Towards continualized task-based resolution modeling in PET imaging

Saeed Ashrafinia[a], Nicolas. Karakatsanis[a], Hassan Mohy-ud-Din[a,b], Arman Rahmim[a,b,*]
[a]Department of Radiology, Johns Hopkins University, 601 N. Caroline St, Baltimore, MD USA 21287; [b]Dept. of Electrical & Computer Engineering, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD USA 21218
[*]arahmim1@jhmi.edu; Phone (410) 502-6173; Fax (410) 955-0696; www.jhu.edu/rahmim

ABSTRACT

We propose a generalized resolution modeling (RM) framework, including extensive task-based optimization, wherein we continualize the conventionally discrete framework of RM vs. no RM, to include varying degrees of RM. The proposed framework has the advantage of providing a trade-off between the enhanced contrast recovery by RM and the reduced inter-voxel correlations in the absence of RM, and to enable improved task performance. The investigated context was that of oncologic lung FDG PET imaging. Given a realistic blurring kernel of FWHM h ('true PSF'), we performed iterative EM including RM using a wide range of 'modeled PSF' kernels with varying widths $\hat{h}$. In our simulations, h = 6mm, while $\hat{h}$ varied from 0 (no RM) to 12mm, thus considering both underestimation and overestimation of the true PSF. Detection task performance was performed using prewhitened (PWMF) and non-prewhitened matched filter (NPWMF) observers. It was demonstrated that an underestimated resolution blur ($\hat{h}$ = 4mm) enhanced task performance, while slight over-estimation ($\hat{h}$ = 7mm) also achieved enhanced performance. The latter is ironically attributed to the presence of ringing artifacts. Nonetheless, in the case of the NPWMF, the increasing inter-voxel correlations with increasing values of $\hat{h}$ degrade detection task performance, and underestimation of the true PSF provides the optimal task performance. The proposed framework also achieves significant improvement of reproducibility, which is critical in quantitative imaging tasks such as treatment response monitoring.

Keywords: PET, image reconstruction, resolution modeling, task-based image assessment, detection, quantification.

1. INTRODUCTION

Resolution modeling (RM), also known as the point-spread function (PSF) modeling, has attracted considerable interest in PET, especially in the past decade (see recent review by Rahmim et al. [1]). Unlike post-reconstruction partial volume correction (PVC) methods [2-4], RM models resolution-degrading phenomena within the reconstruction step. It is an attractive alternative to a range of PVC methods that make simplifying assumptions. RM improves resolution and therefore achieves enhanced contrast recovery. At the same time, the following three issues pose concerns:

(1) RM considerably modifies the noise texture. It reduces voxel variances, while increasing inter-voxel correlations [5-8]. These two effects impact noise differently based on the very definition of noise [6]. For instance, image roughness ($\sigma_{spatial}$) will decrease, but the ensemble standard-deviation of ROI mean uptake(σ_{ens}) used to assess reproducibility, can remain unchanged [6] or even be significantly amplified especially for small ROIs [9, 10]. RM images, as such, can be deceptive, because they are likely to be assessed visually of having higher quality [11], since contrast is enhanced and spatial noise $\sigma_{spatial}$is reduced. However, the potential degradation in reproducibility (also observed when utilizing PVC [10, 12]) adversely impacts quantitative imaging tasks as in pharmacokinetic imaging or treatment response monitoring.

(2) The noise texture is more thoroughly captured by the noise power spectrum (NPS), which in fact determines detection task performance [1]: with RM, the NPS is amplified at frequencies in which the modulation transfer function (MTF) is improved, and turns out to compete against it, limiting detection performance in RM [8].

(3) RM can lead to notable edge artifacts, reminiscent of the Gibbs ringing overshoot at the edges, as extensively reviewed recently [1]. An approach to address this issue was to use a reconstruction kernel that blurred the true PSF [14, 15], i.e. underestimating its extent, and was shown to be effective at suppressing edge artifacts (e.g. see [9, 16]). Watson [17] suggested that direct use of measured PSF values within reconstruction actually over-estimated the PSF because reconstructions include additional 'numerical' mechanisms of point spread, primarily due to image discretization and the projection operations. Nonetheless, it can be shown [13, 16] (as we also show in this work) that even when the 'true' PSF actually has the exact same structure as the modeled PSF, edge artifacts can occur.

