A System for the 3D Reconstruction of Retracted-Septa PET Data Using the EM Algorithm

Calvin A. Johnson[1], Yuchen Yan[2], Richard E. Carson[2], Robert L. Martino[1], and Margaret E. Daube-Witherspoon[2]

1- Division of Computer Research and Technology
2- PET Department, Clinical Center
National Institutes of Health
Bethesda, MD 20892

Abstract

We have implemented the EM reconstruction algorithm for volume acquisition from current generation retracted-septa PET scanners. Although the software was designed for a GE Advance scanner, it is easily adaptable to other 3D scanners. The reconstruction software was written for an Intel iPSC/860 parallel computer with 128 compute nodes. Running on 32 processors, the algorithm requires approximately 55 minutes per iteration to reconstruct a 128× 128×35 image. No projection data compression schemes or other approximations were used in the implementation. Extensive use of EM system matrix (C_{ij}) symmetries (including the 8-fold in-plane symmetries, 2-fold axial symmetries, and axial parallel line redundancies) reduces the storage cost by a factor of 188. The parallel algorithm operates on distributed projection data which are decomposed by base-symmetry angles. Symmetry operators copy and index the C_{ij} chord to the form required for the particular symmetry. The use of asynchronous reads, lookup tables, and optimized image indexing improves computational performance.

I. INTRODUCTION

The expectation maximization (EM) method for reconstructing PET data is an iterative solution based on a maximum likelihood criterion. The EM algorithm is known in principle to yield more accurate reconstructions than filtered backprojection [1]. EM can incorporate the resolution limitations of the detection system into the reconstruction model and thereby remove the partial volume effect inherent in filtered backprojection [2]. New generation PET scanners allow for the retraction of the lead septa shields that prevent coincidence events from being detected outside the axial plane of emission. Retracting the septa increases the angle over which coincidence events can be accepted and consequently improves the scanner's sensitivity [3]. However, retracted-septa reconstructions are not separable into multiple 2D reconstructions. Consequently, the size of the reconstruction problem grows geometrically when using the EM algorithm for these wider acceptance angles. In a 3D EM reconstruction for a typical scanner geometry, the number of projection elements increases by an order of magnitude, and the size of the system matrix C_{ij}, which is used throughout the reconstruction, can grow by four orders of magnitude or more from that of the 2D problem.

Kaufman [4,5] has investigated redundancies in the system matrix and has proposed solutions to the EM problem that take advantage of a vector computer architecture. Chen et. al. have proposed solutions to the EM problem on a message passing parallel computer. Their solutions incorporate Kaufman's in-plane symmetries as well as 2-fold axial symmetries [6,7].

In this work, we present a system for EM reconstruction from volume acquisition on current generation retracted-septa PET scanners. The reconstruction software was written for the Intel iPSC/860 parallel computer with 128 processor elements (PEs). Although the software was designed for a GE Advance scanner, it is easily adaptable to other 3D scanners. No projection data compression schemes or other approximations (as for example in [8]) were used in the implementation, since this work is intended as a "baseline" for future studies. Our 3D reconstruction system includes a spatially-variant resolution model but does not currently include corrections for attenuation, normalization, randoms, or scatter [9]. We demonstrate the feasibility of using the EM algorithm to reconstruct simulated and actual volume PET data and show initial results from a comparison of the convergence properties of the 3D algorithm to that of 2D.

II. METHODS

The iterative solution of the EM algorithm is [1,10]

$$\lambda_j^{(n+1)} = \frac{\lambda_j^{(n)}}{\sum_{i \in J_j} C_{i,j}} \sum_{i \in J_j} \frac{C_{i,j} \cdot Y_i}{\sum_{k \in I_i} C_{i,k} \cdot \lambda_k^{(n)}} \qquad (1)$$

where Y_i is the ith projection of coincidence events, $\lambda_j^{(n)}$ is the jth element of the reconstructed volume at the nth iteration, and $C_{i,j}$ is the probability of an event in voxel j being detected by detector pair i (the i,j element of system matrix **C**). The parallel implementation decomposes the problem by partitioning the projection space among the PEs. The projections **Y** and the system matrix **C** are distributed across the PEs. As such, this implementation is similar to the "partition by tube"

(i.e., partition in projection space) approach of Chen et. al. [6,7] The image objects are replicated on all PEs.

