Binned Time-of-Flight Positron Emission Tomography

László Szécsi[1], László Szirmay-Kalos[1], Győző Egri[2], and Gergely Patay[2]

[1] Budapest University of Technology and Economics
szecsi@iit.bme.hu, szirmay@iit.bme.hu
[2] Mediso Medical Imaging Systems
Egri.Gyozo@mediso.hu, patay.gergely@mediso.hu

Abstract. This paper describes the *binned time-of-flight* reconstruction scheme for incorporating the measured time difference of detected gamma photon pairs in PET scanners, without list mode processing of individual events. Events are not summed for every detector crystal pair (forming a Line Of Response or LOR), but for time bins along LORs, retaining time difference information — which can only be measured with low accuracy — to a reasonable degree. Thus our approach avoids extreme storage and processing requirements, but allows for faster convergence and lower residual error of the reconstructed volume. We describe how various computation steps of non-time-of-flight reconstruction must be adjusted to accommodate for the binned time information.

1 Introduction

In emission tomography we need to find the spatial density of radioactive tracer materials [6]. The tracer material undergoes radioactive decay, producing a positron, which, upon meeting an electron, annihilates. A photon pair is emitted into (nearly) opposing directions. The photons may undergo scattering in the measured volume, absorption, scattering in the detector crystals, where they can be finally detected. Thus, tomography reconstruction is the *inverse problem* of particle transport calculation in scattering and absorbing media, which requires the iteration of particle transport calculations and corrective back projections.

In *Time-of-Flight* (TOF) reconstructions, we exploit the measured time difference of the two coinciding hits detected in each *Line of Response* (LOR). Traditionally, this required that detection events are handled individually [4], which is known as *list mode reconstruction.* List mode reconstruction algorithms require an input which is proportional to the number of detected events in size, and the running times are in the order of hours at least [5]. In our *binned reconstruction* approach, time of flight data of individual events are aggregated in *bins*, where a LOR is decomposed into N_{TOF} bins. This can be seen as creating histograms of TOF measurements in each LOR. This will increase the input data size only by a factor of the number of the bins per LOR, and the overhead above non-TOF reconstruction will be manageable. The accuracy of time measurement in PET scanners is around 300 picoseconds, corresponding to a spatial variance of 90 mm. Assuming Gaussian noise, this measurement error makes the estimator of the event location a wide bell curve, which can be approximated with small additional error using quadrangles.

The time difference of photon impact events must be considered a signed quantity to avoid ambiguity. The sign depends on the choice of primary and secondary detector modules, thus we need to uphold a convention on where a LOR starts or ends. It is straightforward to choose the detector with lower index as primary and the one with the higher index as secondary.

During measurements, in LORs, we collect $(y_1^i, y_2^i, \ldots, y_{N_{\mathrm{LOR}}}^i)$, where y_L^i is the number of hits in LOR L observed in TOF window $[t_i, t_{i+1})$. Generally, the time binning $[t_i, t_{i+1})$ may be different in each LOR or may be globally defined. In practice, we opt to keep it globally the same, as this avoids introducing extra data or logic, and allows us to effectively perform summation or filtering operations over the LOR histograms.

The expectation of the number of detected hits in LOR L in time window $[t_i, t_{i+1})$ is:

$$\tilde{y}_L^i = \int\limits_{t=t_i}^{t_{i+1}} \int\limits_{\mathcal{V}} x(\boldsymbol{v}) \mathcal{T}(\boldsymbol{v} \to L) p_L(\boldsymbol{v}, t) \mathrm{d}v \mathrm{d}t, \tag{1}$$

where $\mathcal{V}$ is the volume of interest, $\mathcal{T}$ is the system sensitivity describing the probability density of detecting a pair of photons by LOR L provided that they were born in point $\boldsymbol{v}$, and $p_L(\boldsymbol{v}, t)$ is the probability density that photons born in point $\boldsymbol{v}$ arrive in LOR L with time delay t.

