Comparison between TOF and non-TOF PET using a scan statistic numerical observer

Lucreţiu M. Popescu and Robert M. Lewitt

***Abstract*— Two dimensional time-of-flight (TOF) and non-TOF PET image sets are analyzed using a scan statistic (or maximum noise nodule search) observer, as well as a fixed position observer. The results reveal significant differences between the two approaches and indicate that the usual fixed position detectability evaluation techniques, often referred as signal known exactly background known exactly (SKE/BKE) evaluation tasks, insufficiently capture the true nature of image quality improvement that TOF information brings. We review here several results of the detection with localization theory and present a generalization using the noise nodule distribution in a given search area.**

I. Introduction

THe quality of medical images obtained from reconstructed data is difficult to evaluate due to the complex nature of their noise. Because the values of the image elements are correlated, a single scalar is not enough to describe the noise, as is often done in the literature. The differences between the image elements correlations are particularly important in positron emission tomography (PET) when time-of-flight (TOF) and non-TOF images are compared. Theoretically the noise can be described by the covariance matrix, but this is very difficult to evaluate (both theoretically and experimentally) and is also very difficult to work with, being usually very large. In conditions of (local) uniformity the covariance matrix can be substituted by more compact forms of expressing the same information, such as the image points standard deviation and the autocorrelation function (or the power spectrum). However, even in this case it is difficult to derive directly how well will be accomplished a given task (e.g. detection of small nodules), with the exception of a few oversimplified situations.

Because it is difficult to translate the imaging system physical characteristics into unequivocal image quality indicators, the task based performance evaluations have become a standard image quality evaluation practice. The most widely accepted task based evaluations are those that use human observers (or readers) performing a clinically relevant task. Unfortunately, the studies involving human observers are laborious and expensive, being used only in the advanced stages of a new medical imaging procedure development. For the routine evaluation needs, such as reconstruction algorithm optimization, researchers need to work with surrogate evaluation procedures that are easy to use and yet able to offer the necessary insight about the task performance.

Let A and B be two imaging methods we want to compare. And let T be the true task performance and M the task performance measured by our surrogate method. The task performance measuring method M should have the following properties relative to T.

- Monotonicity $M(A) > M(B)$ if $T(A) > T(B)$. This property is necessary to preserve the ranking order.
- Proportionality, or the relation with the true performance should be known. This property is necessary for a cost-benefit analysis.

The monotonicity property, in an acceptable degree, can be achieved by a range of evaluation procedures, from simple to complicated. The proportionality property is more difficult to achieve. We believe that the way towards this goal involves more realistic modeling of the detection mechanisms.

For meaningful task based image quality evaluations another important factor is the choice of the task itself. The radiological diagnostic tasks often require the detection of features (also called targets or signals) consisting of small regions with higher than usual activity concentration. Due to the data noise and the limited detector resolution these features appear in the reconstructed images as small bumps of higher concentration than the surrounding background. Such features sometimes appear the same as arbitrary noise clumps that normally occur in the image background. Therefore distinct detection criteria have to be considered, depending on whether the feature position is a priori known, or the feature must be searched for in a given image area (or volume).

In the localization known case we need to know the variations of a reduced region of the image. In this situation one can go to great lengths and consider in detail the statistical variations of our image spot, obtaining "ideal" observers. However, the variations with the position, within the same image, of equivalent image spots are more important. Here must be also noted that the position exactly known case rarely occurs in practice. In the case when the suspicious position is known, either from a different radiological modality, or a prior diagnostic, we still are confronted with spatial uncertainties due to the image registration errors, and a search of the suspicious feature within a small region of the order of a few cm^2 (or cm^3) still takes place.

In the localization unknown case we have signal-like features that randomly appear in the image background. For a detectability study we need to know the frequency of appearance, and the distribution of such features with the variables used also to describe the signal (contrast, size, etc.). Such information can be obtained by scanning the images.

II. Theory

A. The image scanning procedure

The image scan is done by smoothly sliding a scanning window over the area/volume of interest, and for each point

The authors are with University of Pennsylvania, Department of Radiology, 423 Guardian Drive, 4-th floor Blockley Hall, Philadelphia, PA 19104-6021, USA, e-mail: popescu@mipg.upenn.edu

computing some relevant local statistic of the image elements inside the scanning window. For instance the scanning window can be a disk (or a ball) of a fixed radius R and the statistic can be simply the sum of the values of the pixels inside the disk. Or, in order to eliminate the influence of background nonuniformity, one can use an enlarged scanning window and as statistic the local contrast given by the ratio between the average image values inside the disk and the average over an exterior annulus. More complex scanning windows and statistics can be devised, which can account for the searched feature size variation, or the rotation asymmetry, or their known typical activity profiles. However, when the searched feature is small, due to the limited system resolution its shape and even its original size become less relevant, therefore using a scanning window with a fixed radius seems a reasonable approach.

