BAYESIAN PET IMAGE RECONSTRUCTION INCORPORATING ANATO-FUNCTIONAL JOINT ENTROPY

Jing Tang, Benjamin M. W. Tsui, and Arman Rahmim

Department of Radiology, The Johns Hopkins University, Baltimore, MD 21287, USA

ABSTRACT

We developed a maximum *a posterior* (MAP) reconstruction method for PET image reconstruction incorporating MR image information, with the joint entropy between the PET and MR image features serving as the prior. A non-parametric method was used to estimate the joint probability density (JPD) of the PET and MR images. The sampling rate for Parzen window estimation of the JPD was studied for both simulated phantom and clinical FDG PET brain images. Using realistic simulated PET and MR brain phantoms, the quantitative performance of the proposed algorithm was investigated. In particular, variations in the weighting factor on the MAP prior as well as the variance in the Parzen window were examined. Incorporation of the anatomical information via this technique was seen to noticeably improve the noise vs. bias tradeoff in various regions of interest.

Index Terms— positron emission tomography, Bayesian image reconstruction, anatomical priors, joint entropy

1. INTRODUCTION

PET imaging technique provides important physiological and biochemical information for clinical diagnosis and scientific studies. However, even state-of-art PET imaging continues to be affected by limited spatial resolution and inherently noisy data. Incorporation of anatomical information obtained from high-resolution MR/CT anatomical images may potentially improve the PET image quality and quantitative accuracy. The anatomical information is usually applied as the priors in the Bayesian PET image reconstruction framework. These techniques often consist of using segmented anatomic images, in which regional boundaries or labels are produced, to penalize inter-voxel noise within the boundaries [1, 2].

Nevertheless, while many types of PET imaging tasks exhibit strong correlations between anatomy and radiopharmaceutical uptake, the relationship can be complex and indirect (e.g. increased lesion uptake in otherwise uniform anatomical region, or areas of increased uptake in PET brain stimulation studies, which appear the same as any other portion of gray matter in MR images). As such, an area of growing interest is that of developing reconstruction algorithms designed to take into account differences between PET and anatomical segmentation [3, 4]. In particular, Somayajula *et al.* [5] proposed to use the mutual information between features (e.g. intensities or gradients) of MR/CT and PET images as priors in the reconstruction task to address the aforementioned issue. It was later suggested by Nuyts that the use of only the joint entropy component of the mutual information metric was better conditioned to minimize introduction of bias into the reconstructed images [6].

In this paper, we have designed and investigated a one-step-late (OSL) maximum *a posteriori* (MAP) algorithm with the joint entropy between features (such as intensity) of the anato-functional image pairs as the prior. We tested the algorithm using a simulated brain phantom. Non-parametric estimation of the joint probability density (JPD) was performed [7, 8]. Noise vs. bias tradeoff curves were generated for different regions of interest (ROIs) of the brain and the entire brain, and were used to evaluate the proposed reconstruction framework.

2. METHODS

2.1. MAP Reconstruction Algorithm

The MAP estimate of the functional image $\mathbf{f}$ from emission sinogram data $\mathbf{g}$ is given by

$$\hat{\mathbf{f}} = \arg\max_{\mathbf{f}\geq 0} \frac{P(\mathbf{g}|\mathbf{f})P(\mathbf{f})}{P(\mathbf{g})}. \quad (1)$$

The expression $P()$ indicates a probability of the variable enclosed in parenthesis.

Let the N feature vectors extracted from the functional and anatomical images be represented by $\mathbf{x}_i$ and $\mathbf{y}_i$, respectively for $i = 1, 2, \ldots, N$. Note if only one feature, e.g. the intensity of the two images are used, $\mathbf{x}_i$ and $\mathbf{y}_i$ represent the functional and anatomical images, respectively. The feature vectors, $\mathbf{x}_i$ and $\mathbf{y}_i$, can be considered as realizations of the random feature vectors $\mathbf{X}$ and $\mathbf{Y}$. The prior $P(\mathbf{f})$ is assumed to have a Gibbs distribution of the form

$$P(\mathbf{f}) = \frac{1}{W}\exp(-\beta H(\mathbf{X}, \mathbf{Y})), \quad (2)$$

This work was supported by the Siemens Medical Solutions USA, INC and the National Institutes of Health under grant R01-EB000168. The authors wish to thank Dr. Hiroto Kuwabara for consultation on realistic MRI simulations.

where W is a normalization factor, β is a positive constant, and $H(\mathbf{X}, \mathbf{Y})$ is the joint entropy of the two random feature vectors.