Discretizing the volume to voxels, i.e. expressing the unknown activity distribution as a finite function series

$$x(\boldsymbol{v}) = \sum_{V=1}^{N_{\mathrm{voxel}}} x_V b_V(\boldsymbol{v}), \tag{2}$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_{N_{\mathrm{voxel}}})$ are the unknown coefficients and $b_V(\boldsymbol{v})$ $(V = 1, \ldots, N_{\mathrm{voxel}})$ are *basis functions*, this can be written as

$$\tilde{y}_L^i \approx \sum_{V=1}^{N_{\mathrm{voxel}}} x_V \cdot \int\limits_{\mathcal{V}} b_V(\boldsymbol{v}) \mathcal{T}(\boldsymbol{v} \to L) \mathrm{d}v \cdot \int\limits_{t=t_i}^{t_{i+1}} p_L(\boldsymbol{v}_V, t) \mathrm{d}t. \tag{3}$$

where $\boldsymbol{v}_V$ is the center of voxel V.

Probability density $p_L(\boldsymbol{v}, t)$ is assumed to be Gaussian of mean l/c where l is the distance difference traveled by the two photons and c is the speed of light, and of constant variance σ_T^2. The constant variance is a system parameter that describes the accuracy of time measurement of the tomograph:

$$p_L(\boldsymbol{v}, t) = \frac{1}{2\pi\sigma_T} e^{-\frac{(t-(l_1-l_2)/c)^2}{2\sigma_T^2}},$$

where l_1 and l_2 are the distances traveled by two photons.

Note that in the discretized formula (Equation 3), we need the integral of this probability density in different time bins:

$$\int_{t=t_i}^{t_{i+1}} p_L(\boldsymbol{v}_V, t)\mathrm{d}t = F(l_1 - l_2, t_i, t_{i+1}),$$

where will call F the *TOF coefficient*.

If the same time bins are used in all LOR, just a pre-computed array (texture) is needed that encodes the TOF coefficient for each bin as a function of the distance difference. Instead of a using a texture, values can be computed on-the-fly using a fast approximation of the cumulative distribution function of the Gaussian[1]:

$$\mathrm{erf}(t) \approx 1 - \frac{1}{(1 + 0.278393t + 0.230389t^2 + 0.000972t^3 + 0.078108t^4)^4}. \tag{4}$$

Equation 3 can also be written in a matrix form:

$$\tilde{\mathbf{y}}^i = \mathbf{A}^i \cdot \mathbf{x}$$

where the system response is characterized by a *system matrix* $\mathbf{A}^i$ [3],

$$\mathbf{A}^i_{LV} = \int_{\mathcal{V}} b_V(\boldsymbol{v})\mathcal{T}(\boldsymbol{v} \to L)\mathrm{d}v \cdot F(l_1 - l_2, t_i, t_{i+1}) = \mathbf{A}_{LV} \cdot F(l_1 - l_2, t_i, t_{i+1}). \tag{5}$$

which defines the correspondence between voxel intensities $\mathbf{x} = (x_1, x_2, \ldots, x_{N_{\mathrm{voxel}}})$ and expected LOR bin values $\tilde{\mathbf{y}} = (\tilde{y}^i_1, \tilde{y}^i_2, \ldots, \tilde{y}^i_{N_{\mathrm{LOR}}})$. The meaning of matrix element $\mathbf{A}^i_{LV}$ is the probability that a photon pair born in a random point distributed with density $b_V(\boldsymbol{v})$ is detected by LOR L in time window $[t_i, t_{i+1})$. Note that the TOF coefficients that modify the original non-TOF system matrix $\mathbf{A}_{LV}$ only depend on the geometry, thus, when the system matrix is factored, only the geometric elements must be corrected with the TOF probabilities. Attenuation, in-detector scattering, and detector sensitivity will affect all bins along a LOR identically.