If we expect to find only one feature per searched area, or we are interested only in the most suspicious one, a detection strategy would consist of scanning the interest area and selecting only the location with the maximum scan value as suspicious. Further in the text this value will be referred to as the maximum scan (or max-scan) result.

By determining the distribution $g(c)$ of such maximum scan returns c for empty background cases (images with no features present) we can assess how usual or unusual a suspicious finding can be if we assume that the image is normal.

Such procedures are often called scan statistics tests and they are part of a statistics chapter actively studied [1]–[3]. These methods are preferred in applications where random fields must be searched for abnormal occurring local situations. Scan statistic techniques are commonly used in biostatistics and epidemiology [4] and more recently have also been applied in astronomy [5] and high energy physics [6]. In the medical image evaluation field Swensson [7] has used such maximum search value tests as basis for his localization receiver operating characteristic (LROC) theory.

B. Feature detection test

If the search features are of known size we can determine the scan response distribution $f(c)$ from multiple realizations, or from the the image reconstruction algorithm behavior, if such theory is available. With both distributions f and g known we can study the properties of the detection tests and plot the ROC type curves.

We denote the cumulative distributions of f and g as $F(c) = \int_{c_0}^{c} f(c')\mathrm{d}c'$, $G(c) = \int_{c_0}^{c} g(c')\mathrm{d}c'$, where c_0 is the lower limit of the scan response value c.

If the scan procedure returns a value c greater than a certain threshold value d we declare the location corresponding to the value c as a positive detection result, otherwise we declare that no feature/target/signal is detected.

If no feature is present the probability of the scan procedure to return a false positive result, that is a result c greater than the threshold d, is

$$P_0(d) = \mathrm{Prob}(c > d) = \int_d^{\infty} g(c)\mathrm{d}c = 1 - G(d) \qquad (1)$$

In the case when one feature is present, we have two random variables: a the feature realization value, and b the maximum scan obtained by searching the background. The scan will return the value $c = \max(a, b)$. The probability density that the scan procedure returns the value $c = a$ corresponding to the true target location (a true positive) is

$$h_{1\mathrm{L}}(c) = f(a = c) \times \mathrm{Prob}(b < c) = f(c)G(c)\,. \qquad (2)$$

And the probability of detecting a true target location, that is $c > d$, is

$$P_{1\mathrm{L}}(d) = \int_d^{\infty} h_{1\mathrm{L}}(c)\mathrm{d}c = \int_d^{\infty} f(c)G(c)\mathrm{d}c\,. \qquad (3)$$

It is assumed that the presence of a feature has negligible effect on the background and vice versa, which is a reasonable assumption if the feature is small and the background area is large compared to the feature size.

C. Image abnormality test

In many image evaluation tests one is interested only in whether or not the image is abnormal without requiring the correct localization of the suspicious features. The test is: if the returned max-scan value c exceeds the threshold d then the image is abnormal (positive), otherwise we declare it normal (negative).

In this case the probability for a false positive result is the same as in previous case given by (1). While if a feature is present the probability density of the returned scan values is

$$\begin{aligned} h_1(c) &= f(a = c) \times \mathrm{Prob}(b < c) + g(b = c) \times \mathrm{Prob}(a < c) \\ &= f(c)G(c) + g(c)F(c) \end{aligned} \qquad (4)$$

The probability of a true positive result (correct localization not required) is

$$\begin{aligned} P_1(d) &= \int_d^{\infty} h_1(c)\mathrm{d}c = \int_d^{\infty} f(c)G(c)\mathrm{d}c + \int_d^{\infty} g(c)F(c)\mathrm{d}c \\ &= 1 - F(d)G(d)\,. \end{aligned} \qquad (5)$$

If $P_1(c)$ is plotted against $P_0(c)$ we have the receiver (or relative) operating characteristic (ROC) curve, while if $P_{1\mathrm{L}}(c)$ is plotted against $P_0(c)$ we obtain the localization receiver operating characteristic (LROC) curve.

