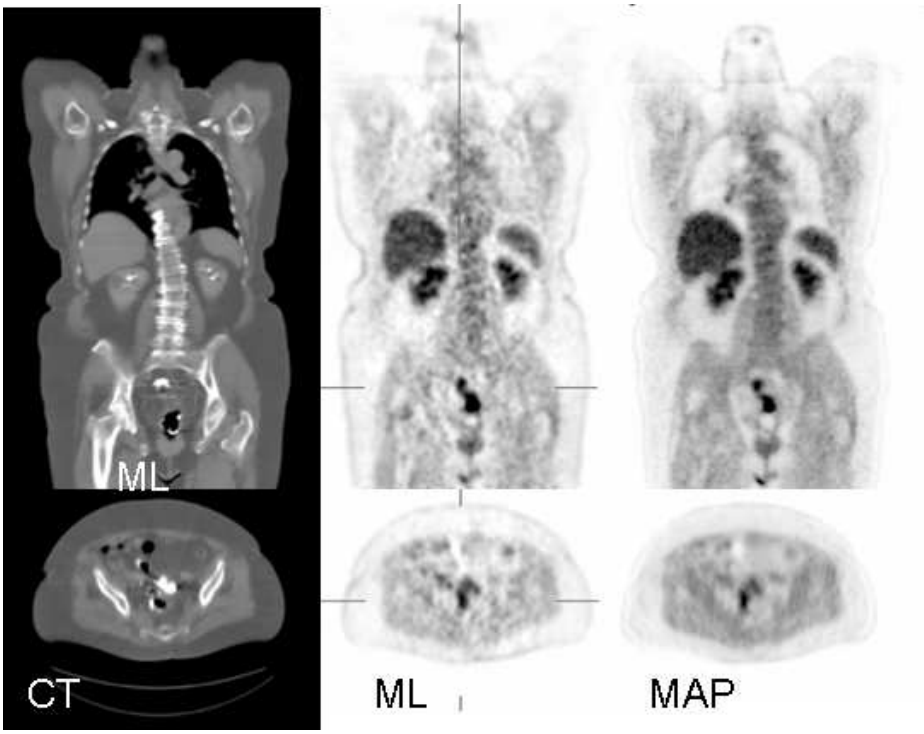

Figure 6. PET-CT result. Left: CT, center: post-smoothed ML reconstruction, right: MAP with combination of JE-prior and relative difference prior.

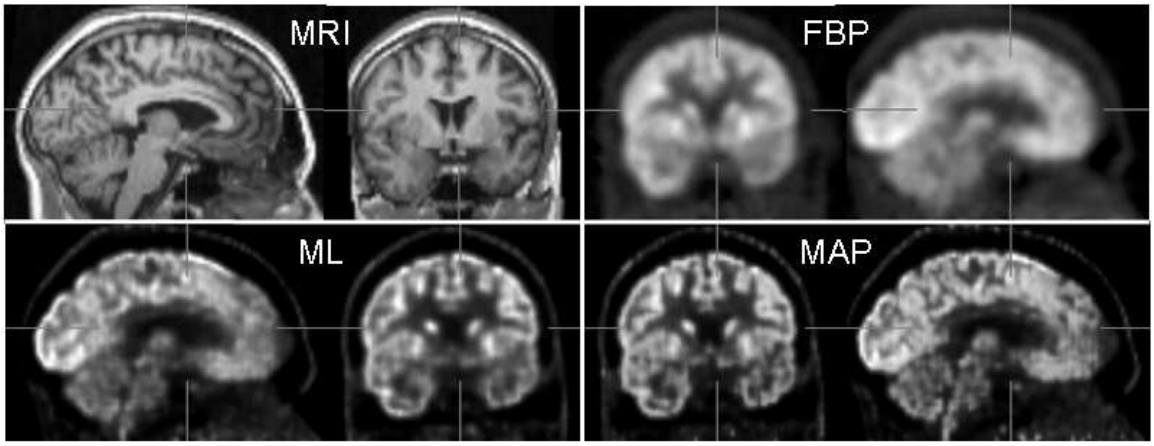

Figure 7. PET-MR result: sagittal and coronal slices of MRI images and PET reconstruction with FBP, ML and MAP (using JE-prior).