The task of the reconstruction is to find voxel intensities of $\mathbf{x}$ based on the *measured* incidents in TOF bins. The iterative optimization process alternates *forward-projection*:

$$\tilde{y}^i_L = \sum_{V=1}^{N_{\mathrm{voxel}}} \mathbf{A}^i_{LV} x_V, \tag{6}$$

then *back-projection*:

$$x'_V = \frac{x_V}{\sum_i \sum_{L=1}^{N_{\mathrm{LOR}}} \mathbf{A}^i_{LV}} \cdot \sum_i \sum_{L=1}^{N_{\mathrm{LOR}}} \mathbf{A}^i_{LV} \frac{y^i_L}{\tilde{y}^i_L} \tag{7}$$

in each of the $n = 1, 2, \ldots$ iteration step.

As F is a probability density, its integral is 1, thus the precomputed normalization factor

$$\mathbf{n} = \sum_i \sum_{L=1}^{N_{\mathrm{LOR}}} \mathbf{A}_{LV}^i = \sum_{L=1}^{N_{\mathrm{LOR}}} \mathbf{A}_{LV}$$

is the same as the one used in non-TOF reconstruction.

2 Computation steps with TOF

During the reconstruction process, when the direct contribution is computed in forward-projection (Equation 6) or when the back-projection is evaluated (Equation 7), we have to compute the distances between the voxel and the two detectors. In forward-projection, the contribution computed for the LOR must be distributed into bins according to their respective probabilities F. In back-projection, the same factors should be used to get the correction factor for a voxel from the per-bin correction ratios $y_L^i/\tilde{y}_L^i$. While in forward-projection this requires an effort proportional to the size of the output, which cannot be spared, in back-projection the evaluation of the sum over the bins is a performance-critical computation. As the back-projection model is not so crucial in the accuracy of the reconstruction, we can make further concessions. A finite support approximation of the Gaussian, or even a Dirac-delta can be used to decrease computational load. In the latter case, which we call *zero-order approximation*, only a single bin of a LOR is considered for a voxel. The back-projection can be formulated as

$$x'_V \approx \frac{x_V}{\sum_{L=1}^{N_{\mathrm{LOR}}} \mathbf{A}_{LV}} \cdot \sum_{L=1}^{N_{\mathrm{LOR}}} \mathbf{A}_{LV} \frac{y_L^i}{\tilde{y}_L^i}, \tag{8}$$

where i is the index of the bin in which the voxel is. Such an approximation is bound to produce edge artifacts at bin boundaries, as different LOR ratios are used to compute neighboring voxels in the output data. However, we observed that these artifacts vanish quickly as the iteration converges. Also, the rate of convergence in error measures was identical, if not slightly better, with the zero-order approximation. We present results later in Section 5.1.

In addition to the direct contribution, the reconstruction also uses terms responsible for scattering and random contributions. Random contribution is measured without time data, so no TOF information is available for it. Although it would be possible to compute the TOF distribution also for the scattered component, this calculation would significantly slow down the reconstruction and would result in high variance results. Thus, we ignore TOF for both random and scattered contributions and distribute the events according to simple rules. Two rules are worth considering:

1. The scattered and random contributions are uniformly partitioned into time slots.
2. The scattered and random contributions are partitioned into time slots proportionally to the direct contribution.

3 Binning convention

In order to avoid introducing a dependence on the investigated region of interest into the measurements, the binning must cover all of the length of the LORs. It is likely that the measured object will be located in the middle of the LORs, with sizeable regions devoid of activity near the detectors. Thus, a uniform subdivision of a LOR into same-size bins is not practical. A non-general, LOR-dependent subdivision would have its merits: e.g. *interleaved binning*, where the set of bin boundaries on all LORs is uniformly distributed, would alleviate any quantization artifacts that could arise from binning. However, as we also need to perform filtering over LORs to compensate for intra-crystal scattering, we argue that a general scheme for all LORs is more practical. Thus bin intervals are defined in terms of t, the signed time difference of photon detections.