D. Fixed position feature detection test

In the case when the feature position is exactly known *a priori*, no search is required. In that case the scanning window is placed directly at the known position and the scan value is computed for that position only. This procedure can be seen as equivalent with a zero area search. We denote with $g_0(c)$ the background scan distribution for this case (and with $G_0(c)$ its cumulative distribution).

For the fixed position (or very small area search) detection we have the classical hypothesis test case. We have to determine

whether the scan return random variable c is due to the distribution $f(c)$ (the feature contrast variation) or $g_0(c)$ (normal background variation).

The probabilities for false positive and true positive results respectively, are

$$P_{0,\mathrm{fix}}(d) = \int_d^\infty g_0(c)\mathrm{d}c = 1 - G_0(d)\,, \quad (6)$$

$$P_{1,\mathrm{fix}}(d) = \int_d^\infty f(c)\mathrm{d}c = 1 - F(d)\,. \quad (7)$$

If $P_{1,\mathrm{fix}}(c)$ is plotted against $P_{0,\mathrm{fix}}(c)$ we have the particular case of the ROC curve for a priori known location case.

E. Generalization: noise nodule distribution model

In the scan procedure, instead of picking only the maximum contrast noise nodule from a scanned region of a given area, we can obtain a list of all noise nodules with values greater than a certain limit c_0. We assume that such noise nodules are a relatively rare occurrence and they are independent of each other.

Let us assume that after scanning image regions of a total area A_t we gather a list with a total N_t noise nodules. The average density of noise nodules with $c > c_0$ is $n_0 = N_t/A_t$.

By histogramming (or using other density estimation techniques) we can obtain the distribution $p(c)$ of the noise nodules contrast (for $c > c_0$) in the scanned image set. We have $\int_{c_0}^\infty p(c)\mathrm{d}c = 1$.

For a region of a given size A, that may be different from the size of the scanned regions, we have $N = n_0 A$ for the average number of noise nodules with $c > c_0$.

The average number of noise nodules with contrast larger than c (with $c > c_0$) in the area A is $\nu(c) = Nq(c)$, where $q(c) = \int_c^\infty p(c')\mathrm{d}c'$. Since the number k of noise nodules in a given image area is a random variable occurring at a constant rate, it is natural to assume (as in [8]) that it follows the Poisson distribution

$$P(k;c) = \frac{\nu^k}{k!}\mathrm{e}^{-\nu}\,. \quad (8)$$

This is mainly valid for large enough areas A. Alternatively and with similar results the analysis can be based on the binomial distribution $P_{\mathrm{B}}(k;c) = \binom{N}{k}[1-q(c)]^{N-k}q^k(c)$ that converges to the Poisson distribution for large N values.

The probability of having at least one noise nodule with contrast greater than c is

$$Q(1;c) = 1 - P(0;c) = 1 - \mathrm{e}^{-\nu} \quad (9)$$

The likelihood of having k nodules with contrast $c_1, \ldots, c_k$ greater than c is

$$L(c_1, \ldots, c_k) = P(k;c)\prod_{j=1}^{k}\frac{p(c_j)}{q(c)} = \frac{N^k}{k!}\mathrm{e}^{-Nq(c)}\prod_{j=1}^{k}p(c_j)\,. \quad (10)$$

We can derive the max-scan distribution $g(c)$ for a given search area A, from the noise nodule distribution $p(c)$. The probability of obtaining a max-scan result with the value c is given by the probability density of having one noise nodule with contrast c and all the other noise nodules with contrast less than c.

$$g(c) = P(1;c)\frac{p(c)}{q(c)} = Np(c)\mathrm{e}^{-Nq(c)} \quad (11)$$

The cumulative distribution is

$$G(c) = \mathrm{e}^{-Nq(c)} \quad (12)$$

We assumed that the searched area A is large enough so that $\mathrm{e}^{-N} \approx 0$.

III. Simulation study

We have studied the above feature detection approaches for two dimensional PET image reconstruction configurations. That enabled us to conveniently produce large numbers of image reconstruction (100-400) realizations and determine the distributions f, g_0, and g for different search area sizes by histogramming the results obtained by analyzing this large pool of images.

As phantoms we have considered activity distributions in the shape of disks of radius $R_p = 16\,\mathrm{cm}$. We have obtained hot feature phantoms by placing on the circle of radius $R_f = R_p/2$ a set of uniform circular features of radius $R = 0.5\,\mathrm{cm}$ and contrast 3:1 relative to the uniform background of the surrounding disk. The features were placed symmetrically at distances of minimum 6 cm one from another. Background only phantoms were obtained by omitting the hot features. In this study we have not considered any attenuation, scatter or randoms.

