Efficient Rendering of Regions of Response in List-Mode Reconstruction for PET

Giancarlo Sportelli, *Member, IEEE,* Juan E. Ortuño, and Andrés Santos, *Senior Member, IEEE*

***Abstract*—We have developed a new algorithm for fast region of response rendering in Positron Emission Tomography (PET). The algorithm can be applied efficiently to Expectation Maximization (EM) techniques that include per-event probability distributions with shapes more complex than symmetric tubes of response. The algorithm is generic for any system geometry and draws the scalar field associated to a generic input kernel with a constant cut-off threshold independent from the actual kernel shape. In the current version it has been tested for reconstructing List-Mode (LM) data simulated with GATE on the rPET small animal scanner and used with a 2D Gaussian kernel specifically designed for the rPET. EM parallelization has been achieved at per-event level on a 8 core dual CPU. Results show that the new algorithm allows for LM reconstruction in low times with complex kernel functions of the relative position of the voxel with respect to the line of response.**

I. Introduction

THE ability of reconstructing images iteratively by considering full raw acquired Positron Emission Tomography (PET) data, instead of compressing it in histograms, is known as one the promises of next PET reconstruction techniques. However, histogram mode reconstruction is still today the way to go in typical real world applications. The main reason comes from the high computational complexity in deriving a full emission probability spatial distribution for each acquired event. It is possible to reduce complexity by introducing approximations, as in Siddon based algorithms [1], but the trade off in accuracy can exceed the advantages of using per-event information.

Recently, highly parallelized, fast List-Mode (LM) reconstruction has been shown to be possible with more accurate tubes of response (TOR) [2], [3]. However, the proposed implementation is efficient only for tubes of response that are radially symmetric and whose section is constant along the line of response (LOR). If asymmetric tubes with 2D Gaussian section are used, the algorithm would benefit from a more specific tailoring of the region of response (ROR). The introduction of additional information as acolinearity, time of flight and energy correction would make a specific region tailoring approach even more advantageous.

In this work, we present a new efficient algorithm suitable for parallel LM reconstruction, able to produce accurate regions of response defined with a constant probability cut-off on its elements and a generic modeling kernel. The algorithm, with a variable 2D Gaussian kernel, has shown in simulations that can outperform image quality in traditional reconstruction algorithms based on precomputed system matrices with a very low CPU time per processing unit [4]. Our tests also showed how tuning the ROR size can have dramatic effects in reconstruction speed with negligible effects on image quality.

Manuscript received November 15, 2011. This work was partially supported by Spain's Ministry of Science & Innovation through CDTI - CENIT (AMIT project) and INNPACTO (PRECISION project), Instituto de Salud Carlos III (PI09/91058 & PI09/91065) and projects TEC2010-21619-C04-03 &TEC2008-06715-C02-02, Comunidad de Madrid (ARTEMIS S2009/DPI-1802), and the European Regional Development Funds (FEDER). CIBER-BBN is an initiative funded by the VI National R&D&i Plan 2008-2011, Iniciativa Ingenio 2010, Consolider Program, CIBER Actions and financed by the Instituto de Salud Carlos III with assistance from the European Regional Development Fund.

G. Sportelli, J. Ortuño and A. Santos are with the CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Spain, and with the Biomedical Image Technologies, Depto. de Ingeniería Electrónica, Universidad Politécnica de Madrid, Spain (e-mail: gsportelli, juanen, andres@die.upm.es).

II. Materials and Methods

The reconstruction process is based on the one-pass list-mode EM algorithm [5], in which the emission probabilities are derived through a fast voxel search and compute algorithm. The probability associated to each voxel is calculated with a generic projector model, or kernel, that is defined according to the modeled system response [2], [3], [4]. The novelty of the introduced method consists in the possibility of identifying the ROR dynamically depending on the modelled probabilities, without any assumptions on its axial section. The advantage is that the processing time per processor unit can be reduced at no cost in terms of used memory or accuracy.