Proposed approach: the abovementioned three issues and observations have prompted us to propose extensive task-based assessment of a generalized RM framework, wherein we continualize the conventionally discrete framework of

Medical Imaging 2014: Physics of Medical Imaging, edited by Bruce R. Whiting, Christoph Hoeschen, Despina Kontos,
Proc. of SPIE Vol. 9033, 903327 · © 2014 SPIE · CCC code: 1605-7422/14/$18 · doi: 10.1117/12.2043973

RM vs. no RM, to include varying degrees of RM, including overestimation and underestimation of the 'true PSF'. Such a generalized scheme allows consideration of a much wider array of images, which are subsequently analyzed in the context of different imaging tasks. We elaborate upon these next.

2. THE GENERALIZED FRAMEWORK AND TASK-BASED ASSESSMENT

2.1 Continualized RM imaging

The investigated context was that of oncologic FDG PET imaging. In our initial assessment, we focused on lung tumor imaging, with SUV images at 45min, simulated based on kinetic parameters extracted from the literature ([18], Table II). Blurring was simulated in the image-space domain (being generalized in subsequent work to incorporate detection-space blurring), using a Gaussian blurring kernel with FWHM h ('true PSF'). We then performed iterative EM including RM using a wide range of 'modeled PSF' kernels with FWHM denoted by $\hat{h}$. In our simulations, h=6mm, while $\hat{h}$ varies from 0 (no RM) to 12mm, thus performing both underestimation and overestimation of the true PSF. We then focus on two different types of tasks: (i) detection and (ii) quantification, and seek to assess and optimize performance of the generalized RM framework. We discuss these two tasks next.

2.2 Detection task performance

RM results in increased inter-voxel correlations, as demonstrated by a number of groups including ours [5-8]. In fact, a number of experiments were performed in the distant past to quantify the impact of correlated noise (as induced by tomography) on human performance [19-23]. Overall, it was observed that human observer task performance was degraded by noise correlations. In fact, prior to these studies, it had been hypothesized by Wagner [36] that human observers would be inefficient when presented with noise-correlated structure, and that a non-prewhitened matched filter (NPWMF) as an observer for detection might better predict human performance compared to an observer performing prewhitening, which turned out to be true in later studies [22].

In the present work, we perform assessment of the impact of RM using both PWMF and NPWMF observers. Let us consider a signal- and background-known-exactly (SKE/BKE) ideal observer study of an ROI of size n with detected signal-absent s1 and signal-present s2 distributions (which are blurred versions of original signals f1 and f2), and with the addition of noise become g1 and g2, respectively. For a given image vector g of size n, not knowing whether it is signal-absent or present, the PWMF operator performs the following:

$$\lambda_{PWMF}(\boldsymbol{g}) = \Delta \boldsymbol{s}^T \boldsymbol{K}^{-1} \boldsymbol{g} \tag{1}$$

by first applying the inverse of the n by n noise covariance matrix K, followed by the matched filter Δs = s2 – s1 (for an ideal observer. This difference signal is known exactly; otherwise, this can be computed by training, performing averaging and subtraction of a set of signal-present and signal-absent images, arriving at $\Delta\bar{g}$). By contrast:

$$\lambda_{NPWMF}(\boldsymbol{g}) = \Delta \boldsymbol{s}^T \boldsymbol{g} \tag{2}$$

To perform ROC analysis, two general approaches can be utilized, following generation of various noise realizations of images with signal absent vs. present, and application of the abovementioned PWMF or NPWMF operators: (Method 1) One may compute the inter-class (signal absent vs. present) signal-to-noise ratio (SNR). Then, under assumption of normality of the observer metric, e.g. a linear observer under Gaussian noise, the area-under-curve (AUC) of ROC analysis (true positive fraction vs. false positive fraction) can be shown to be given by [25]:

$$AUC = \frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(\frac{SNR}{2}\right) \tag{3}$$

where erf(z) represents the error function. For a general linear operator $\lambda(\mathbf{g}) = \mathbf{w}^{\mathrm{T}}\mathbf{g}$, where $\mathbf{w}$ has the same dimensions as the image vector g, the inter-class SNR can actually be analytically derived ([25], see Eqs. 13.118 and 13.180):

$$SNR_{\mathbf{w}}^2 = \frac{\left[\mathbf{w}^{\mathrm{T}}\Delta\mathbf{s}\right]^2}{\mathbf{w}^{\mathrm{T}}\mathbf{K}\mathbf{w}} \tag{4}$$