PET image significantly, which is indicative for partial volume correction. This could be quantitatively verified with Monte Carlo simulations using realistic MRI images and PET tracer distributions.

In CT images, the relevant information is concentrated in relatively small intensity ranges. Consequently, it is helpful to transform the intensities to ensure that this information will be represented by different histogram bins. In the patient study, this was done by applying a simple grey value window. Other, possibly non-linear grey value transformations may produce better results. Similarly, additional experiments will be needed to determine optimal values for the number of bins and the width of the Parzen window. However, it is encouraging already that after very limited parameter tuning, promising results have been obtained in simulations as well as in patient images.

V. Conclusion

This study shows that mutual information (applied to image intensities) is not optimal for the incorporation of anatomical side information in maximum-a-posteriori reconstruction. Deleting the terms involving the marginal histograms yields the joint entropy prior, which performed well in the simulation experiments and the two patient studies. In a preliminary bias-noise study using a very simple phantom, the prior improved the quantitative accuracy of the reconstructions. In the simulations, good results were obtained, and no deterioration of quantitative accuracy was observed when the anatomy did not match with the regions of increased tracer uptake.

VI. Acknowledgement

The author is grateful to Kathleen Vunckx for proofreading the manuscript.

References

[1] B Lipinski, H Herzog, E Rota Kops, W Oberschelp, HW Müller-Gärtner, "Expectation maximization reconstruction of positron emission tomography images using anatomical magnetic resonance information", *IEEE Trans Med Imaging*, vol. 16, no. 2, pp. 129-136, 1997.

[2] C Comtat, PE Kinahan, JA Fessler et al., "Clinically feasible reconstruction of 3D whole-body PET/CT data using blurred anatomical labels", *Phys Med Biol* vol 47 pp 1-20, 2002

[3] PP Bruyant, HC Gifford, G Gindi, MA King. "Numerical observer study of MAP-OSEM regularization methods with anatomical priors for lesion detection in 67Ga images", *IEEE Trans Nucl Sci* vol 51, pp 193-197, 2004.

[4] JE Bowsher, H Yuan, LW Hedlund et al., "Using MRI information to estimate F18-FDG distributions in rat flank tumors", *Conference Record IEEE Nucl Sci Symp Med Imag Conf*, Rome, pp 2488-2492, 2004.

[5] K Baete, J Nuyts, K Van Laere et al., "Evaluation of anatomy based reconstruction for partial volume correction in brain FDG-PET", *NeuroImage*, vol 23, pp 305-317, 2004.

[6] A Rangarajan, IT Hsiao, G Gindi, "A Bayesian joint mixture framework for the integration of anatomical information in functional image reconstruction", *J Math Imaging Vision*, vol 12, pp 199-217, 2000.

[7] S Somayajula, E Asma, RM Leahy, "PET Image Reconstruction using Anatomical Information through Mutual Information Based Priors", *Conf. Record: IEEE Nucl Sci Symp Med Imag Conf* , Puerto Rico, pp 2722-2726, 2005.

[8] AM Alessio, PE Kinahan. "Improved quantitation for PET/CT image reconstruction with system modeling and anatomical priors", *Med Physics*, vol 33, no 11, pp 4095-4103, 2006.

[9] S Ahn, RM Leahy. "Analysis of Region of Interest Quantification for PET Image Reconstruction with Selective Regularization". *IEEE Nuclear Science Symposium Conference Record*, vol 3, pp 1781-1786, 2006

[10] S Kulkarni, P Khurd, I Hsiao, L Zhou, G Gindi. "A channelized Hotelling observer study of lesion detection in SPECT MAP reconstruction using anatomical priors", *Phys Med Biol* vol 52, pp 3601-3617, 2007

[11] F Maes, A Collignon, D Vandermeulen et al., "Multimodality image registration by maximization of mutual information," *IEEE Trans Med Imaging.*, vol 16, pp. 187-198, 1997.

[12] J Nuyts, D Bequé, P Dupont, L Mortelmans. "A concave prior penalizing relative differences for maximum-a-posteriori reconstruction in emission tomography", *IEEE Trans Nucl Sci*, vol 49, pp 56-60, 2002.