A. Problem Size

Computationally, the primary difference between 2D and 3D EM reconstructions is the size of the problem. The incorporation of out-of-plane information imposes severe requirements on storing the C_{ij} and on subsequently performing forward- and back-projections with the C_{ij}. Table 1 lists the characteristics of the GE Advance scanner that relate to problem size. Table 2 lists the object size-dependent parameters related to the C_{ij} matrix. Although the full size of the sparse C_{ij} is listed, in practice only the nonzero elements of C_{ij} are stored in "chords" as shall be described in Section B. The use of symmetry-related redundancies in C_{ij} reduces the storage requirement (but not the computational requirement) by a factor of 188 as shall be described in Section C. The actual cost of storing only the base-symmetry chords is listed in the bottom row of Table 2. Because of the extremely heavy computational and storage costs associated with reconstructing a full-FOV object, we have not yet attempted such a large reconstruction.

Table 1: GE Advance scanner characteristics [11].

ring diameter	930 mm
detectors per ring	672
transverse field of view	550 mm
axial field of view	153 mm
number of rings	18
maximum axial separation	11 rings
transaxial angles per sinogram	336
sinograms (without collapsing axial angles)	282

Table 2: C_{ij} parameters which are object-size specific.

characteristic	full-FOV object	brain-sized object
object size, mm	550×550×148.75	256×256×148.75
object size, voxels	256×256×35	128×128×35
voxel size, mm	2.15×2.15×4.25	2×2×4.25
rays per angle	283	119
projections per sinogram	95,088	39,984
total projections	$26.8 \cdot 10^6$	$11.275 \cdot 10^6$
full size of sparse C_{ij},	$6.151 \cdot 10^{13}$ elem.	$6.425 \cdot 10^{12}$ elem.
average size of chord (see Section 4)	12,600 elements 25,200 bytes	3500 elements 7000 bytes
size of all C_{ij} chords without symmetries	675.3 Gbytes	78.9 Gbytes
actual storage requirement	3.592 Gbytes	434 Mbytes

B. System Matrix Generation and Storage

Only the nonzero elements of C_{ij} are stored in "chords", each of which stores the C_{ij} for a particular detector pair i over all contributing voxels j. The chord is stored as a three-dimensional array, a long and narrow box whose primary index is along the direction of greatest traversal (x or y) with respect to the length of the chord box. An example of a chord is illustrated in Fig. 1.

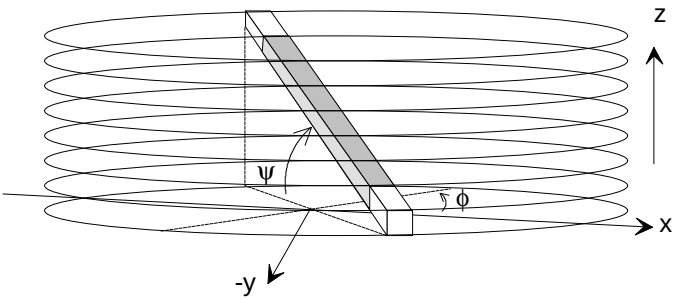

Figure 1: For a particular detector pair i, the nonzero elements of C_{ij} are stored as a long, narrow chord with a leading index along the primary direction. This diagram depicts a nearly central-ray tube at azimuthal angle f and axial lift angle y with maximum traversal along the y direction. The darkened portion of the chord represents where the object contributes; the white portion of the chord is outside the field of view.

The chord model explicitly includes the spatially-variant detector resolution in the chord density calculation. The secondary widths of the chord (y or x and z) are just large enough to contain the point spread to 5% of maximum. Because the point spread is larger at higher transaxial radii, reconstructions of full-FOV objects require considerably larger chords. Specifically, the average chord size for a full-FOV object is estimated to be 200×9×7, while that for a brain-sized object is 100×7×5 (assuming the object dimensions listed in Table 2).

A ray-based approach [10] was selected to compute the chord values. The chord is partitioned into an array of lines that cover the entire spread, i.e., the entire cross-sectional area of the chord. The distance between these lines is 1/7 the voxel size in the secondary transverse and axial directions, respectively. Thus a brain-sized object reconstruction, for which the chord cross section is 7×5 voxels, requires a total of 49·35=1715 lines per chord. The sum of the lengths (weighted by the detector point spread function) of all lines through voxel j determines the value of C_{ij}. Only the base-symmetry chords are calculated and stored; the others are generated through symmetry operators (Section C). Due to the large numbers of lines, computing all chords requires nearly 2 hours on 32 PEs for brain-sized objects. The chords are computed only once and are stored for subsequent use by EM reconstructions.