Seeking the solution $\mathbf{f}$ to (1) is equivalent to finding $\mathbf{f}$ that maximizes the log-posterior probability:

$$\log P(\mathbf{g}|\mathbf{f}) - \beta H(\mathbf{X}, \mathbf{Y}) + k, \tag{3}$$

where the constant k represents the contribution of the W term in (2) and the denominator in (1).

The joint entropy of two random feature vectors $H(\mathbf{X}, \mathbf{Y})$ can be evaluated by constructing the vector random variable $\mathbf{Z} = [\mathbf{X}, \mathbf{Y}]^T$ and evaluating $H(\mathbf{Z})$. The estimate of this entropy will be dependent on the covariance $\psi_{\mathbf{Z}}$ of the multi-dimensional Parzen window functions that are used in the density estimator for $\mathbf{Z}$. Assume that this covariance matrix is diagonal $\psi_{\mathbf{Z}} = \text{DIAG}(\psi_{\mathbf{XX}}, \psi_{\mathbf{YY}})$, $\psi_{\mathbf{XX}}$ and $\psi_{\mathbf{YY}}$ being the diagonal covariance matrices for the functional and anotomical image feature vectors, respectively. The entropy of the random variable $\mathbf{Z}$ may be expressed as

$$H(\mathbf{Z}) = -E_{\mathbf{Z}}[\ln p(\mathbf{Z})] = -\sum_{\mathbf{Z}} p(\mathbf{Z}) \ln p(\mathbf{Z}). \tag{4}$$

The probability density (PD) $p(\mathbf{Z})$ is approximated by a superposition of Gaussian densities centered on the elements $\mathbf{z}_j$ of a sample S drawn from $\mathbf{Z}$

$$\hat{P}(\mathbf{Z}) = \frac{1}{N_S} \sum_{j \in S} G_{\psi}(\mathbf{Z} - \mathbf{z}_j), \tag{5}$$

where N_S is the size of the sample S,

$$G_{\psi}(\mathbf{Z}) = (2\pi)^{\frac{-n}{2}} |\psi|^{-\frac{1}{2}} \exp(-\frac{1}{2} \mathbf{Z}^T \psi^{-1} \mathbf{Z}), \tag{6}$$

and n is the dimension of $\mathbf{Z}$. The use of the Gaussian density (6) in the Parzen density estimate (5) simplifies the subsequent analysis but is not necessary [8].

Applying the Poisson statistics to the first term of (3), we derive the following iterative procedure as what was done by Green [9]

$$f_i^{\text{new}} = \frac{f_i^{\text{old}}}{\sum_j c_{ij} + \beta \frac{\partial H(\mathbf{X},\mathbf{Y})}{\partial f_i}|_{f_i = f_i^{\text{old}}}} \sum_j \frac{c_{ij} g_{ij}}{\sum_i c_{ij} f_i^{\text{old}}}. \tag{7}$$

The OSL strategy is applied in calculating the derivative of the joint entropy.

If only the image intensity is used as the feature considered, the derivative of the joint entropy $H(\mathbf{Z})$ with respect to the functional image intensity f_i becomes $\frac{\partial H(\mathbf{Z})}{\partial x_i}$. The JPD can be represented by

$$\begin{aligned} \hat{P}(\mathbf{X}, \mathbf{Y}) =& \frac{1}{N_S} \sum_{j \in S} [(2\pi)^{-1} (\psi_{\mathbf{XX}}^j \psi_{\mathbf{YY}}^j)^{-\frac{1}{2}} \\ & \cdot \exp(-\frac{1}{2} \frac{(\mathbf{X} - x_j)^2}{\psi_{\mathbf{XX}}^j} - \frac{1}{2} \frac{(\mathbf{Y} - y_j)^2}{\psi_{\mathbf{YY}}^j})] \end{aligned} \tag{8}$$

and the derivative of the joint entropy is

$$\begin{aligned} & \frac{\partial H(\mathbf{X}, \mathbf{Y})}{\partial x_i} = -\sum_{\mathbf{X},\mathbf{Y}} \left(\ln \hat{P}(\mathbf{X}, \mathbf{Y}) + 1 \right) \frac{\partial \hat{P}(\mathbf{X}, \mathbf{Y})}{\partial x_i} \\ =& -\sum_{\mathbf{X},\mathbf{Y}} [-\ln N_s + \ln \sum_{j \in S} ((2\pi)^{-1} \cdot (\psi_{\mathbf{XX}}^j \psi_{\mathbf{YY}}^j)^{-\frac{1}{2}} \cdot \\ & \exp(-\frac{1}{2} \frac{(\mathbf{X} - x_j)^2}{\psi_{\mathbf{XX}}^j} - \frac{1}{2} \frac{(\mathbf{Y} - y_j)^2}{\psi_{\mathbf{YY}}^j})) + 1] \cdot \\ & \frac{1}{N_s} (2\pi)^{-1} \cdot (\psi_{\mathbf{XX}}^i \psi_{\mathbf{YY}}^i)^{-\frac{1}{2}} \cdot \\ & \exp(-\frac{1}{2} \frac{(\mathbf{X} - x_i)^2}{\psi_{\mathbf{XX}}^i} - \frac{1}{2} \frac{(\mathbf{Y} - y_i)^2}{\psi_{\mathbf{YY}}^i}) \cdot (\frac{\mathbf{X} - x_i}{\psi_{\mathbf{XX}}^i}) \end{aligned} \tag{9}$$