It thus follows for the case of PWMF, that since $\mathbf{w}^{\mathrm{T}} = \Delta\mathbf{s}^{\mathrm{T}}\mathbf{K}^{-1}$ and $\mathbf{w} = \mathbf{K}^{-1}\Delta\mathbf{s}$:

$$SNR^2_{PWMF}(\boldsymbol{g}) = \Delta \boldsymbol{s}^T \boldsymbol{K}^{-1} \Delta \boldsymbol{s} \quad (5)$$

In fact, (4) attains its maximum in this case ([25], Eq. 13.184). By contrast, for NPWMF, where $\mathbf{w} = \Delta\mathbf{s}$, we have:

$$SNR^2_{NPWMF} = \frac{\left[\Delta \mathbf{s}^{\mathrm{T}} \Delta \mathbf{s}\right]^2}{\Delta \mathbf{s}^{\mathrm{T}} \mathbf{K} \Delta \mathbf{s}} \quad (6)$$

(Method 2) Alternatively, one may use varying diagnostic thresholds or cut-offs to generate ROC curves, from which the area-under-curve (AUC) may be computed. In the present work, we implemented both approaches and verified nearly similar performances. However, we prefer the latter approach: this is especially because it does not make the assume normality, as mentioned before Eq. (3), and furthermore, for NPWMF, it does not require computation of the covariance matrix K, nor its inverse, as it only computes (2), followed by AUC computation, unlike (6) as needed by (3).

2.3 Quantification task performance

We generated noise vs. bias trade-off curves, as generated with increasing iterations into the various generalized RM algorithms. We also performed convergence analysis, in which an algorithm was quantified as converged, when the last 10 iterations altered the bias no more than a certain threshold. We then computed the coefficient-of-variability (COV), which is the percentage of $\sigma_{ensemble}$ compare to the mean uptake. We further include the contrast vs. noise trade-off. The two noise metrics were specifically defined as: (i) Spatial variance $\sigma^2{}_{spatial}$, calculated for an image at a given noise realization (in the case of multiple noise realization measurements, this expression is subsequently averaged) via the following:

$$\sigma^2_{spatial} = \frac{1}{N-1}\sum_{i=1}^{N}\left(s_i - m\right)^2 \quad (7)$$

Where s_i denotes the image values at any voxel i within a given ROI (e.g. tumor) consisting of N voxels and having a mean m. In the case of multiple noise realization measurements, this expression is subsequently averaged.

(ii) Ensemble variance of ROI mean uptake m_r across multiple noise realizations r=1…R, with the average ROI mean uptake $\bar{m}$:

$$\sigma^2_{ensemble} = \frac{1}{R-1}\sum_{r=1}^{R}\left(m_r - \bar{m}\right)^2 \quad (8)$$

3. RESULTS

Figures 1-2 show comparison 3D trade-off plots of AUC vs. $\hat{h}$ FWHM (in mm) and deconvolution iterations, including 3000 noise realizations, for the NPWMF and PWMF, respectively. Figures 3-4 depict the calculated AUC for a single iteration from SNR using error function (method 1), and from ROC curves (method 2), respectively, both yielding nearly similar curves. Overall, we observe optimal performance when $\hat{h}$ (4mm) underestimates the true h = 6mm, or slightly overestimates it (7mm); the latter is attributed to edge artifacts (Fig. 5) which can actually enhance AUC. Nonetheless, it is clearly seen that NPWMF, a more realistic model of human performance, performs increasing poorly with increasing $\hat{h}$ values, as they introduce increased inter-voxel correlations that can only be accounted for properly using PWMF.

