Optimal speed gain for CUDA implementation of iterative emission tomography (SPECT) image reconstruction

Jakub Pietrzak1,2 and Krzysztof Kacperski3

1Department of Medical Physics, Maria Skłodowska-Curie Memorial Cancer Center - Institute of Oncology, Warsaw, Poland
2Institute of Experimental Physics, Faculty of Physics, University of Warsaw

Abstract

GPU implementation can greatly accelerate iterative techniques of 3D reconstruction in nuclear medicine imaging. To obtain high quality images in Single Photon Emission Computed Tomography (SPECT) within reduced scanning times, high sensitivity collimators need to be used and their response function modeled in the reconstruction. This is in general very computationally intensive and unfeasible with conventional PCs and algorithm implementations. Our software is able to perform the reconstruction of patient data within clinically acceptable times (18 s vs 17 mm on CPU) using relatively low-cost and widely available hardware.

Single photon emission computed tomography (SPECT)

Principle of radionuclide imaging:

- Administer the patient a drug of interest, labelled with a short-lived gamma-emitting radioisotope
- Measure a set of parallel projections of the gamma emission field with a detector fitted with a parallel hole collimator
- Reconstruct the radionuclide distribution to obtain clinically relevant information, e.g. sites of malignant tumors, blood flow in the heart, etc.

Gamma camera

The MLEM Algorithm

To reconstruct the 3D activity distribution in the patient from the set of planar projections measured typically on a flat panel scintillating detector, rotating around the long axis of the patient, one needs to apply a reconstruction algorithm.

Why statistical reconstruction algorithms?

- Better modeling of the physics of data acquisition process - collected data are treated as set of statistical samples and the reconstructed image as the estimator of underlying activity distribution.
- Modeling of the physics of scanning system: collimator and detector point response function, attenuation, scatter.
- Possibility to optimize the collimator and scanning geometry to reduce scanning time and/or patient dose.

Projections: $g = Mf$

$m$: measured projections,
$f$: image activity distribution
$M$: system matrix.

Maximum Likelihood expectation maximization (MLEM algorithm)

The most popular statistical reconstruction method is the Maximum Likelihood Expectation Maximization algorithm, in which the pixels of the image, $f_i$, are updated iteratively according to the formula:

$$f_{i+1} = \sum_{k} M_{ik} \left[ \frac{g_k}{\sum_j M_{jk}} - \frac{M_{ik}}{\sum_k M_{ik}} \right]$$

Why do we use CUDA?

- Projection operator is simple to parallelise, values of projection pixels can be computed independently at the same time
- Fast floating point mathematical and logical, thus parallelized. edge detection processes.
- Linear Texture Interpolation (used in rotations)
- Optimal memory size – reconstructed image is a 4 byte pixel matrix of the size 128x, so it is possible to keep all the data on device memory and reduce data transfer process to just sending a set of projections to the device and to get reconstructed data back.

Implementation

Main programming objectives:

- Keep all cores busy
- Minimize host-device/memory data transfer
- Use device features for caching and optimizations (shared memory, cudaKernels, fast math functions, texture linear interpolation)

Results

Reconstruction times on GPU and CPU have been compared for 100 iterations of OSEM algorithm (20 subsets) with a high sensitivity collimator (sensitivity = 2.2 x standard HR collimator) at different image matrix sizes. The algorithm has been implemented on CUDA GPU and standard CPU (without multithreading and SIMD optimization) for comparison. Tests were performed on GTX 480 and Intel Core i7 930 installed on the same machine.

Conclusions

- GPU implementation reduces the computation time of the statistical image reconstruction in SPECT by a factor of about 50 for typical image sizes.
- The speed-up factor grows with the size of image size; it may approach 200 for large images.
- For GPU implementation reducing the dimensions of image matrix may not be an effective way of decreasing the reconstruction time. Similarly, using non-pixel based image representations, e.g. blocks, may be suboptimal to just applying a finer pixel grid.
- For standard SPECT 128x seems to be the optimal image size, readily suitable to apply the fast FFT based filtering.
- Even higher speed gains could be obtained in CT, where typical image sizes are of the order of 1000x1000.