Performance of the Relative Difference Prior for Hot Lesion Detection in Whole-body PET/CT: an Evaluation with Numerical and Real Observers

Johan Nuyts *Member, IEEE,*, Christian Michel *Member, IEEE,*

***Abstract*— For PET (positron emission tomography) imaging, different reconstruction methods can be applied, including ML (maximum likelihood) and MAP (maximum a-posteriori) reconstruction. Post-smoothed ML images have approximately position and object independent spatial resolution, which is advantageous for (semi) quantitative analysis. However, the complex object dependent smoothing obtained with MAP might yield improved noise characteristics, beneficial for lesion detection. In this contribution, MAP and post-smoothed ML are compared for hot spot detection by the channelized Hotelling observer (CHO) and by human observers (using a "multiple alternative forced choice" method). For MAP, a prior penalizing relative rather than absolute differences was applied.**

MAP resulted in markedly better performance for the CHO. However, no significant performance difference was detected with the human observer study. Addition of internal noise improved the agreement between the CHO and the human observers.

The prior can be adjusted to obtain a smaller penalty for large differences, resulting in improved tolerance for strong edges. An additional CHO study indicated that this edge tolerance decreases observer performance. Similarly, performance was better for the quadratic prior than for the Huber prior.

I. Introduction

IT is generally accepted that statistical reconstruction of PET images using maximum-likelihood (ML) or maximum-a-posteriori (MAP) algorithms results in improved image quality when compared to filtered backprojection. When iterated to convergence, ML yields very noisy images. However, when these images are filtered with a well-chosen low pass filter, they have very attractive features: they are quantitatively accurate, they have nearly position and object independent spatial resolution and high visual quality. In contrast, if the prior is well tuned, MAP can be iterated to convergence, yielding excellent images that need no further processing. If a shift-invariant prior is used, however, the resolution is object and position dependent, which is usually considered a disadvantage for clinical use. When a prior is tuned to impose uniform resolution, the images produced by MAP and post-processed ML are equivalent [1], [2], [3].

The non-isotropic and object dependent resolution of MAP is a direct result of the non-uniform variance in the projections. Shift invariant priors smooth more when the data are less reliable. This may be beneficial for non-ideal observers who are not able to optimally process correlated noise: they may need more smoothing when the data are noisier.

To avoid the aforementioned problems with spatial resolution, we currently use post-smoothed ML for our clinical images. This study investigates if the use of MAP would be recommended for clinical PET applications where hot spots must be detected in a noisy background, as is typically the case in oncological PET examinations.

The next section describes the relative difference prior and the tools used to evaluate it: the signal-to-noise ratio (SNR) of the channelized hotelling (CHO) observer and the MAFC (multiple alternative forced choice) human observer study. It also explains how the CHO was extended with internal noise, and how it was applied to the MAFC data set. The third section then describes the application of these tools in three experiments: A) computation of CHO SNR to compare MAP and post-smoothed ML, B) MAFC study comparing the same algorithms, and C) a CHO SNR study to examine the effect of using priors with edge tolerance. The results are presented in section IV and discussed in section V.

II. Methods

A. Relative difference prior

The ML reconstruction is obtained by the iterative maximization of (the logarithm of) the likelihood, assuming that the PET data are a realization of a Poisson distribution. In MAP, the logarithm of a prior distribution is added to the log-likelihood term. The prior is mostly used to suppress noise by favoring smooth reconstructions. The objective function then becomes:

$$\text{posterior}(\Lambda, Y) = L(\Lambda, Y) + M(\Lambda), \tag{1}$$

where L is the likelihood, M is the prior, Y are the PET data and Λ is the image to be reconstructed by maximizing the posterior. With convex priors (such as the ones considered in this paper), the posterior has a unique maximum.