Figures 6 and 7 depict SUV_{mean} COV and SUV_{max} COV vs. SUV_{mean} bias; respectively. In these two figures, neither of no RM $\hat{h}$=0 nor full RM ($\hat{h}$= h=6) achieve the best noise-bias trade-off; yet Fig. 6 demonstrate the best performance achieved through $\hat{h}$=4. The higher SUV_{max} COV in Fig. 7 compared to SUV_{mean} COV in Fig. 6 is due to the greater Gibbs ringing artifacts for higher $\hat{h}$. From Figures 6 and 7 we observe the superiority of $\hat{h}$ values close to h, as well as slightly overestimated and underestimated values as described above. Figure 8 contains noise-contrast trade-off plots, and shows higher contrast for higher $\hat{h}$ as a result of greater overshoot of the edges. This effect can be observed more in detail in Fig. 9, where the CRC vs. iterations is plotted. This plot shows how contrast increases faster for higher $\hat{h}$ for a fixed iteration.

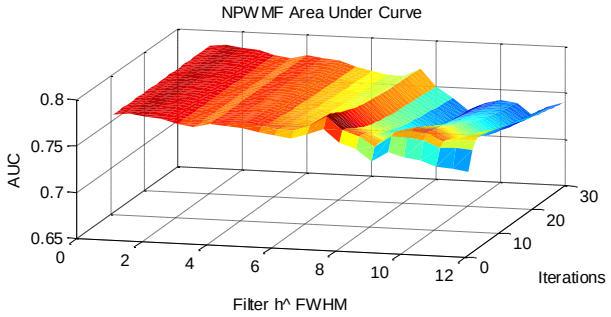

Figure 1: comparison trade-off plots of AUC vs. $\hat{h}$ FWHM (in mm) and deconvolution iteration

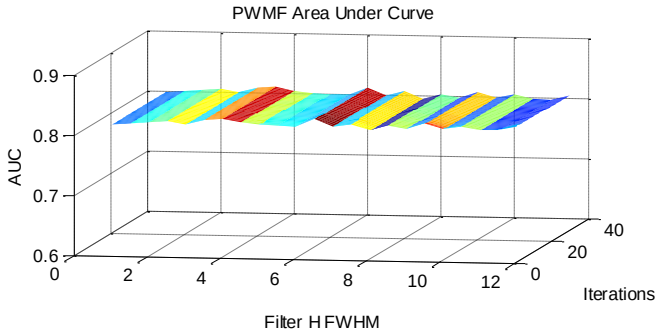

Figure 2: comparison trade-off plots of AUC vs. $\hat{h}$ FWHM (in mm) and deconvolution iteration

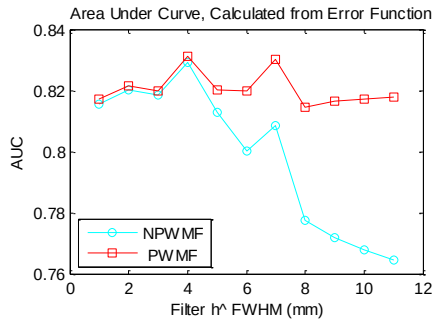

Figure 3: AUC calculated from SNR (method 1) and evaluated for both PWMF and NPWMF.

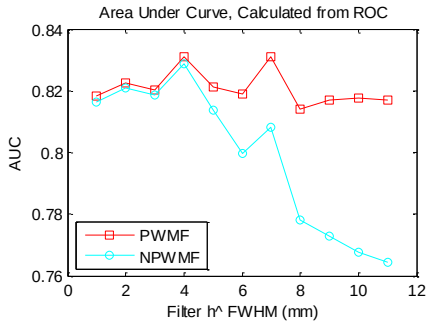

Figure 4:AUC of PWMF and NPWMF calculated from ROCs, respectively (preferred method 2).

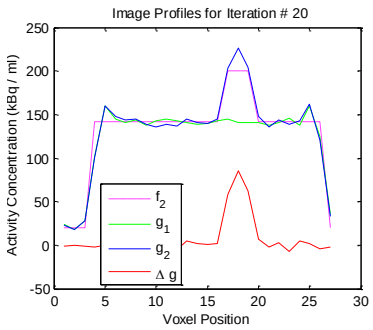

Figure 5: Single profile of $\boldsymbol{g_1}$, $\boldsymbol{g_2}$ and $\boldsymbol{\Delta g}$ and the true image for $\hat{h}$=*7mm*.