A. The LOR Space Frame

The projector model, i.e. the scalar field that describes the TOR, has been derived with respect to a reference system in which the z axis is aligned with the ideal LOR, x and y axes orthogonal to the z axis and projected from the detectors axes. For any LOR, a 3D transformation is calculated to convert the coordinates of every voxel center $\mathbf{p} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ on the new reference frame $\mathbf{p}' = x'\mathbf{i}' + y'\mathbf{j}' + z'\mathbf{k}'$. In homogeneous coordinates, the transformation can be expressed as a 4×4 matrix M, composed of a rotation matrix R and a translation vector $\mathbf{t}$:

$$\begin{bmatrix} \mathbf{p}' \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} \mathbf{p} \\ 1 \end{bmatrix} = \mathbf{M} \begin{bmatrix} \mathbf{p} \\ 1 \end{bmatrix} \qquad (1)$$

The elements of the rotation matrix are the normal vectors of three perpendicular planes: a) the plane perpendicular to the LOR; b) the plane that contains the LOR and principal axis of the pixelated crystals; c) the plane perpendicular to a) and b). Thus, it holds that:

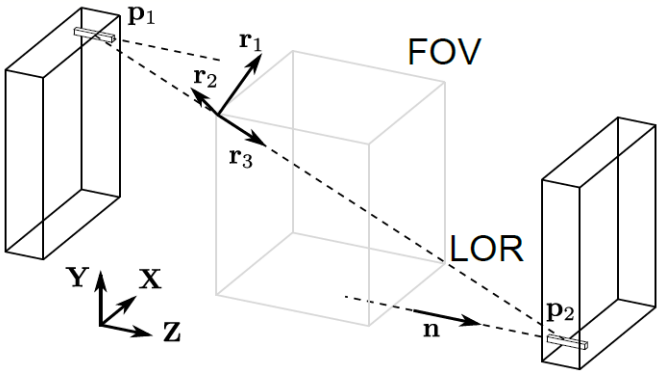

Fig. 1: Geometric diagram of the vectors involved in the change of reference system.

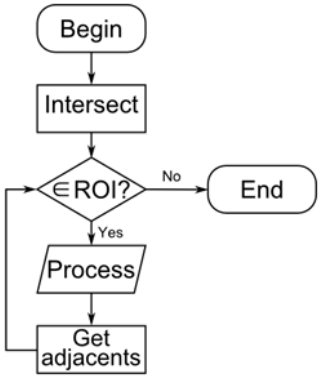

Fig. 2: Simplified flow chart of the voxel search algorithm.

$$\mathbf{r}_3 = \frac{\mathbf{p}_2 - \mathbf{p}_1}{|\mathbf{p}_2 - \mathbf{p}_1|}, \qquad \mathbf{r}_2 = \frac{\mathbf{r}_3 \times \mathbf{n}}{|\mathbf{r}_3 \times \mathbf{n}|}, \qquad \mathbf{r}_1 = \mathbf{r}_2 \times \mathbf{r}_3 \quad (2)$$

where $\mathbf{p}_1$ and $\mathbf{p}_2$ are the vertices of the LOR and $\mathbf{p}_3$ is a point belonging to the principal axis of one of the pixelated crystal associated to the LOR. R and $\mathbf{t}$ are computed as:

$$\mathbf{R} = \begin{bmatrix} \mathbf{r}_1^T \\ \mathbf{r}_2^T \\ \mathbf{r}_3^T \end{bmatrix}, \qquad \mathbf{t} = -\mathbf{R}\mathbf{p}_0 \quad (3)$$

where $\mathbf{p}_m$ is a point belonging to the LOR, that will be the origin of coordinates of z'. A geometric diagram of the vectors involved in the change of reference system is illustrated in Fig. 1.

The derived transformation is used repeatedly on each voxel center coordinates whose contribution to the projection is above a given threshold.

B. Search Algorithm

Given an ideal LOR, derived as the pair of most probable positions of photons-crystals interaction for an acquired event, the search algorithm provides the coordinates of all image voxels whose position has a probability of emission of the detected photons higher than a constant cut-off value. This a posteriori probability is evaluated by means of a projector function $p(i, j, D)$, where i is the voxel center, j is the LOR and D is additional useful data as crystals efficiency, skewness etc.