Most MAP implementations use a Gibbs prior, which is a function of the absolute difference of the pixel values of neighboring pixels. In [4], a prior was proposed that penalizes relative differences instead of absolute differences. The prior can be written as

$$M(\Lambda) = -\beta \sum_j \sum_{k \in N_j} \frac{(\lambda_j - \lambda_k)^2}{(\lambda_j + \lambda_k) + \gamma|\lambda_j - \lambda_k|}, \tag{2}$$

This work was supported by F.W.O. grant G.0174.03 and by Siemens

J. Nuyts is with Nuclear Medicin, K.U.Leuven, Belgiume (e-mail: Johan.Nuyts@uz.kuleuven.ac.be), C. Michel is with CPS Innovations, Siemens, Knoxville, TN, USA, (e-mail: Christian.Michel@cpspet.com).

where λ_j and λ_k are values of the pixels j and k of image Λ, and β is the strength of the prior. The parameter γ determines the edge tolerance. With $\gamma = 0$, the prior is very similar to the quadratic prior, except that its strength is not constant but inversely proportional to the local pixel intensity [4].

Mumcuoğlu et al [5] proposed to use the Huber prior instead of the quadratic prior. This prior is identical to the quadratic prior for small differences, but becomes linear rather than quadratic for large differences. The decreased penalty for large differences results in better tolerance for strong edges. A similar effect is obtained by setting γ to a non-zero value in (2): the prior becomes linear rather than quadratic when the relative difference $|\lambda_j - \lambda_k|/(\lambda_j + \lambda_k)$ is large compared to $1/\gamma$.

In this study, we compared the performance of MAP with the relative difference prior to post-smoothed ML. In addition, we evaluated the influence of edge tolerance on lesion detection performance.

B. Channelized hotelling observer

Currently, the CHO is often used as a model for the human observer in lesion detection tasks. As a measure of detection performance, we computed the CHO signal-to-noise ratio for an SKE-BKE experiment (signal known exactly, background known exactly), for different lesion positions in the image.

In our study, the CHO was similar to that of [6], [7], it used three adjacent rectangular bandpass filters, defined by the four frequencies $[q^{-3}B, q^{-2}B, q^{-1}B, qB]$, with B = 0.4 per pixel and q = 2.3. The corresponding impulse responses in the spatial domain were computed and centered at the lesion position. To detect the presence or absence of a lesion at a particular position, the CHO applies the three filters to that location and obtains a column matrix A with three values, one for each filter. The CHO is assumed to know exactly the mean and covariance matrix of the filter outputs, and estimates the probability for the observation A as

$$p(A|\bar{A}_i) = N_i \exp\left(-\frac{1}{2}(A-\bar{A}_i)' C_{A_i}^{-1}(A-\bar{A}_i)\right) \quad (3)$$

where $\bar{A}_i$ is the mean under the hypothesis $i = 0, 1$ (lesion present or absent), $C_{A_i}^{-1}$ is the covariance matrix, prime denotes matrix transpose and N_i is a normalization constant. The CHO computes the logarithm of the likelihood and select the most likely hypothesis. Assuming that $C_{A_0} = C_{A_1} = C_A$, one obtains:

$$q(A) = (\bar{A}_1 - \bar{A}_0)' C_A^{-1} A - \frac{1}{2}(\bar{A}_1' C_A^{-1} \bar{A}_1 - \bar{A}_0' C_A^{-1} \bar{A}_0). \quad (4)$$

The signal is $q(\bar{A}_1) - q(\bar{A}_0)$, the variance of q is readily derived from (4) if one assumes that the noise is Gaussian. This yields a signal-to-noise ratio which can be used as a measure of image quality [6]:

$$\text{SNR} = \sqrt{(\bar{A}_1 - \bar{A}_0)' C_A^{-1} (\bar{A}_1 - \bar{A}_0)}. \quad (5)$$

C. Multiple alternative forced choice study

For the human observer study, we used the MAFC (multiple alternative forced choice) approach [8]. Images with exactly one lesion are presented, the observer has to select the lesion position from a finite set of possible positions. This approach does not yield an ROC or LROC curve. However, it is far less demanding for the observers, mainly because they don't have to rate their certainty. Similar to the LROC-experiment, the MAFC experiment can be used to compute a detectability index or equivalent number, which can be used as a quantitative measure of image quality [8].