The data were generated [9] in list-mode with precise event positioning information. This fact enabled us, by randomly altering the longitudinal position information, to consider time-of-flight (TOF) with various timing precisions, as well as non-TOF image reconstruction situations.

The image reconstructions were performed using a list-mode ML-EM algorithm described in [10], with images represented using blobs — smooth overlapping basis functions [11] — on a rectangular grid (grid spacing $\Delta x = 0.4\,\mathrm{cm}$, Kaiser-Bessel blobs with radius $r_b = 2.4\Delta x$, and with parameters $m = 2$ and $\alpha = 6$). We used iterations over subsets made from consecutive events in the list. Further in the text a complete pass through the data will be referred to as an iteration. Multiple image realizations were obtained by using distinct data sets containing 6×10^4, 12×10^4 and 18×10^4 counts, split in 12 equal consecutive subsets, and using a relaxation parameter $\lambda = 0.8$ applied as an exponent to the update quantity [10].

For image analysis we have considered as statistic the local contrast c given by the ratio between the image average on a disk of radius $R_a = 0.5\,\mathrm{cm}$ and the image average on the annulus of inner and outer radii $R_{b1} = 0.6\,\mathrm{cm}$ and $R_{b2} = 2.0\,\mathrm{cm}$ concentric with the disk.

We have applied this local contrast calculation procedure on the hot features and background images in order to obtain the following results:

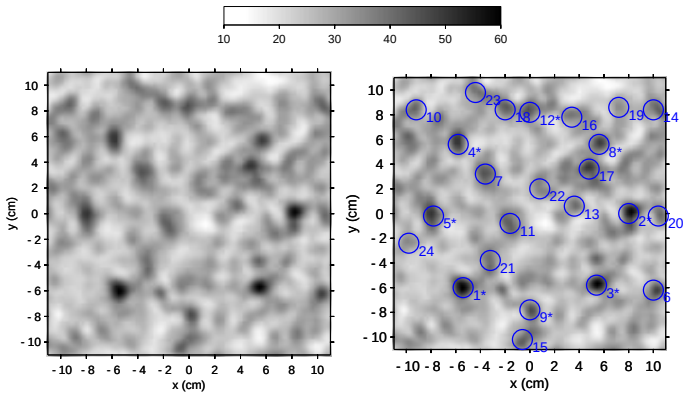

Fig. 1. Example of image scan (image reconstructed from 4×10^4 counts, non-TOF, iteration 3). On the left side is the original unmarked image. On the right side the nodules returned by the scan procedure are marked and numbered in decreasing order of their contrast. All nodules with contrast value greater than $c_0 = 1.2$ are returned. The eight hot features present in the original phantom are marked with a '*' symbol and are returned with the following order numbers: 1, 2, 3, 4, 5, 8, 9, 12.

- The hot feature realizations contrast distribution $f(c)$;
- The max-scan distributions $g(c, A)$, obtained by scanning regions of area A of the background-only images. We have considered an initial area A_0 in the form of a square of size $22 \times 22\,\mathrm{cm}^2$ and then we took regions of area size A_0 divided by 2, 4, 8, 16 and 32 respectively.
- The distribution $p(c)$ and the density n_0 of the noise nodules with $c \geq c_0$.
- The background variation $g_0(c)$ for a fixed position.

The scanning procedure comprises two steps. The first step is analogous with a filtering procedure and consists of computing an auxiliary scan image in which each point has the value of the scan result obtained with the scanning window centered on that point. In the second step we start by determining the maximum point of the scan image. Once this point is found the value is entered in a list and the pixels in the disk of radius R_a corresponding to the nodule at that position are masked. The procedure continues with the rest of the unmasked image pixels as long as the maximum values found exceed the lower limit c_0. In this manner a list is produced with the non-overlapping nodules in the decreasing order of their contrast values. An example is shown in Fig. 1.