C. Symmetry Reductions

Extensive use of redundancies present in the C_{ij} matrix makes its storage possible. In-plane symmetries [4,5] reflect a chord with respect to $\phi=45^o$ within its sinogram. The original chord and its reflection are rotated by $\phi=90^o$, 180^o, and 270^o within the sinogram (Fig. 2a). These in-plane operators provide a factor of 8 reduction. The axial symmetry reflects a chord with respect to the plane z=0 [6,7]. For $\theta\neq0$ sinograms, this symmetry operator provides a factor of 2 reduction, but for $\theta=0$, there is no reduction (Fig. 2b). Parallel axial chord symmetries are justified under the assumption that the axial point spread function is invariant to axial position (Fig. 2c). This operator provides a reduction factor of (N_R-s), where N_R is the number of detector rings, and s is the ring separation of rays in the sinogram. The average reduction over all chords from symmetry operators on the GE Advance is 188. The base-symmetry chords are thus defined by $\phi<45^o$, $\psi\geq0$, and include the top ring of the scanner.

The parallel implementation of the 3D EM reconstruction partitions the projection space according to base-symmetry angle. PEs are each assigned a set of base symmetries, which they read from separate files. Chords are grouped into files of similar base-symmetry angle. All possible symmetry-related chords are constructed from the base-symmetry chord and used for forward- and back-projection before the PE reads a new chord (ray index) from file. Constructing the symmetry-related chords from the base chord generally requires re-ordering the chord data. The number of base-symmetry angles is large enough to ensure that efficiency of the parallel implementation is not compromised.

D. Computational Performance Enhancements

A number of features have been written into the software to improve computational performance. These features include asynchronous reads of the chords to eliminate I/O waits, lookup tables to check image bounds while traversing chords during forward- and back-projections, an image indexing scheme designed to minimize indexing arithmetic, and precalculation of trigonometric functions. These and other measures have reduced the computation time from 113 to 52 minutes per iteration on 32 PEs.

III. FEASIBILITY STUDIES

To demonstrate the feasibility of the 3D EM reconstruction method, we have performed reconstructions on data simulated from a CT image and on measured PET data of a uniform cylinder and point sources.

In the simulation study, PET data were simulated from a CT image of a dog thorax using the GE Advance scanner geometry. The image was forward projected, with 6mm blur-ring but without additive noise, to create the simulated sinogram data. Four voxels from the reconstructed image were chosen for a convergence study (Fig. 3). The number of iterations taken for the 3D reconstruction was limited due to the amount of time required. The 2D reconstruction at 1000 iterations appears to be nearly convergent. The results of the convergence study (Fig. 4) indicate that the additional data acquired in 3D mode may not significantly alter the convergence rate of the EM algorithm over 2D mode.

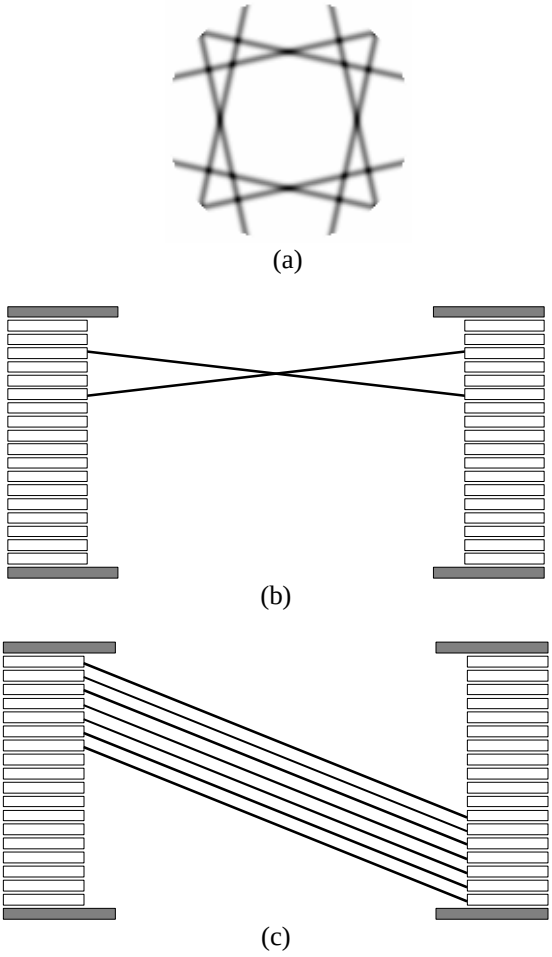

Figure 2: Schematic representations of the symmetry operators. (a) In-plane symmetries, shown as an image of a cross section (top axial slice) of the superposition of the chords described by {$\theta=0$, $\phi=12.6^o$, ray index = 16}. (b) Two-fold axial symmetry. (c) Parallel axial chord redundancies, shown for a maximum ring separation of 11.