To compare human and CHO performance, the CHO was applied to the MAFC data as well. For that purpose, (4) was computed for every lesion position, and the position with maximum q was selected. This requires a covariance matrix C_A for every lesion position, which was estimated by averaging over all available noisy images, and using the noise-free images as the mean. This approach assumes that the covariance matrix is independent of the contrast and is the same for lesion present and absent (or at least that the numerical observer believes so).

For CHO with internal noise (see next paragraph), the CHO was applied a 100 times for every image with different internal noise realizations.

D. CHO with internal noise

The direct comparison between the CHO and human observers on the same detection task revealed superior performance for the CHO. This was particularly true for MAP reconstruction: CHO performance increased almost monotonically with increasing strength β of the smoothing prior, while human performance reached a maximum for an intermediate value of β (see eq (2)). Heavy smoothing with the prior decreases both the signal and the noise. Apparently, the noise decreases faster, yielding improved SNR and lesion detection by the CHO. However, it seems that humans, unlike the CHO, are unable to detect the small intensity changes. This can be modeled by adding internal noise to the CHO [9], [7], [10]. We have added noise with a standard deviation proportional to the average amplitude of the channel outputs when there is no lesion:

$$C_A^*[i, j \neq i] = C_A[i, j] \quad (6)$$

$$C_A^*[i, i] = C_A[i, i] + \alpha\left(C_A[i, i] + \left(\bar{A}_0[i]\right)^2\right) \quad (7)$$

where C_A^* is the adjusted covariance, and α is a parameter, tuned with trial and error to obtain a good agreement with human performance. Thus, an amplitude change is only reliably detected if it is sufficiently large compared to the average background amplitude.

III. Experiments

In the first experiment, CHO tumor detection performance in a simulated PET slice of the thorax was studied to compare post-smoothed ML and MAP for $\gamma = 0$ in (2). The second experiment used the MAFC approach to compare the same algorithms in a human observer experiment. The CHO was applied to the same MAFC-image set to tune the internal noise

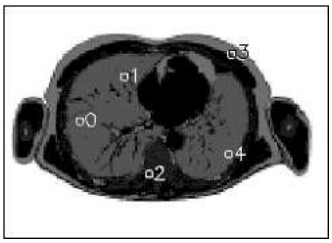

Fig. 1. Thorax phantom used for the channelized Hotelling observer study

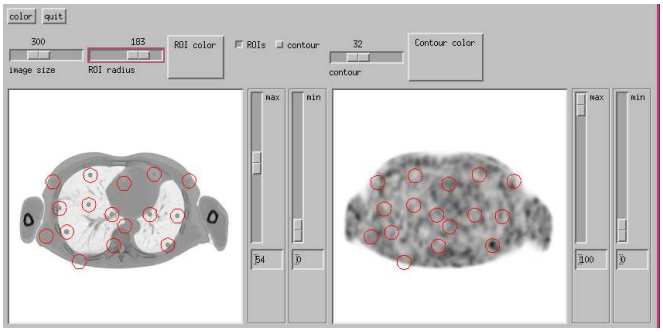

Fig. 2. The interactive tool for the human observer study. The 16 lesion positions are indicated.

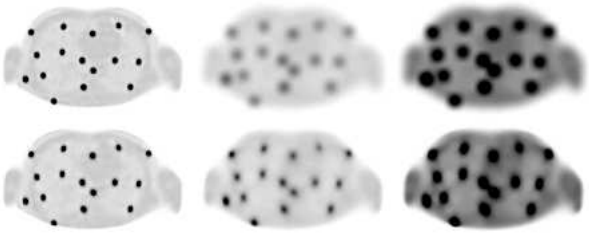

Fig. 3. Noise-free images of the thorax phantom used for the human observer study. Top row is ML, bottom row is MAP. The left images have the weakest regularization, the central ones the strongest. The left and middle images are scaled to the same maximum. The two right images are the same as the central ones, but scaled to a lower maximum, illustrating the difference in local impulse response.

parameter α. In the third experiment, the influence of MAP with increased edge tolerance on tumor detection was studied using the CHO. To estimate the error on the CHO SNR, the bootstrapping method of Efron was used [11].