IV. Results

A. Nodule detectability evaluation using scan statistic

In Fig. 2 we show comparisons between the local contrast distribution for the background at fixed positions $g_0(c)$, the background max-scan (or scan statistic) distribution obtained for several search area sizes $g(c, A)$, and the true feature contrast distribution $f(c)$. The plotted distributions have been obtained in a smooth form over a refined grid by using the density kernel estimation technique with the kernel $w(\xi) = (1-\xi^2)^2$, where $\xi = (c-c')/h$ with a window width $h = 0.08$. In Fig. 3 are compared the ROC curves for the feature detection

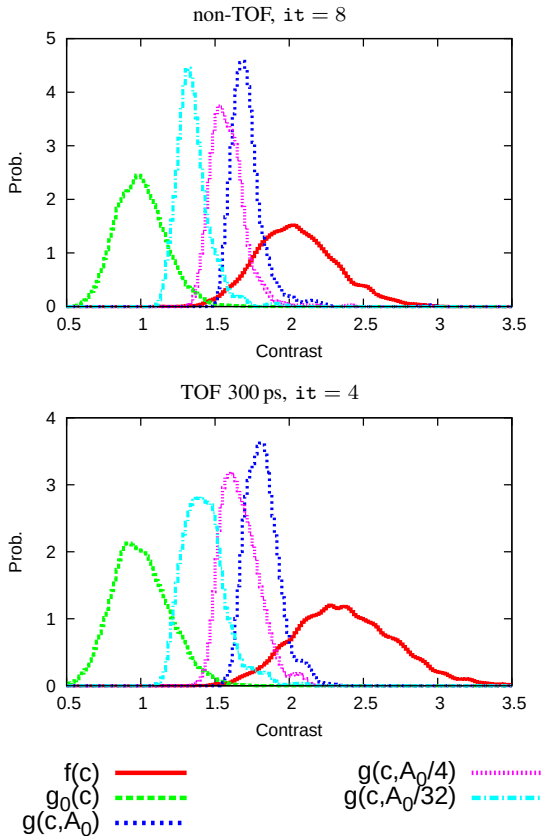

Fig. 2. Comparison between the local contrast distribution for the background at fixed positions ($g_0(c)$), background max-scan for several search area sizes ($g(c, A), A = A_0, A_0/4, A_0/32$) and the feature realization ($f(c)$). Reconstructions using $N = 6 \times 10^4$ counts, for non-TOF case (top) and TOF case for 300 ps timing resolution (bottom).

tests for the fixed location test and max scan test for several search area sizes.

In Fig. 4 we compare the max-scan distributions for search area size A_0, as well as for the background fixed position variation, and feature contrast variation for the non-TOF and the 300 ps TOF case. We show results obtained at iteration 8 in the case of non-TOF, and iteration 4 in the TOF case, results that are close to the convergence point of each case, respectively.

We can notice that there is little difference between the fixed position background variation curves for TOF and non-TOF. Some improvement can only be seen in the high count case $N = 18 \times 10^4$. The background max-scan distributions show that for the low counts case $N = 6 \times 10^4$ the TOF reconstructions produce noisier backgrounds than the non-TOF. However, as the number of counts increases to $N = 12 \times 10^4$ and $N = 18 \times 10^4$ this situation is reversed, and the TOF reconstruction tends to produce noise nodules with smaller contrast than the non-TOF case. Also the TOF reconstructions systematically produce higher contrast features, leading to the difference in detectability, as shown by the LROC curves in Fig. 5.

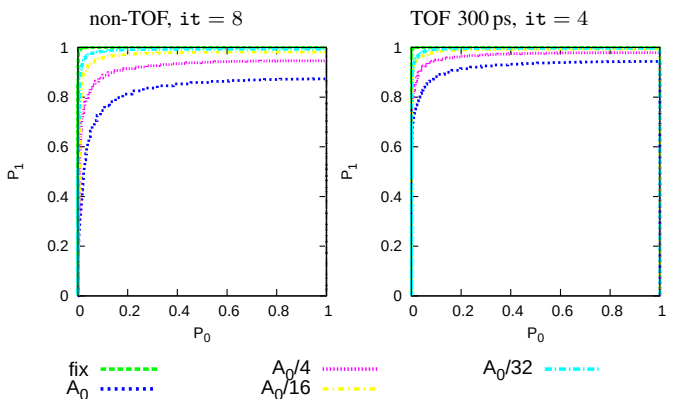

Fig. 3. Comparison between the LROC curves for the fixed position case versus background max-scan for several search area sizes A (same cases as in Fig. 2).

B. Direct determined max-scan distribution versus estimated max-scan distribution

The main drawback of the scan-statistic approach, as it was applied in the previous examples, is that it requires a large number of image samples in order to determine the distribution $g(c)$. One of the benefits of the generalized approach presented in section II-E is its more economic way of using the sample images, since from one scan are obtained multiple values for the determination of the noise nodule distribution $p(c)$, while with the max-scan procedure only one value per scan is obtained. In Fig. 6 we show comparisons between the noise nodule distributions $p(c)$ obtained for a reduced set of only 8 images and the full set of 400 images. As shown in the figure, both experimental histograms are well fitted by the tail of a Gaussian distribution $a \exp\left[-\frac{(c-\mu)^2}{2\sigma^2}\right]$ with mean $\mu = 1$.