Bins are located symmetrically, mirrored on the midpoint of the LOR (at $t = 0$). Please refer to Figure 1 for a depiction. The number and length of the bins, as well as the variance of time difference measurements, must be specified as a feature of the measurement setup along with detector geometry specifications. The first and the last bins are extended to include any remaining parts of LORs, or even the range outside of detectors, as outlying time difference readings cannot be excluded.

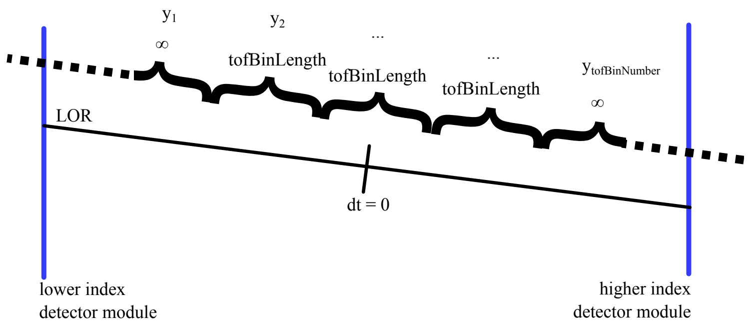


Fig. 1. Time-of-flight binning convention.

The input measurement file must specify readings for bins consecutively for every LOR. When read, the measurements are stored in memory in a similar manner, the only difference being that instead of the final bin reading, the sum of readings over the LOR is stored. This allows us to find the total without summing the bins, and in particular, to quickly identify LORs without activity. The value of the final bin can be computed quickly when all the bin values of the measurement are read anyway. Note that this trick should not be used for any other LOR sets, only the measured data, because it would not allow efficient vector operations over all bins.

4 Implementation

Forward-projection samples the volume either along LORs or using volume sampling. For every sample point on the LOR, time difference t must be computed, and the contribution to bins must be obtained using the TOF coefficients. We use an approximating function "cerf()" (see Equation 4) for the cumulative distribution of the Gaussian to evaluate TOF coefficients.

For back-projection the ratio of measured and computed LOR values must be computed. This can be done for all bins separately. If the measured summed LOR activity is zero, all bins can be skipped, writing zeros into the result LOR image. Care must be taken to compute the measured activity value in the final bin subtracting the rest of the bins from the sum. Voxel-centric back-projection evaluates all LORs crossing a voxel. The t time difference corresponding to the voxel location on the LOR must be computed. The correction ratios computed for all bins must be weighted according to TOF coefficients.

We may use the zero order approximation from the TOF coefficients in back-projection, as measurements show that the performance gain is enormous while error measures do not suffer, and artifacts vanish rapidly with an increasing number of iterations.

5 Results

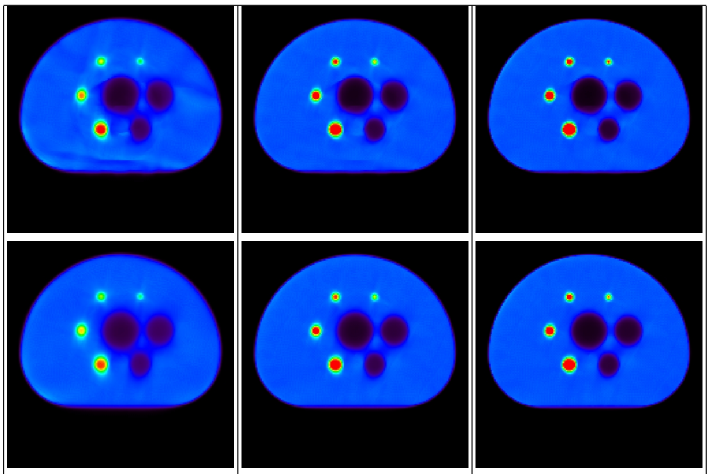
In this section we list and evaluate measurement results. All reconstructions used five bins. We used *Ordered Subset Expectation Maximization*[2], meaning that only a subset of LORs were used in every iteration, taking six iterations to cover all of them. Such six iteration are called an *OSEM cycle*.