Table I. Converged Ensemble Noise Values.

$\hat{h}$ FWHM	0	1	2	3	4	5	6	7	8
COV	3.44	3.42	3.47	3.54	3.58	3.93	4.12	4.75	5.01
%change COV	0	-0.74	0.14	2.77	4.08	14.23	19.58	37.73	45.36

The converged ensemble noise values based on the criteria described in section 2.3 is shown in Table I. In this table, we see that $\hat{h}$ = *4mm* results in minimal increases in COV, with respect to no RM, while the impact is considerably higher for $\hat{h}$ = *6mm* and $\hat{h}$ = *7mm*. Overall comparison of above results reveals that underestimation of the true PSF can result in actually enhanced detection capabilities, while considerably lowering COV compared to full RM.

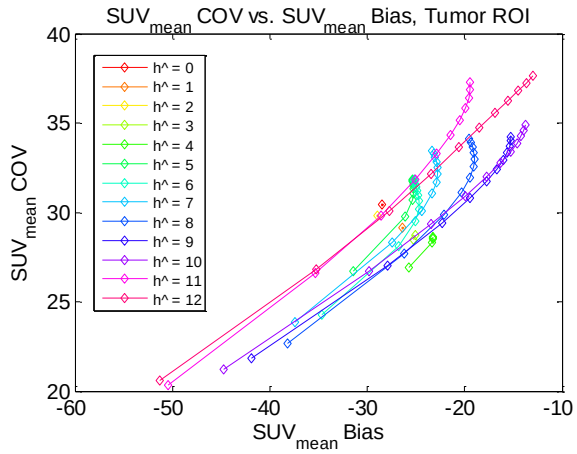

Figure 6: SUV_{mean} COV vs. SUV_{mean} bias trade-off for iterations 1-20

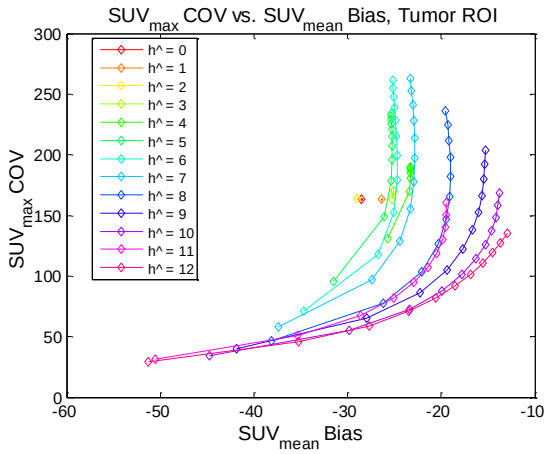

Figure 7: SUV_{max} COV vs. SUV_{mean} bias trade-off for iterations 1-20

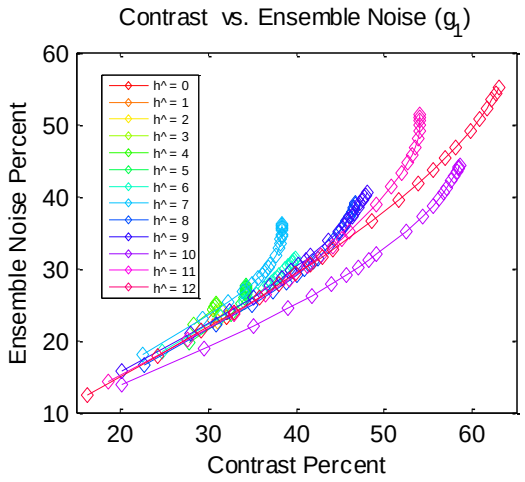

Figure 8: Ensemble noise percent of g_1 vs. contrast percent trade-off for iterations 1~100