In Fig. 7 and Fig. 8 are shown comparisons between the experimental max-scan distributions $g(c)$ obtained for different search area sizes and the estimates of $g(c)$ obtained from equation (11) using the curves fitting the experimental noise nodule distribution $p(c)$ from scanning areas of size A_0 with $c > 1.1$. The right column of each figure shows comparisons between experimental cumulative distributions $1-G(c)$ and the corresponding estimates from equation (12). The plots reveal a good agreement especially for the tail of the distributions and for the large search area cases.

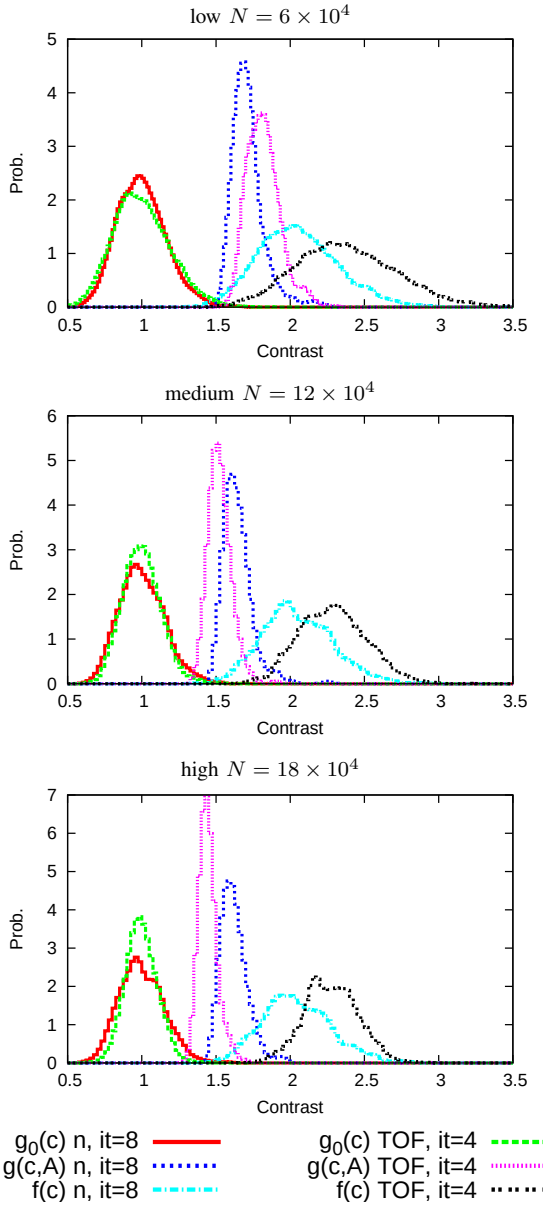

Fig. 4. Comparison between the TOF (300 ps) and non TOF local contrast distribution for the background at fixed positions ($g_0(c)$ n, $g_0(c)$ TOF), max-scan for a search area of size $A = A_0$ ($g(c, A)$ n, $g(c, A)$ TOF), and the feature contrast variation ($f(c)$ n, $f(c)$ TOF). We show here results for $N = 6 \times 10^4$, 12×10^4, and 18×10^4 counts.

V. Conclusions

The comparisons between scan-statistic based image evaluations and fixed position evaluations (the widely used SKE/BKE task evaluation case), reveal a reduced sensitivity of the latter with the TOF timing improvement. This is in agreement with similar conclusions reached in [12] where various image regularization techniques were tested.

In general the dependence on the search area size of the detectability of the features at unknown positions indicates that the image area size is a factor that should be carefully considered when the image evaluations are meant to produce results proportional to the real performance of the evaluated methods. If only the ranking order is of interest this factor is less important, yet the area search methods are more sensitive than the fixed position evaluations.