5.1 Zero order approximation in back-projection

As discussed in Section 2, processing TOF data impacts the performance of back-projection heavily, but our zero-order approximation can sidestep this. Table 1 offers a visual comparison of artifacts during the reconstruction on the human IQ dataset. Initial artifacts vanish very fast with only a few OSEM cycles.

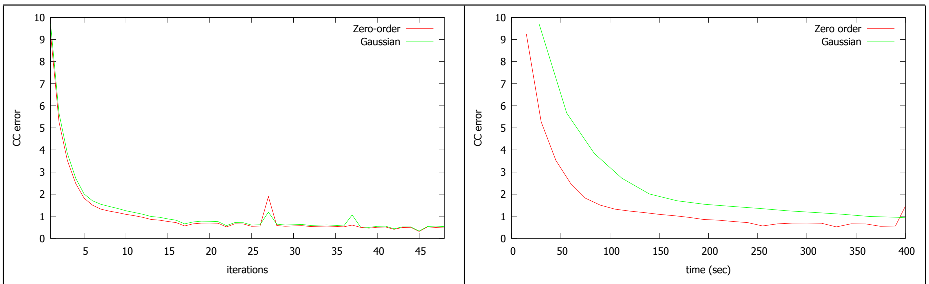
In Table 2 we compare the L2 and CC error measures describing the convergence of the reconstruction with and without the zero-order approximation. When observing convergence with the iteration count, the zero-order method does not perform any worse. While accurate computation slows down back-projection by a factor of the number of bins per LOR, the zero-order approximation only adds a smaller constant overhead.

The underlying assumption behind zeroth order approximation is that the correction ratios obtained by Gaussian smoothing of the values computed for the bins are close to those obtained from a single bin. If the variance of the time measurement is small, this is a good approximation, as the contribution of bins further away would be small. If the variance is large, than the computed

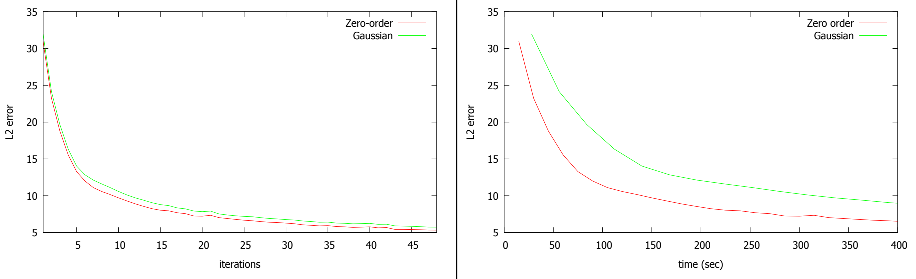


1 OSEM cycle. | 2 OSEM cycles | 3 OSEM cycles

Table 1. Comparison of reconstructed volumes for the first few iterations using and not using the zero order approximation (upper row) and full Gaussian (lower row) in the TOF back-projection.



CC error versus iteration count. | CC error versus reconstruction time.



L2 error versus iteration count. | L2 error versus reconstruction time.

Table 2. Comparison of zero order approximation in the back-projection kernel vs. applying the formulaic Gaussian. Convergence with iteration count is even slightly better and execution times are drastically reduced.

(forward projected) LOR values will be smooth, because forward projection applies the Gaussian formula. Therefore, whether our assumption holds depends on the characteristics of the measurement. If the empirical distribution of events measured in the bins does not follow the theoretical distribution (which would be very low-frequency), then the zeroth order approximation results in artifacts which do not diminish with iterations. We can conclude that the zeroth order approximation should not be used for measurements of poor characteristics. However, simulated measurements we could work with were well above the quality requirements needed for zeroth order approximation to work without persistent artifacts.