2.2. Phantom Simulation

We used a new mathematical brain phantom [10] for the purpose of performing realistic simulations. The brain phantom was constructed using subdivision surfaces [11], which can be used to efficiently model structures with an arbitrary topological type, such as the brain, skull, muscle tissue, and vasculature. Surfaces were modeled based on a segmented MRI dataset of a normal subject, with 100 structures identified and modeled.

To test the MAP algorithm described above, we performed analytical simulations of the brain phantom, where the activities in the brain were based on a clinical raclopride study. Data were simulated for the geometry of the Discovery RX PET/CT scanner [12], but with the transaxial dimensions (crystal dimensions, field of view) scaled by 0.5 to simulate a dedicated brain scanner. (We already have in place a dedicated Monte Carlo simulation package for the brain-only HRRT scanner [13], which will be applied later.)

2.3. Evaluation Metrics

To evaluate the reconstructed images quantitatively, we used the tradeoff between normalized mean squared error (NMSE) and normalized standard deviation (NSD) for individual areas of the brain (caudate, putamen, gray matter, white matter, cerebellum, and brain stem with exact boundaries) and the tradeoff between average NMSE and average NSD of the individual areas over the whole brain.

The NMSE_r for each ROI r (ranging from 1 to 6) covering one area of the brain was calcuated using

$$\text{NMSE}_r = \left(\frac{\bar{f}^r - \bar{\mu}^r}{\bar{\mu}^r} \right)^2, \tag{10}$$

where $\bar{f}^r = \frac{1}{n} \sum_{i=1}^n f_i^r$ and $\bar{\mu}^r = \frac{1}{n} \sum_{i=1}^n \mu_i^r$; f_i^r and μ_i^r denote the ith reconstructed and reference true activity value, respectively; and n is the number of voxels in the ROI r. We adopted such an ROI-based definition as we believe it should minimize the effect of voxel noise on this bias-measuring

metric. For each ROI, the NMSE value was plotted against the NSD value, as calculated using

$$\text{NSD}_r = \frac{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(f_i^r - \bar{f}^r)^2}}{\bar{f}^r}, \quad (11)$$

where f_j^r, $\bar{f}^r$, and n were defined as those in (10).

The overall bias was measured by

$$\text{NMSE} = \frac{1}{R}\sum_{r=1}^{R}\text{NMSE}_r, \quad (12)$$

where $R = 6$ including all the ROIs. The average NMSE value was subsequently plotted against the average NSD value over the regions, i.e. $\langle \text{NSD}_r \rangle$.

3. RESULTS

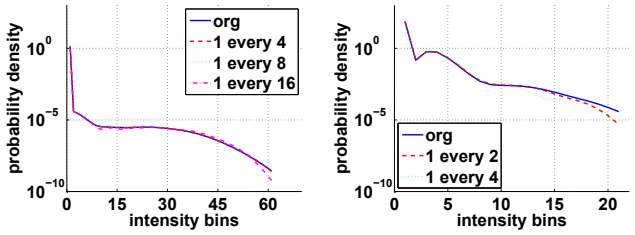
3.1. Probability Density Function Estimation

When non-parametric estimation of PD function is applied as in (5), the number of sampled voxels and the variance of individual Gaussian function are major variables to determine. Based on the statistical noise properties of emission tomography images [14], we use scaled squared individual PET voxel intensity as the variance for each Gaussian function and the scaling factor is named α. We tested a wide range of α and landed on a range of $\alpha = 0.002$ to 0.01 with resulting probabilities matching well with those calculated from image histogram. The effect of α on the PD estimation therefore on the image reconstruction will be presented in the next subsection.

(a) (b)

Fig. 1. Center slice of a clinical FDG brain image (a) and a simulated OS-EM reconstructed brain image (b).