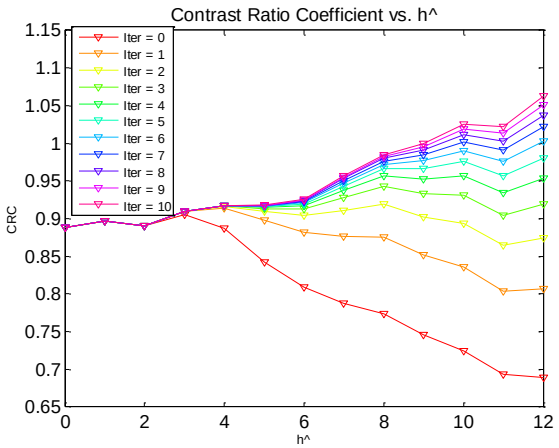

Figure 9: CRC vs. different $\hat{h}$ for iterations 1~100

4. ONGOING WORK AND CONCLUSIONS

We have proposed a generalized RM framework with extensive task-based optimization, to continualize the conventionally discrete framework of RM vs. no RM and apply various degrees of RM. We showed how this framework provides a trade-off between the enhanced contrast recovery by RM and the reduced inter-voxel correlations for no RM, and improves task performance. We run detection task performance for NPWMF and PWMF observers, while demonstrating two approaches to calculate AUC that converge to the same values. We assuming a blurring kernel with FWHM of h, and performed iterative EM including RM with varying widths $\hat{h}$, to demonstrate that an underestimated resolution blurring $\hat{h}$ enhances task performance. A slight overestimation also achieves enhanced performance for an ideal observer because of the Gibbs edge artifacts; however, in the case of the NPWMF – a realistic model of human performance – increasing values of $\hat{h}$ degrade performance. Overall, the results reveal that underestimation of the true PSF can result in actually enhanced detection capabilities, while considerably lowering COV compared to full RM, thus providing an attractive solution for both diagnostic and treatment response monitoring applications.

In future work, we consider a diverse range of tumor geometries, locations and contrasts. Furthermore, in addition to SUVmean analysis, we will assess SUV_{peak} in the context of the proposed framework. We will also extend abovementioned ROC analysis, to additionally include detection + localization as performed in LROC analysis, which should enable a more realistic assessment of the proposed framework in oncologic imaging.

5. ACKNOWLEDGMENTS

This work was in part supported by the NIH grant 1S10RR023623. We wish to thank Drs. Charles Watson, Mike Casey and Dan Kadrmas for helpful discussions.

REFERENCES

[1] Rahmim, A., Qi, J., and Sossi, V., "Resolution modeling in PET imaging: Theory, practice, benefits and pitfalls," Med Phys, 40(6), 064301-064442 (2013).
[2] Soret, M., Bacharach, S.L., and Buvat, I., "Partial-volume effect in PET tumor imaging," Journal of Nuclear Medicine 48(6), 932-945 (2007)
[3] Rousset, O.G., Rahmim, A., Alavi, A., Zaidi, H.,"Partial volume correction strategies in PET," PET Clinics (2)1, 235-249 (2007)