The PET slice was derived from a CT image (pixel size 0.825 mm). The image was segmented with thresholding, and PET activities and attenuation values were assigned to the different tissue types. Attenuated projections were computed for 144 angles. These projections were smoothed to simulate a detector resolution of 5 mm FWHM and rebinned to obtain a PET sinogram of 150 detectors and 144 angles with a pixel size of 3.3 mm, with a total count of 245000. For ML and MAP reconstruction, the equivalent of 103 iterations were used, using a scheme with a gradually decreasing number of subsets.

A. CHO study comparing ML and MAP without edge tolerance.

Fig 1 shows the thorax image that was used for the studies. For the channelized Hotelling observer (CHO) study, a lesion was inserted at one of 5 positions, shown also in fig 1. For each lesion, 100 Poisson noise realizations of attenuated PET projections were computed and reconstructed with ML and with MAP (using $\gamma = 0$). The regularization was varied from "not smooth enough" to "too smooth" for clinical practice, as judged by simple visual inspection. For ML, 5 different widths of the Gaussian post-filter were applied (FWHM = 2,3,5,8,16 pixels, pixel size 3.3 mm), for MAP, 5 different values of β were used ($\beta = 1, 3, 10, 30, 100$).

B. MAFC human observer study

For the human observer study a multiple alternative forced choice experiment was designed [8]. The same thorax image was used, but now there were 16 lesion positions. The lesions had a diameter of 1 cm and the tumor to background ratio was varied from 5 to 2.5 in 10 steps. In each image presented to the observer, 15 lesions were "benign", having the same activity as the surrounding tissue, and 1 lesion was malign, having increased activity. Together with the simulated PET image, the registered simulated CT image was shown to the observer. Figure 2 shows a snapshot of the interactive tool. The observer could select the color lookup table, adjust its minimum and maximum, change the image size on the screen etc., similar as with the viewers used for clinical purposes. Optionally the possible lesion positions were indicated with a circle of adjustable size, as shown in the figure. Scoring was done by clicking in one of the lesion positions. The PET image was shown alternatively at the left and the right, in an attempt to prevent the observer from rapidly browsing through the images, using changes between subsequent images to detect tumors. For each contrast, 8 randomly selected malign lesions were shown, and the images were presented in order of decreasing contrast. Thus, for each session, the observer had to localize 80 lesions. There were 12 sessions: 6 for post-smoothed ML (FWHM of 2, 3, 4, 5, 6 and 8 pixels) and 6 for MAP (β = 3, 5, 7, 10, 20, 40).

The extreme regularizations are shown for noise-free images with all lesions present in fig 3. The smoothing effects of the prior and the Gaussian are very different. The prior has a local impulse response with a sharp peak and heavy tails [2], while the Gaussian has a wider peak but is more compact. No attempt was undertaken to match the regularizations.

The CHO was applied to score the same data, and the parameter α was adjusted to obtain good agreement with human performance.

The output of the MAFC experiment is the fraction of correct responses, which we use as a measure of detection performance. The detectability index is a monotonic function of the fraction of correct responses [8], so both numbers are equivalent for ranking algorithms.