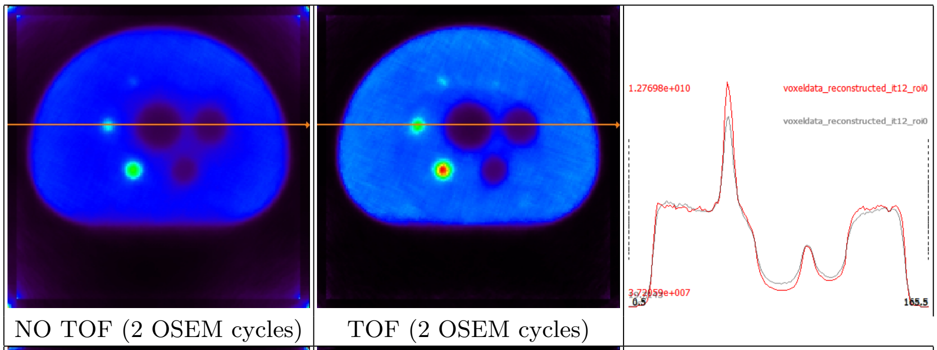
5.2 Comparison of TOF and non-TOF reconstruction

We compare reconstructions using binned TOF data against reconstructions of conventional measurements without such binning. Where GATE simulated measurements were available, we used them with appropriate settings, a bin size of 300 mm and time difference measurement variance also of 300 mm. Where reference projections were reconstructed (meaning that the input data was computed using our own forward-projector from the reference volume), we set the bin size to 80 mm in order to make the best use of temporal data in the activity region of the human IQ phantom. The variance was set to 160 mm. Thus, these can be considered to be idealistic TOF settings with little activity measurement noise. We used the zero order approximation in the back projection. Note the edge artifact along the circular arc in the middle of low-iteration count TOF-reconstructed volumes, and how it disappears in later iterations.

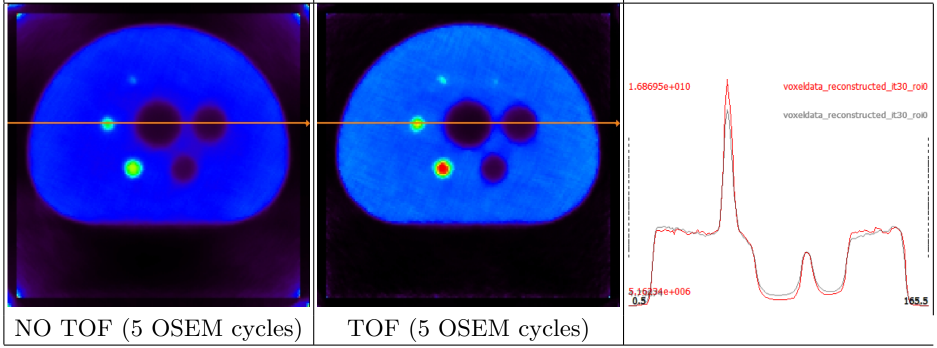
Table 3 presents result for geometry-only reconstruction, while in Table 4 results using the full reconstruction model can be seen. As only the geometric part of the system matrix is influenced by TOF in our model, gains are much more visible in the geometry only case, and are somewhat obscured by overhead of simulating physical phenomena. It can be concluded that TOF reconstructions consistently achieve lower errors in a converged state, and also faster convergence before that. Even if we account for the slower iterations, and chart errors versus reconstruction time (not including normalization array computation, which is independent of TOF) convergence is still superior. Interestingly, in case of noisy measurements, TOF reconstruction also increased stability, avoiding oscillation around the converged result. Visually, and also visibility in the line profiles, TOF reconstructions achieve clearer empty regions and more pronounced high-activity spots in the same number of iterations.