A schematic diagram of the method is illustrated in Fig. 2. The first intersection between the LOR and the field of view (FOV) is computed geometrically and produces the voxel of entrance. Starting from this voxel, a recursive loop looks for its neighbors and processes all the elements that produce a kernel output above the cut-off threshold. Given that a single voxel can be neighbor of multiple voxels, a specific mechanism is needed in order to prevent voxel reprocessing. A voxel marks memory has been initially developed, but it showed to be infeasible for massive parallelization on graphics procecessors (GPU), given the low memory resources available per single GPU core. Therefore a specific FOV sampling system has been developed. The new voxel search algorithm

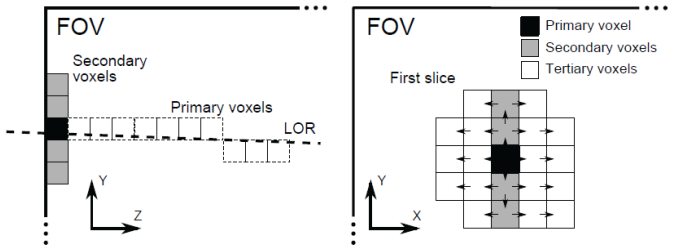

Fig. 3: Geometric diagrams of the ROR propagation technique. In the figure the primary direction is along axis Z, the secondary along axis Y and the tertiary along axis X.

is based on a hierarchical approach, in which each identified voxel corresponds to a node of a tree. Nodes are divided in 3 partitions (primary, secondary, tertiary) and processed sequentially. Primary voxels are all intersected by the LOR, each with a different coordinate in the axis Z, secondary voxels are obtained by propagating primary ones in the Y direction; tertiary voxels are obtained by straight propagation of secondary ones in the X direction Fig. 3. Propagation stops when the cut-off condition on the current voxel is met. With this scheme all the voxels of the ROR that are within the cut-off boundary can be obtained by keeping in memory only the coordinates of the current voxel and those of its parents.

It is important to note that the search algorithm involves only sum operations and no global memory access in write mode, which is the necessary condition for its implementation on graphical processing units.

C. Projector Model

Three different projector models have been used in order to assess quality and speed of the proposed reconstruction method. The simplest one is the constant Gaussian model, in which all TORs are modelled as tubes with symmetric Gaussian section with a fixed standard deviation derived from the mean point spread function of the system. We also used circular and elliptical (2D) Gaussian kernels where the standard deviations are calculated on-line per LOR based on GATE simulations [4]. Image reconstructions have been performed on simulated Derenzo-type and NEMA quality phantoms in order to assess the contrast recovery coefficient (CRC) and noise capabilities of the proposed method (Fig. 4 and 5).

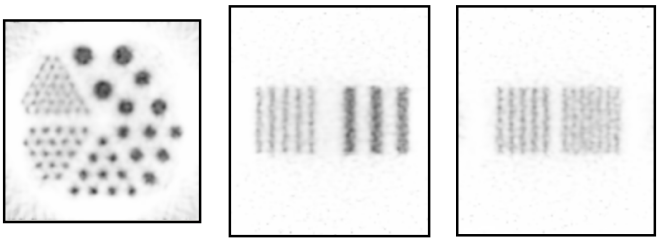

Fig. 4: Axial, coronal and sagittal views of a Derenzo-type phantom reconstructed with the elliptical Gaussian model.

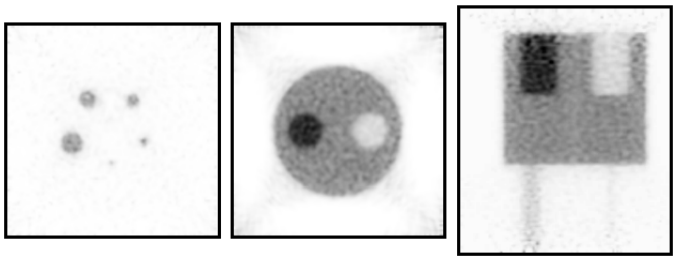

Fig. 5: Transaxial and axial slices of the NEMA quality phantom reconstructed with the elliptical Gaussian kernel.

D. Parallelization

In order to parallelize the OSEM reconstruction algorithm we subdivide the LM dataset into several partitions, each processed by a different processing unit. Separated results are then joined and split again after each iteration. This kind of parallelization pattern, commonly referenced as reduction pattern, provides negligible computation overhead, although is not feasible for graphics processing units. A finer parallelization pattern is under testing for GPU processing. In the GPU parallelization is achieved at LOR level by means of atomic operations on a shared memory space.