The generalized scan-statistic approach presented here has several advantages compared with the direct determination of the scan statistic (or max-scan) distribution from multiple realizations. One advantage is the greater efficiency, since one scan search obtains multiple values for the determination of the noise nodules distribution $p(c)$, while the max-scan procedure obtains only one value per scan. As was shown it can be applied

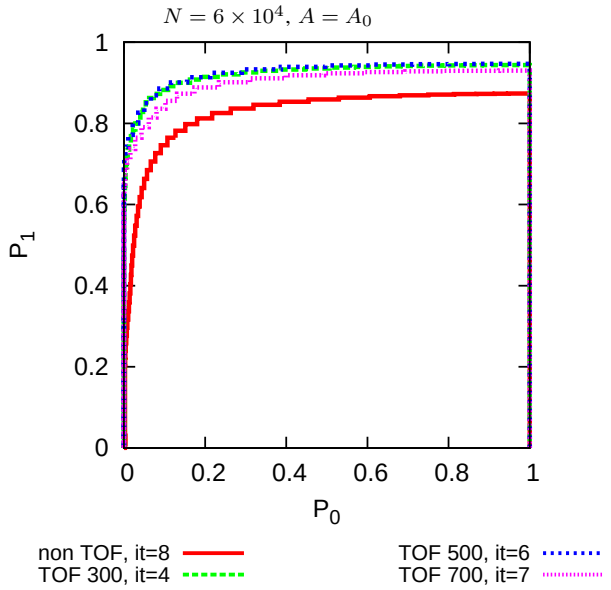

Fig. 5. Comparison between the LROC curves for images for different TOF resolutions (300, 500 and 700 ps and non-TOF) for a search area of size A_0. Results shown are obtained for $N = 6 \times 10^4$ counts.

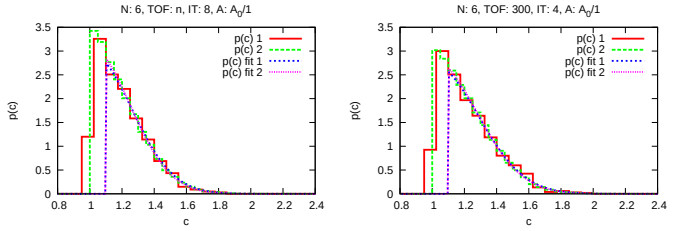

Fig. 6. Comparison between the noise nodule distributions $p(c)$ obtained for a reduced set of 8 images (1) and the full set of 400 images (2). Both reduced set and full set histograms are well fitted by the tail of a Gaussian distribution with mean set to unity. On the left is a non TOF case (6×10^4 counts, iteration 8). On the right is a TOF case (6×10^4 counts, 300 ps timing resolution, iteration 4.)

with only a few image samples. A further advantage is the greater flexibility, since the $p(c)$ distribution can be determined by searching areas of arbitrary size, and then the theory can be applied to a certain reference area size. Also, based on the noise model proposed the multiple targets per image case can be addressed using the equations presented here as we have shown in more detail in [13].

Acknowledgment

This work was supported by the National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health under grants R21-EB005434 and R33-EB001684.

References

[1] E. Parzen, *Modern Probability Theory and Its Applications*. New York: John Wiley and Sons, 1960.

[2] J. I. Naus, "Clustering of random points in two dimensions," *Biometrika*, vol. 52, pp. 263–267, 1965.

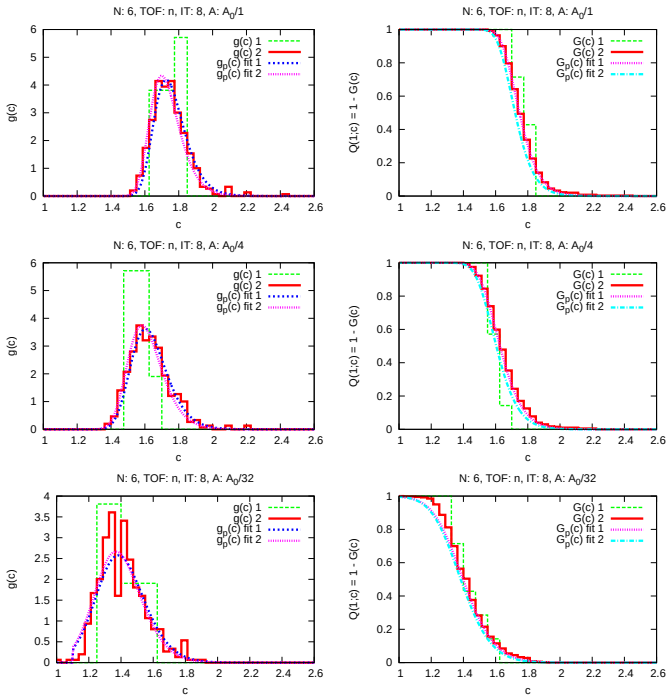

Fig. 7. Comparisons between the experimental and the estimated (through the $p(c)$ fit) max-scan distributions $g(c)$ (left) and the cumulative distributions $1-G(c)$ (right) obtained for different search areas. Pairs of curves show results obtained from both the reduced set of only 8 images (1) and the full set of 400 images (2). The non-TOF case (6×10^4 counts, non-TOF, iteration 8).