6 Conclusions

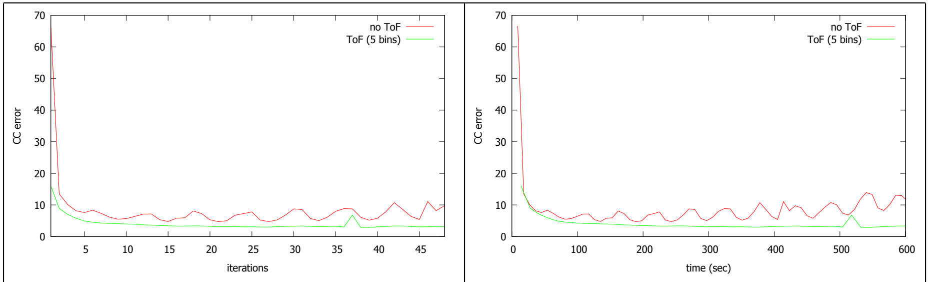
We have shown a binned approach to time-of-flight PET reconstruction, which does not require a list mode approach, but allows for an implementation which can even be faster than non-TOF reconstruction in reaching the same level of error. This way, even tomographs which feature lower time measurement resolution can take advantage of TOF information with little overhead.



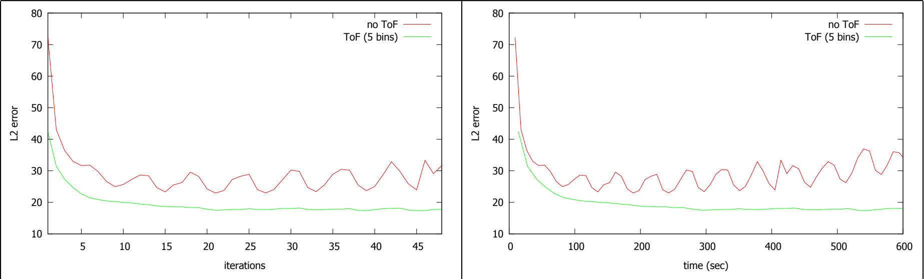
NO TOF (2 OSEM cycles) TOF (2 OSEM cycles)



NO TOF (5 OSEM cycles) TOF (5 OSEM cycles)

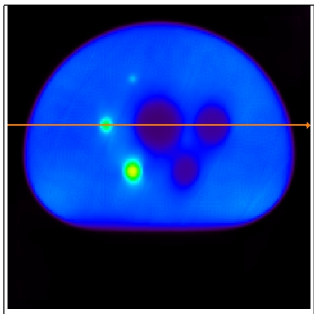
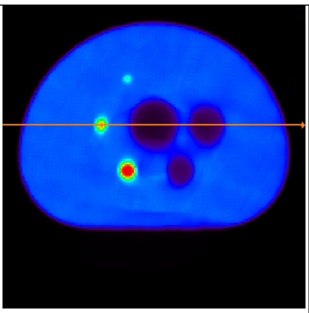
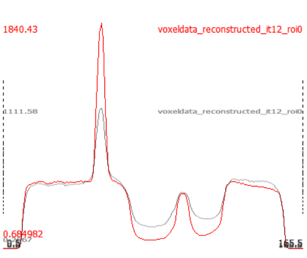
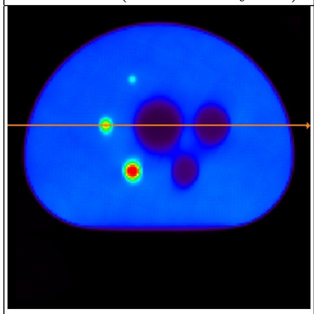
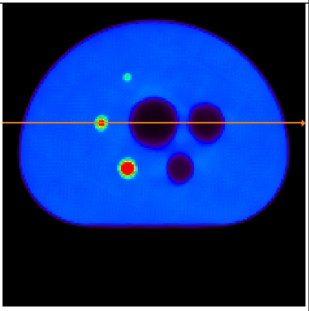
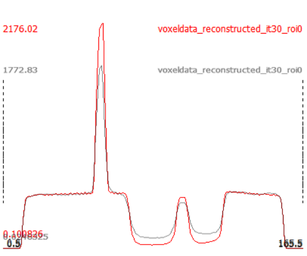


CC error versus iteration count. CC error versus reconstruction time.