C. CHO-study for MAP with edge tolerance

The first experiment CHO SNR study (which was done with $\gamma = 0$) was repeated for $\gamma = 10$ and 20. In addition, the experiment was repeated for the quadratic prior and the Huber

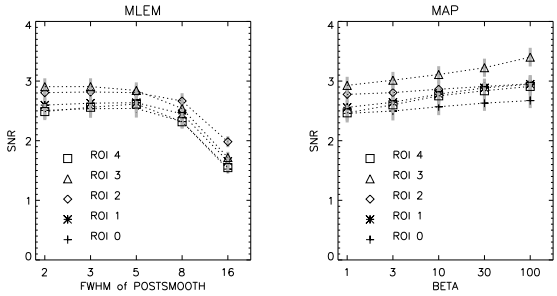

Fig. 4. The SNR of the **CHO without internal noise** as a function of regularization, for MAP and postsmoothed ML. There is one curve per lesion, see 1 for the lesion positions. The vertical bars represent the standard deviations obtained with the bootstrapping method.

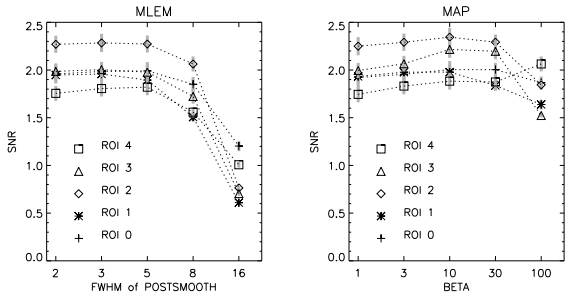

Fig. 5. The SNR of the **CHO with internal noise** as a function of regularization, for MAP and postsmoothed ML. The vertical bars represent the standard deviations obtained with the bootstrapping method.

prior. As argued above, the relative difference prior with $\gamma = 0$ is similar to a quadratic prior. With $\gamma > 0$, it becomes more similar to the Huber prior.

IV. Results

A. CHO study comparing ML and MAP without edge tolerance.

Figure 4 shows the SNR of the CHO experiment without internal noise for each of the five regions. For MAP, the CHO performance increases monotonically with increasing β. In contrast, there is a maximum performance for intermediate smoothing with ML. When the CHO was handicapped with internal noise, CHO performance was best at intermediate regularization for both algorithms (figure 5), which agrees better with human performance.

B. MAFC human observer study

Nine observers have completed the study, the fraction of correct responses is plotted in fig 6. The CHO was applied to the same data, its response is shown in the same figure. The parameter α in (7) was set to 0.6 to obtain good agreement with the human performance. The CHO with the same parameters was used to produce the curves of fig 5 and 7.

C. CHO-study for MAP with edge tolerance

Figure 7 shows the response of the CHO with internal noise for different MAP algorithms, using the thorax image

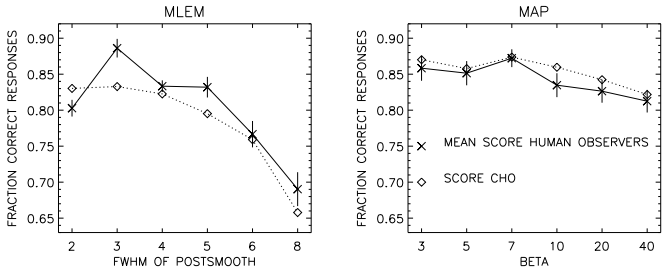

Fig. 6. Fraction of correct responses in the MAFC study averaged over the 9 observers, left ML and right MAP (solid line). The vertical lines indicate $\pm$ 1 standard deviation. The CHO response is plotted as a dashed line.

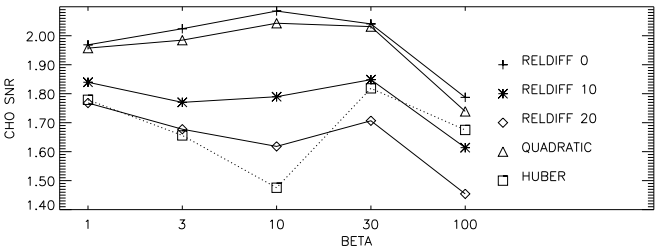

Fig. 7. The response of CHO with internal noise to MAP with different priors, as a function of the prior strength β. For each prior, the CHO response was averaged over the 5 regions positions shown in fig 1