In terms of the sample S in (5), the best we can do would be using every voxel in the image. As it is costly to exhaust the voxels, we studied the effect of sampling rates on the estimated PD aiming to save computing time. A clinical FDG brain image of size $256 \times 256 \times 128$ and a simulated first-iteration OS-EM reconstructed brain image of size $128 \times 128 \times 47$ shown in Fig. 1 were tested. The corresponding estimated PD curves with different sampling rates are presented in Fig. 2. As shown in Fig. 2 (a), the sampling rate of 1 voxel every 16 voxels in the clinical FDG image still reaches PD estimation close to that acquired using every voxel in the entire image. The results in Fig. 2 (b) shows for our simulation study, it is appropriate to use 1 every 4 voxels to estimate the PD.



(a) (b)

Fig. 2. Probability density calculated using the Parzen window method from (a) the clinical FDG image and (b) the simulated reconstructed image shown in Fig. 1, using all voxels in the images (org) and different sampling rates (1 every n voxels), α = 0.01.

3.2. Reconstructed Image Evaluation

The weighting factor β in (7) is a determining factor in the MAP reconstruction. Knowing the scale of system matrix components, we tested β ranging from 0.001 to 10. When β was smaller than $\sim$0.1, the effect of the prior information did not appear in the reconstructed images. When β was larger than $\sim$5, there were outrageous negative values in the reconstructed images, meaning the β weighted term was negative and larger than the $\sum_j c_{ij}$ in (7). The effect of β within the reasonable range was shown in Fig. 3. As shown in the plots for individual ROIs, applying the prior information led to better tradeoff of image noise and image bias. Among the β values tested, better improvement was achieved when β was increased from 0.2 to 1. Further increase of β to 2 started to reverse the improvement. The range from $\beta = 0.2$ to 0.5 seemed an optimal choice for improving NSD vs. NMSE tradeoff.

As mentioned in the last subsection, the effect of α for a fixed β was also studied. The average NSD vs. the average NSME for different α with $\beta = 0.5$ was plotted in Fig. 4. The improvement of NSD-NSME tradeoff was slightly better when decreasing α from 0.01 to 0.002. Decreasing α more, we did not get reasonable reconstructed image. It may infer that using too small a Parzen window variance, we can not estimate the image PD correctly.

To show how the image noise was improved visually using the MAP reconstruction algorithm, we plotted two images that had similar average NMSEs shown in Fig. 4. The 4th iteration from $\beta = 0$ and the 2nd iteration from $\beta = 0.5$ and $\alpha = 0.002$ reconstructed images are shown with the phantom image in Fig. 5. It is quite noticeable that the noise is significantly reduced in the MAP reconstructed image.

4. SUMMARY

We developed an OSL MAP reconstruction algorithm for PET image reconstruction using the joint entropy of anatomical and functional image features as the prior. The algorithm was

tested using simulated PET and MR brain images. The sampling rate of Parzen window estimation of the PD for PET images was studied to save computing time. The effects of the prior weighting factor and the Parzen window variance were studied quantitatively using the NSD-NMSE tradeoff as the measure of the reconstructed image noise and image bias. The MAP reconstructed image had better noise vs. bias tradeoff when the prior weighting factor was chosen appropriately. With the same average NMSE, the noise level reduction in MAP reconstructed images was visually noticeable compared to the OS-EM reconstructed images.

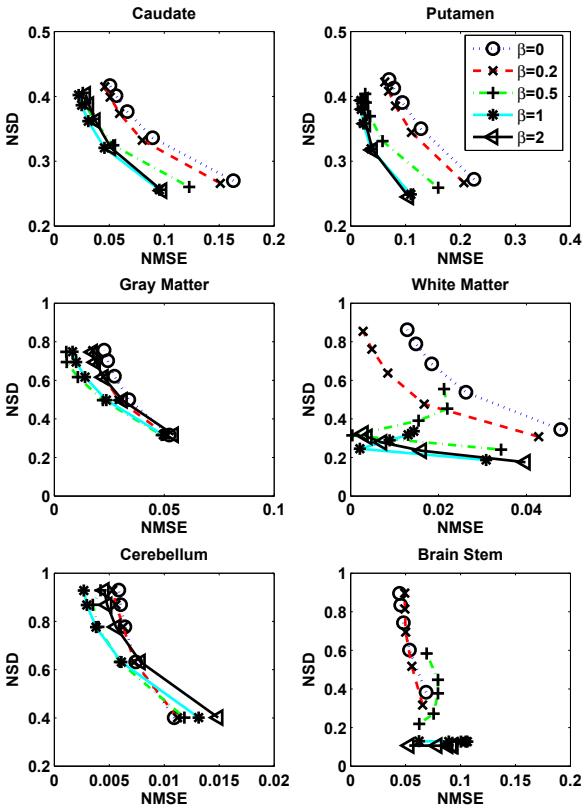


Fig. 3. The NSD vs. NMSE for different areas of the brain image changing with iterations, $\alpha = 0.002$ for the curves other than $\beta = 0$ (OS-EM reconstruction).