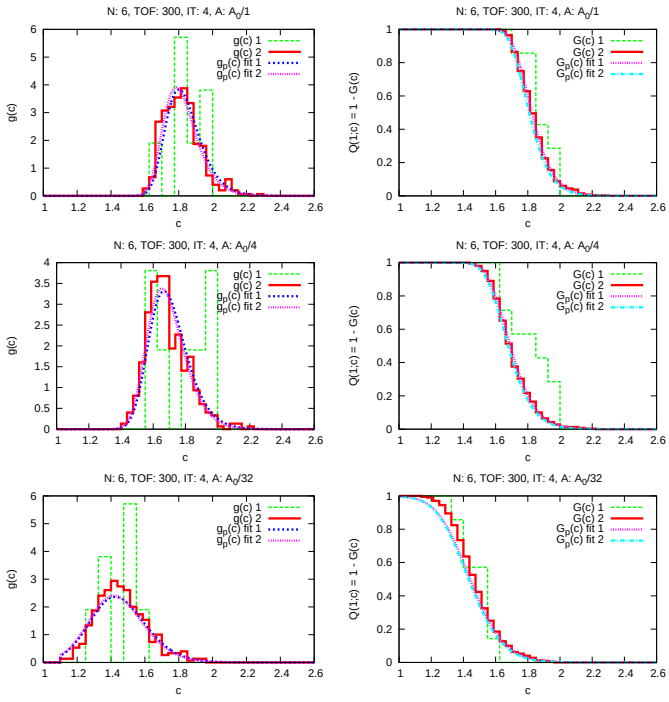

Fig. 8. Comparisons between the experimental and the estimated max-scan distributions $g(c)$ (left) and the cumulative distributions $1 - G(c)$ (right) obtained for different search areas. Pairs of curves show results obtained from both the reduced set of only 8 images (1) and the full set of 400 images (2). The TOF case (6×10^4 counts, 300 ps timing resolution, iteration 4)

[3] J. Glaz, J. Naus, and S. Wallenstein, *Scan Statistics*. New York: Springer, 2001.
[4] M. Kulldorff, "A spatial scan statistic," *Communications in statistics — Theory and Methods*, vol. 26, no. 6, pp. 1481–1496, 1997.
[5] K. J. Orford, "The analysis of cosmic ray data," *J. Phys. G: Nucl. Part. Phys.*, vol. 26, pp. R1–R26, 2000.
[6] F. Terranova, "Peak finding through Scan Statistic," *Nucl. Instr. Meth. Phys. Res. A*, vol. 519, pp. 659–666, 2004.
[7] R. G. Swensson, "Unified measurement of observer performance in detecting and localizing target objects on images," *Med. Phys.*, vol. 23, no. 10, pp. 1709–1725, 1996.
[8] P. C. Bunch, J. F. Hamilton, G. K. Sanderson, and A. H. Simmons, "Free-response approach to the measurement and characterization of radiographic-observer performance," *Journal of Applied Photographic Engineering*, vol. 4, no. 4, pp. 166–171, 1978.
[9] L. M. Popescu and R. M. Lewitt, "A versatile approach for Monte Carlo simulation of tomographic systems," in *IEEE Nucl. Sci. Symp. Conf. Rec.*, vol. 4, pp. 2785 – 2788, 19-25 Oct. 2003.
[10] L. M. Popescu, S. Matej, and R. M. Lewitt, "Iterative image reconstruction using geometrically ordered subsets with list-mode data," in *IEEE Nucl. Sci. Symp. Conf. Rec.*, vol. 6, pp. 3536 – 3540, 16-22 Oct. 2004.
[11] R. M. Lewitt, "Alternatives to voxels for image representation in iterative reconstruction algorithms," *Phys. Med. Biol.*, vol. 37, no. 3, pp. 705–716, 1992.
[12] A. Yendiki and J. A. Fessler, "Analysis of observer performance in known-location tasks for tomographic image reconstruction," *IEEE Trans. Med. Imag.*, vol. 25, no. 1, pp. 28–40, 2006.
[13] L. M. Popescu and R. M. Lewitt, "Small nodule detectability evaluation using a generalized scan statistic model," *Phys. Med. Biol.*, vol. 51, no. 23, pp. 6225–6244, 2006.