of figure 1. These include the relative difference prior with $\gamma = 0, 10, 20$, the quadratic prior and the Huber prior. The curves of fig. 7 were obtained by averaging over the 5 lesion positions. As expected, the performance for the quadratic prior is very similar to that for the relative difference prior. Increasing edge tolerance by increasing γ or using the Huber prior results in decreased performance. Note that for high values of β the hot spot is heavily smoothed. Consequently, the differences between neighboring pixels becomes smaller, entering the range where the priors act in their quadratic mode (small differences are not considered as an edge). As a result, the performance returns to that of the priors without edge tolerance. We believe this explains why the edge preserving priors have a local maximum near $\beta = 30$

V. Discussion

The five lesions in the CHO study were positioned such as to test different factors expected to affect the SNR, including the level and degree of asymmetry in attenuation and measured count. For the human observer study, more regions were used in an attempt to obtain more "global" performance, and to make the task somewhat more realistic. The lesions were presented with random position, but in order of decreasing contrast. The first images are relatively easy to score, and they allow the observer to get used to the image characteristics. This "learning phase" is useful, because the observers were not used to the MAP images.

For practical reasons the entire study used only one image. As a result, the observers rapidly knew the anatomy and tracer distribution much better than they would in a realistic PET

oncology case. However, we believe that this situation is somewhat similar to the analysis of well registered PET and CT images, e.g. as produced by hybrid PET/CT systems. In this situation, the CT image provides virtually exact anatomical knowledge, which allows the experienced observer to make a good estimate of the normal PET image. For that reason, a CT image was included in the observer study as well.

Without internal noise, the CHO performed better than the human observers, and did systematically better with MAP than with post-smoothed ML. As reported by other groups [9], [7], [10], a good agreement with human performance can be obtained by adding internal noise to the CHO. We hypothesized that humans are unable to detect the small differences in the heavily smoothed images. For that reason, we set the internal noise proportional to the average signal amplitude (in the lesion absent case) of each channel. This seemed to work better than simply multiplying the diagonal of the covariance matrix with a constant larger than one, but we have not systematically compared different models. Our noise model may be equivalent to the conversion of the image from floating point to 8 bit integer proposed in [12], which produces internal noise by round-off errors. When extended with internal noise, the CHO performance decreased and was optimal for intermediate regularization. Performance with MAP was still slightly superior, but the difference with ML was much smaller than without internal noise. Similar findings were reported by Oldan et al [7] for the analysis of SPECT MAP images.

The CHO study suggests MAP to be superior, but our human observer study did not reveal a significant difference between post-smoothed ML and MAP, provided that the regularization is tuned for best performance. Currently, post-smoothed ML is the method we apply in clinical routine, because it has the advantage of yielding more uniform resolution than MAP. Our study does not provide arguments to replace post-smoothed ML with MAP for these applications. Note that for other applications, in particular for the inclusion of anatomy, MAP was found to be superior to post-processed ML [13], [14].

As argued above, the relative difference prior is similar to the quadratic prior, except that its strength is adjusted based on the image intensity. This adjustment avoids oversmoothing in regions with high intensity, and makes the prior easier to tune [4]. However, when considering a single lesion position and varying the strength β, a similar performance curve should be obtained for both priors. This was confirmed by the CHO analysis (fig 7). It has been suggested that the quadratic penalty results in excessive smoothing near edges, and for that purpose other priors have been proposed, including the Huber prior. However, our CHO-study indicates that for hot lesion detection, the quadratic penalty is superior to the more edge preserving penalties.

A limitation of our study was the use of a single 2D image. It cannot be excluded that for slices with different anatomy and for a volumetric analysis [15], [16] different results could be obtained. We have attempted to use realistic count levels to generate the Poisson noise, but the doses and scan times in clinical practice vary considerably between PET centers. It is possible that scanning at very different count levels may lead to other findings.