L2 error versus iteration count. L2 error versus reconstruction time.

Table 3. Geometry only. Time bins are 5 × 300mm, the standard deviation of the Gaussian is $\sigma = 300$mm.

		
No TOF (2 OSEM cycles)	TOF (2 OSEM cycles)	
		
No TOF (5 OSEM cycles)	TOF (5 OSEM cycles)	

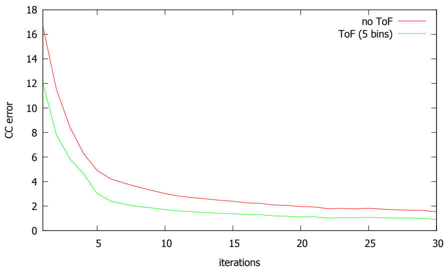
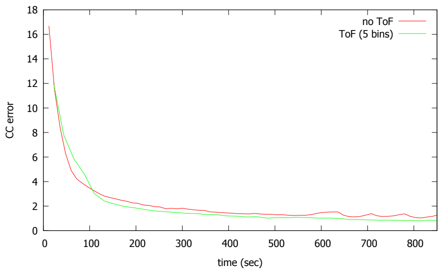
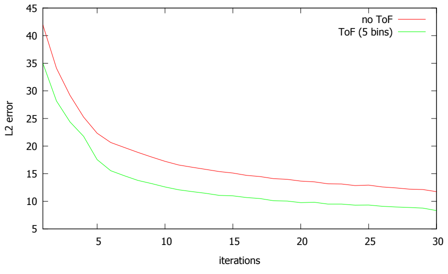
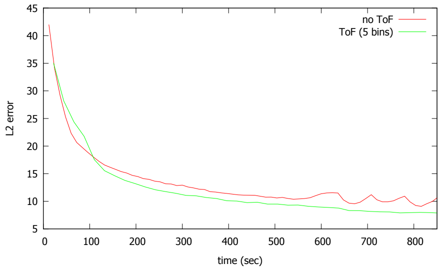
	
CC error versus iteration count.	CC error versus reconstruction time.
	
L2 error versus iteration count.	L2 error versus reconstruction time.

Table 4. Full reconstruction model (using reference projection). Time bins are 5 × 80mm and the standard deviation of the Gaussian is $\sigma = 160$mm.

More research is needed to verify how the method operates on actual measurements with possibly worse characteristics, and how the better visual quality and lower error would contribute to the diagnostic value of the reconstruction.

Acknowledgements

This work has been supported by TÁMOP-4.2.2.B-10/1-2010-0009, OTKA K-101527, OTKA K-104476, OTKA PD-104710, and the János Bolyai Research Grant. Part of the work has been carried out within the SPADnet project (www.spadnet.eu) supported by the European Community within the Seventh Framework Programme ICT Photonics.

References

1. M. Abramowitz and I.A. Stegun. *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, volume 55. Dover publications, 1965.
2. H.M. Hudson and R.S. Larkin. Accelerated image reconstruction using ordered subsets of projection data. *Medical Imaging, IEEE Transactions on*, 13(4):601–609, 1994.
3. C. A. Johnson, J. Seidel, R. E. Carson, W. R. Gandler, A. Sofer, M. V. Green, and M. E. Daube-Witherspoon. Evaluation of 3D reconstruction algorithms for a small animal PET camera. *IEEE Transactions on Nuclear Science*, 44:1303–1308, June 1997.
4. T.K. Lewellen. Time-of-flight pet. In *Seminars in nuclear medicine*, volume 28, pages 268–275. Elsevier, 1998.
5. W.W. Moses. Recent advances and future advances in time-of-flight pet. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 580(2):919–924, 2007.
6. A. J. Reader and H. Zaidi. Advances in PET image reconstruction. *PET Clinics*, 2(2):173–190, 2007.