III. Results and conclusion

Experiments have been conducted in order to assess the trade-off between reconstruction speed and image quality. Two cut-off thresholds have been used in the kernel sampling algorithm while reconstructing GATE simulated NEMA phantoms [4] within a small animal rPET scanner [6]. Results in Fig. 6 show the recovery coefficient vs. noise/signal ratio trade-off when using 1% and 10% as cut-off conditions, where the kernel maximum value at the center is 1. In Fig. 7 reconstruction speeds are plotted for different kernel models and with the two cut-off condtions above. The cut-off value has a clear impact in the quality of results. As it is appreciable in Fig. 7, the 2D Gaussian model reconstructs in almost half the speed if the cut-off is changed from 10% the Gaussian peak to 1%. The parallelization overhead has been measured for multicore CPUs by executing the algorithm with 8 running cores at 2.1 GHz Fig. 7. A preliminar GPU-based algorithm implementation has been also shown to be possible. The acceleration obtained with a commodity graphics adapter (nVIDIA GeForce GTX 480) with 480 processing cores is about 50x with respect to a CPU core.

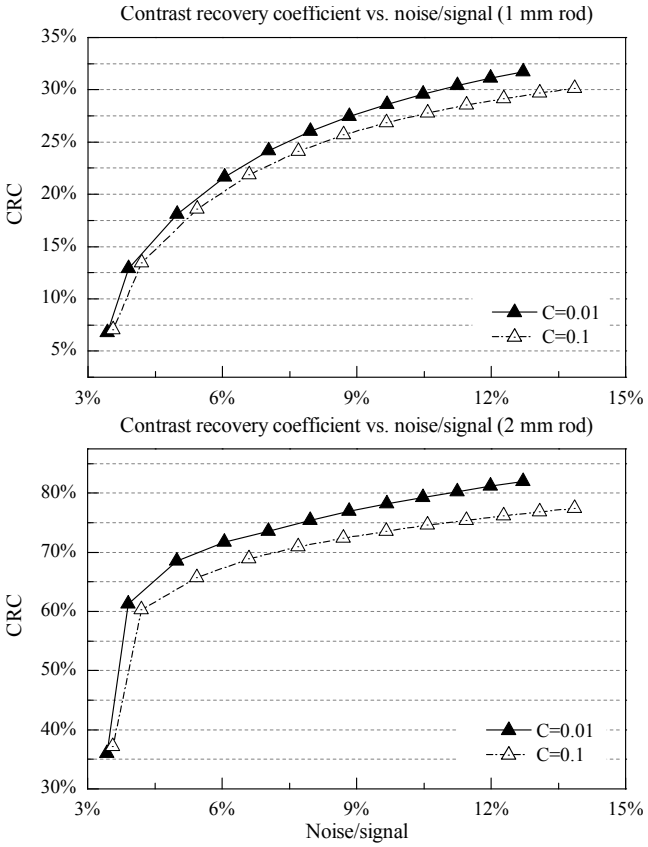

Fig. 6: Recovery coefficient vs noise-signal measured in the 1 mm and 2 mm diameter rod of the quality phantom for LM-OSEM reconstructions using the 2D Gaussian kernel described in [4]. Cut-off values are 1% and 10% of the kernel maximum.

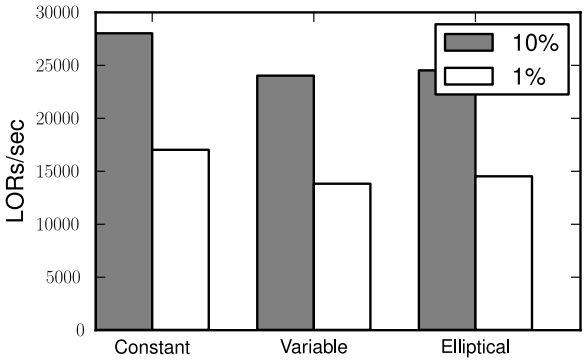

Fig. 7: Reconstruction speed comparison. Speeds for different kernel models and cut-off thresholds are compared when executed in a machine with 8 cores running at 2.1 GHz.

References

[1] R. Siddon, *Med. Phys.*, vol. 12, p. 252, 1985.

[2] G. Pratx *et al.*, *IEEE Trans. Med. Im.*, vol. 28, no. 3, pp. 435–445, 2009.

[3] G. Pratx *et al.*, *IEEE Trans. Nucl. Sci.*, no. 1, pp. 105–109, 2011.

[4] J. E. Ortuño *et al.*, "Projector model for efficient list-mode reconstruction in pet scanners with parallel planar detectors," in *IEEE NSS/MIC Conference Record*, 2011, to be published.

[5] A. Reader *et al.*, *IEEE Trans. Nucl. Sci.*, vol. 49, no. 3, pp. 693–699, 2002.

[6] J. Vaquero *et al.*, in *IEEE NSS/MIC Conference Record*, vol. 5, 2005, pp. 2885–2889.