[4] Erlandsson, K., Buvat, I., Pretorius, PH., Thomas, BA., Hutton, BF., "A review of partial volume correction techniques for emission tomography and their applications in neurology, cardiology and oncology," Physics in Medicine and Biology 57(21), R119-R159 (2012)
[5] Rahmim, A., Cheng J.-C., and Sossi, V., "Improved noise propagation in statistical image reconstruction with resolution modeling," Nuclear Science Symposium Conference Record, 2576-2578 (2005)
[6] Tong, S., Alessio, A.M., and Kinahan, P.E., "Noise and signal properties in PSF-based fully 3D PET image reconstruction: an experimental evaluation," Physics in Medicine and Biology, 55(5), 1453-1473 (2010)
[7] Sureau, F.C., Reader, AJ., Comtat, C., Leroy, C., Ribeiro, MJ., Buvat, I., Trébossen, R., "Impact of image-space resolution modeling for studies with the High-Resolution Research Tomograph". J Nucl Med, 49(6),1000-1008 (2008)
[8] Rahmim, A., and Tang, J., "Noise propagation in resolution modeled PET imaging and its impact on detectability," Med Phys Biol, 58(19), 6945-6955 (2013)
[9] Blinder, S.A., Dinelle, K., and Sossi, V., "Scanning rats on the high resolution research tomograph (HRRT): a comparison study with a dedicated micro-PET," Med Phys, 39(8), 5073-83 (2012)
[10] Hoetjes, N.J., van Velden, FH., Hoekstra, OS., Hoekstra, CJ., Krak, NC., Lammertsma, AA., Boellaard, R., "Partial volume correction strategies for quantitative FDG PET in oncology," European journal of nuclear medicine and molecular imaging, 37(9), 1679-1687 (2010)
[11] Vercher-Conejero, J., Vercher-Conejero J., Sher, A.,Kolthammer, J.,Kohan, A., Rubbert, C., Wojtylak, P.,Partovi, S., Gaeta, M., O'Donnell, J., "Contribution of the modeling of point spread function and time-of-flight technology on the image quality of a PET/CT system," J Nuc Med, 54(2), 101-122 (2013)
[12] Tylski, P., Stute, S., Grotus, N., Doyeux, K., Hapdey, S., Gardin, I., Vanderlinden, B., Buvat, I., "Comparative Assessment of Methods for Estimating Tumor Volume and Standardized Uptake Value in F-18-FDG PET," Journal of Nuclear Medicine, 51(2), 268-276 (2010)
[13] Reader, A.J., Julyan, P.J., Williams, H., Hastings, D.L., Zweit, J., "EM algorithm system modeling by image-space techniques for PET reconstruction," IEEE Trans Nuc Sci, 50(5),1392-1397 (2003)
[14] Snyder, D.L., Miller, MI., Thomas, L.J., Politte, DG., "Noise and edge artifacts in maximum-likelihood reconstruction for emission tomography," IEEE Transactions on Medical Imaging, 6(1), 313-319 (1987)
[15] Politte, D.G. and Snyder, D.L., "The Use of Constraints to Eliminate Artifacts in Maximum-Likelihood Image Estimation for Emission Tomography," IEEE Transactions on Nuclear Science, 35(1), 608-610 (1988)
[16] Tong, S., Alessio, A.M., Thielemans, K., Stearns, C., "Properties and Mitigation of Edge Artifacts in PSF-Based PET Reconstruction," IEEE Transactions on Nuclear Science, 58(5), 2264-2275 (2011)
[17] Watson, C.C., "Estimating effective model kernel widths for PSF reconstruction in PET". 2011 Ieee Nuclear Science Symposium and Medical Imaging Conference (Nss/Mic), 2368-2374 (2011)
[18] Karakatsanis, N.A., Lodge, M., Tahari, A., Zhou, Y., Wahl, R., Rahmim, A., "Dynamic whole body PET parametric imaging: I. Concept, acquisition protocol optimization and clinical application" Phys. Med. Bio, 48(20), 7391-7421 (2013)
[19] Judy, P.F., Swensson, R.G., and Szulc, M.,"Lesion Detection and Signal-to-Noise Ratio in CT Images," Medical Physics, 8(1), 13-23 (1981)
[20] Guignard, P.A., "A Comparative Method Based on ROC Analysis for the Quantitation of Observer Performance in Scintigraphy," Physics in Medicine and Biology, 27(9), 1163-1176 (1982)
[21] Burgess, A.E., "Statistical Efficiency of Perceptual Decisions," Proceedings of the Society of Photo-Optical Instrumentation Engineers 454, 18-26 (1984)
[22] Myers, K.J., Barrett, H.H., and Borgstrom, M.C., "Effect of noise correlation on detectability of disk signals in medical imaging," Journal of the Optical Society of America, 2(1), 1752-9 (1985)
[23] Blackwell, K.T., "The effect of white and filtered noise on contrast detection thresholds," Vision Research, 38(2), 267-280 (1998)
[24] Wagner, R.F., "Decision theory and the detail signal-to-noise ratio of Otto Schade," Photogr. Sci. Eng., 22(1), 41-46 (1978)
[25] Barrett, H.H. and Myers, K.J., [Foundations of Image Science], Wiley and Sons Inc, Hoboken New Jersey, 1540-1540 (2004)