In the uniform cylinder study, we performed a reconstruction on a 16 cm diameter cylinder filled with a solution of ^{18}F (130M counts). The reconstruction shown in Fig. 5 is of the raw scanner data without correction for the physical effects. The voxel intensities of the four points within the cylinder shown in Fig. 5 are plotted as a function of iteration in Fig. 6. Although a spatially-variant resolution model can be included in the system matrix, for these reconstructions a FWHM of 6 mm in all directions was used throughout.

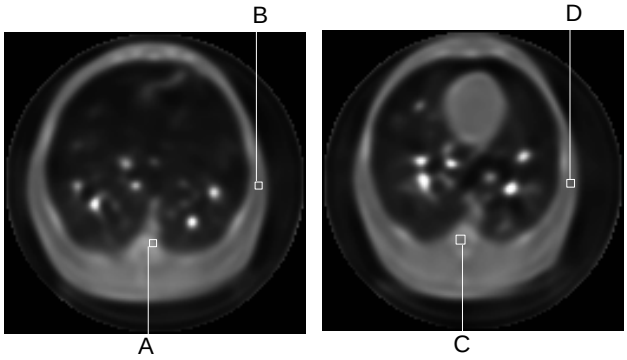

Figure 3: Two slices from a reconstructed image (194 iterations), where PET data were simulated from CT images. Four voxels chosen for the convergence study are highlighted.

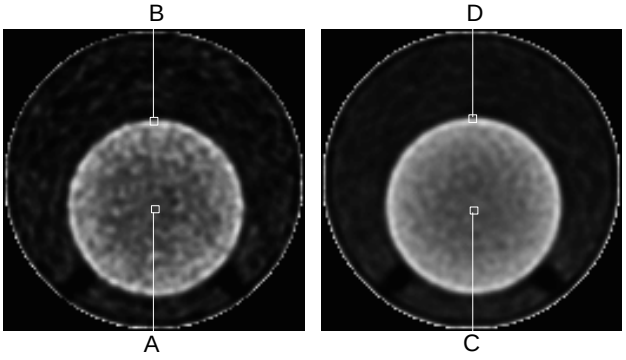

Figure 5: Images of the reconstructed uniform cylinder phantom, uncorrected for the physical effects of attenuation, normalization, randoms, and scatter after 140 iterations. Left- edge slice; right- central slice.

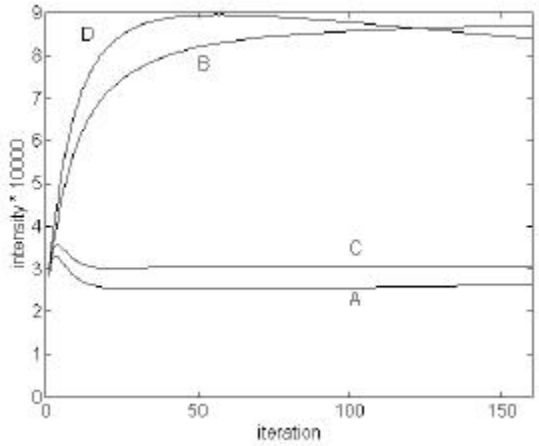

Figure 6: The four voxels highlighted in Fig. 5 are plotted against iteration.

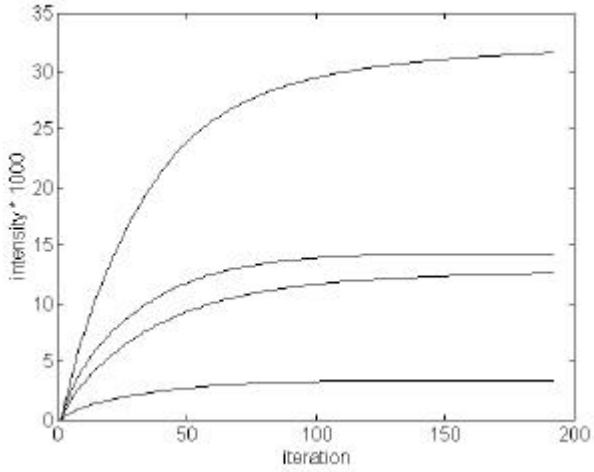

Figure 7: Intensities of the center of four (out of five) reconstructed point sources as a function of iteration.

Figure 4: Voxel intensity as a function of iteration for the four voxels highlighted in Fig. 3. (top)- Retracted-septa (3D) convergence, vs. (bottom)- septa-in (2D) convergence.