VI. Conclusion

The human observer study did not reveal improved lesion detection with MAP relative to post-smoothed ML, although the CHO study suggested MAP to be superior.

The performance of the channelized Hotelling observer agreed well with that of human observers, when internal noise was added. With this tuned CHO, we found that lesion detection performance was better for a quadratic penalty than for penalties with edge tolerance.

VII. Acknowledgement

The authors wish to thank the observers K Baete, L Brepoels, S Ceyssens, L De Ceuninck, G Moulin-Romsee, L Plessers, S Stroobants, K Van Laere, R Verscuren.

References

[1] JW Stayman, JA Fessler, "Compensation for nonuniform resolution using penalized-likelihood reconstruction in space-variant imaging systems," *IEEE Trans Med Imaging*, vol 23, pp 269-84, 2004.

[2] J Nuyts, JA Fessler, "A penalized-likelihood image reconstruction method for emission tomography, compared to post-smoothed maximum-likelihood with matched spatial resolution." *IEEE Trans Med Imaging*, vol 22, pp 1042-52, 2003.

[3] J A Fessler. "Analytical approach to regularization design for isotropic spatial resolution." *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, vol. 3, pp. 2022-6, 2003.

[4] 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.

[5] EU Mumcuoğlu, RM Leahy, SR Cherry. "Bayesian reconstruction of PET images: methodology and performance analysis." *Phys Med Biol*, vol 41, pp 1777-1807, 1996.

[6] HC Gifford, MA King, DJ de Vries, EJ Soares. "Channelized Hotelling and human observer correlation for lesion detection in hepatic SPECT imaging.". *J Nucl Med* vol 41, pp. 514-521, 2000.

[7] J Oldan, S Kularni et al. "Channelized Hotelling and human observer study of optimal smoothing in SPECT MAP reconstruction", *IEEE Trans Nucl Sci* vol 51, pp.733 - 741 , 2004.

[8] AE Burgess. "Comparison of receiver operating characteristics and forced choice observer performance measurement methods", *Med Phys* 1995; 22: 643-655.

[9] KL Gilland, Y Qi et al. "Comparison of channelized Hotelling and human observers in determining the optimum OSEM reconstruction parameters for myocardial SPECT", *IEEE-MIC*, 2003, M11-287.

[10] HC Gifford, MA King, PH Pretorius, RG Wells. "A comparison of human and model observers in multislice LROC studies", *IEEE Trans Med Imaging* vol 24, pp 160-169, 2005.

[11] B. Efron. "Bootstrap methods: another look at the jackknife." *The Annals of Statistics* vol 7, pp 1-26, 1979.

[12] PP Bruyant, HC Gifford, G Gindi, PH Pretorius, MA King. "Human and numerical observer studies of lesion detection in Ga-67 images obtained with MAP-EM reconstructions and anatomical priors." proceedings of *IEEE NSS-MIC, Rome 2004*, M10-320.

[13] K Baete, J Nuyts, K Van Laere, W Van Paesschen, S Ceyssens, L De Ceuninck, O Gheysens, A Kelles, J Van den Eynden, P Suetens, P Dupont. "Evaluation of anatomy based reconstruction for partial volume correction in brain FDG-PET." *NeuroImage*, vol 23, pp. 305-317, 2004.

[14] J Nuyts, K Baete, D Bequé, P Dupont. "Comparison between MAP and post-processed ML for image reconstruction in emission tomography when anatomical knowledge is available." *IEEE Trans Med Imaging*, vol 24, pp 667-675, 2005.

[15] C Lartizien, P Kinahan, C Comtat. "Volumetric model and human observer comparisons of tumor detection for whole-body positron emission tomography", *Acad Radiol* 2004, 11: 637-548.

[16] JS Kim, PE Kinahan, C Lartizien, C Comtat, TK Lewellen. "A comparison of planar versus volumetric numerical observers for detection task performance in whole-body PET imaging". *IEEE Trans Nucl Sci*, vol 51, pp 34-40, 2004.