5. REFERENCES

[1] 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", *Phys. Med. Biol.*, vol. 47, pp. 1-20, 2002.

[2] K. Baete, J. Nuyts, W. Van Paesschen, P. Suetens, and P. Dupont, "Anatomical-based FDG-PET reconstruction for the detection of hypometabolic regions in epilepsy", *IEEE Trans. Med. Imaging*, vol. 23, pp. 510-519, 2004.

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

[4] J. E. Bowsher, V. E. Johnson, T. G. Turkington, R. J. Jaszczak, C. E. Floyd Jr., and R. E. Coleman, "Bayesian reconstruction and use of anatomical a priori information for emission tomography", *IEEE Trans. Med. Imag.*,vol. 15, pp. 673-686, 1996.

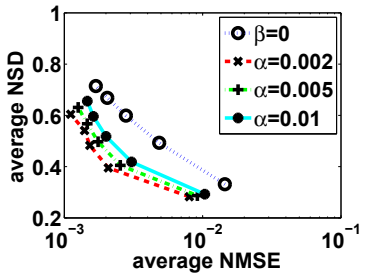


Fig. 4. The average NSD vs. average NMSE for the reconstructed brain image changing with iterations, $\beta = 0.5$ for the curves other than $\beta = 0$ (OS-EM reconstruction).

(a) (b) (c)

Fig. 5. Center slice of the phantom image (a) and reconstructed images from the 4th iteration of $\beta = 0$ (b) and the 2nd iteration of $\beta = 0.5$ and $\alpha = 0.002$ (c).

[5] S. Somayajula, E. Asma, and R. M. Leahy, "PET image reconstruction using anatomical information through mutual information based priors", *IEEE Nucl. Sci. Symp. Conference Record* , pp. 2722-26, 2005.

[6] J. Nuyts, "The use of mutual information and joint entropy for anatomical priors in emission tomography", *IEEE Nucl. Sci. Symp. Med. Imaging Conference Record*, Honolulu, 2007.

[7] R. O. Duda, P.E. Hart and D.G. Stork, Pattern Classification, Second Edition, Wiley, 2001.

[8] P. A. Viola, Alignment by maximation of mutual information. PhD thesis, Massachusetts Institute of Technology. 1995

[9] P. J. Green, "Bayesian reconstructions from emission tomography data using a modified EM algorithm", *IEEE Trans. Med. Imaging*, vol. 9, pp. 84-93, 1990.

[10] A. Rahmim, K. Dinelle, J. C. Cheng, M. Shilov, W. P. Segars, S. C. Lidstone, S. Blinder, O. G. Rousset, H. Vajihoolahi, B. M. W. Tsui, D. F. Wong, and V. Sossi, "Accurate event-driven motion compensation in high-resolution PET incorporating scattered and random events", *IEEE Trans. Med. Imag.*, In Print, 2008.

[11] H. Hoppe, T. DeRose, T. Duchamp, M. Halstead, H. Jin, J. McDonald, J. Schweitzer, W. Stuetzle, "Piecewise smooth surface reconstruction", Computer Graphics, vol. 28, pp. 295-302, 1994.

[12] B. J. Kemp, C. Kim, J. J. Williams, A. Ganin, and V. J. Lowe, "NEMA NU 2-2001 performance measurements of an LYSO-based PET/CT system in 2D and 3D acquisition modes", *J. Nucl. Med.*, vol. 47, pp. 1960-1967, 2006.

[13] V. Sossi, H. W. A. M. de Jong, W. C. Barker, P. Bloomfield, Z. Burbar, M.-L. Camborde, C. Comtat, L. A. Eriksson, S. Houle, D. Keator, C. Knob, R. Krais, A. A. Lammertsma, A. Rahmim, M. Sibomana, M. Teras, C. J. Thompson, R. Trebossen, J. Votaw, M. Walker, K. Wienhard, D. F. Wong, "The second generation HRRT-a multi-center scanner performance investigation", *IEEE NSS/MIC Conference Record*, vol. 4, pp. 23-29, 2005.

[14] H. H. Barrett, D. W. Wilson, and B. M. W. Tsui, "Noise properties of the EM algorithm: I. Theory", *Phys. Med. Biol.* vol. 39, pp. 833-846, 1994.