In the point source study, data taken from a scan of five point sources (2mm×2mm×1mm) located 5 cm below the center of the gantry were reconstructed. The points were spaced 3 cm apart axially. As shown in Fig. 7, the points appear not to have converged after nearly 200 iterations. This slow convergence may be due in part to the use of a spatially invariant 6mm FWHM resolution function, which excessively deblurred the points.

IV. CONCLUSIONS

The current studies demonstrate that 3D EM reconstruction is computationally feasible. The practicality of computing these reconstructions remains an open question. We believe that the brain-sized problem is the largest that can be attempted on the Intel iPSC/860. The time per iteration, memory demands, and storage requirements of the full-FOV problem are much greater than that of the brain-sized problem. We intend to study the larger problem in the future. At least an order of magnitude in computational performance improvement will be required to make such a system feasible in the clinic, thereby requiring advances in parallel computer technology. We may also need to consider convergence acceleration and optimization techniques such as those offered by Kaufman [4,5].

Should the desired computational performance improvement be achieved, the problem will likely become I/O bound. Reconstructing a brain-sized object requires reading 430 MB of chord data per iteration, while reconstructing a body-sized object is estimated to require 3.6 Gbytes per iteration. To avoid becoming I/O limited, the C_{ij} matrix would need to be stored in memory. This would be possible given enough PEs and sufficient memory per PE. Predictive compression techniques may be required to reduce the dynamic range of the C_{ij} data in order to relax memory demands.

In the implementation of the 3D EM algorithm, no compression schemes or approximations were made in order to have a "baseline" for future studies. Further examination of the convergence rates of the 3D algorithm compared with those of the 2D algorithm is planned. Our future studies will investigate image quality, resolution recovery, quantitative accuracy, and variability of the 3D algorithm compared with the 2D algorithm, as well as with other 3D reconstruction methods.

V. REFERENCES

[1] Lange, K., and Carson, R., "EM Reconstruction Algorithms for Emission and Transmission Tomography," *J. Comp. Assist. Tomogr.*, vol. 8, pp. 306-316, 1984.

[2] Carson, R.E., Yan, Y., Chodkowski, B., Yap, T.K. and Daube-Witherspoon, M.E., "Precision and Accuracy of Regional Radioactivity Quantitation Using the Maximum Likelihood EM Reconstruction Algorithm," *IEEE Trans. Med. Imag.*, vol. 13, pp. 526-537, 1994.

[3] Cherry, S.R., Dahlbom, M., and Hoffman, E., "3D PET using a Conventional Multislice Tomograph without Septa," *J. Comp. Assist. Tomogr.*, vol. 15**,** pp. 655-668, 1991.

[4] Kaufman, L., "Implementing and Accelerating the EM Algorithm for Positron Emission Tomography," *IEEE Trans. Med. Imag.*, vol. MI-6, pp. 37-51, 1987.

[5] Kaufman, L., "Solving Emission Tomography Problems on Vector Machines," *Ann. Oper. Res.*, vol. 22, pp. 325-353, 1990.

[6] Chen, C.M., Lee, S.Y., and Cho, Z.H., "Parallelization of the EM Algorithm for 3D PET Image Reconstruction," *IEEE Trans. Med. Imag.*, vol. 10, pp. 513-522, 1991.

[7] Chen, C.M., and Lee, S.Y., "On Parallelizing the EM Algorithm for PET Image Reconstruction," *IEEE Trans. Parallel. Dist. Syst.*, vol. 5, pp. 860-873, 1994.

[8] Guerrero, T.M., Levin, C.S., Huang, S.C., Dahlbom, M., Cherry, S.R., and Hoffman, E.J., "Implementation and Evaluation of the EM Algorithm for 3D PET," *J. Nucl. Med.*, vol. 35, pp. 1871, 1994 (abstract).

[9] Daube-Witherspoon, M.E., Carson, R.E., Yan, Y., and Yap, T.K., "Scatter Correction in Maximum Likelihood Reconstruction of PET Data," in *Proceedings, IEEE Medical Imaging Conf.*, pp. 945-947, 1992.

[10] Shepp, L.A. and Vardi, Y., "Maximum Likelihood Reconstruction for Emission Tomography," *IEEE Trans. Med. Imag.*, vol MI-1, pp. 113-122, 1982.

[11] Degrado, T.R., Turkington, T.G., Williams, J.J., Stearns, C.W., Hoffman, J.M., and Coleman, R.E. "Performance Characteristics of Whole-Body Scanner," *J. Nucl. Med.*, vol. 35, pp. 1398-1406, 1994.