Proceedings of the 18th World Congress
The International Federation of Automatic Control
Milano (Italy) August 28 - September 2, 2011

GPU-based acceleration of the MLEM algorithm for SPECT parallel imaging with attenuation correction and compensation for detector response

Ákos Szlávecz * Gábor Hesz * Tamás Bükki ** Béla Kári *** Balázs Benyó *

* *Department of Control Engineering and Information Technology, Budapest University of Technology and Economics, H-1117 Budapest, Magyar tudósok körútja 2.*
** *Mediso Ltd., H-1022 Budapest, Alsótörökvész 14., Hungary*
*** *Semmelweis University, Department of Diagnostic Radiology and Oncotherapy, H-1082 Üllöi út 78/a*

Abstract: Parallel projection based Single Photon Emission Computed Tomography (SPECT) is one of the most widely used nuclear imaging technique even nowadays. Serious artefacts are produced in the reconstructed images due to the non-homogeneous attenuation medium and the distance dependent spatial resolution (DDSR) of the parallel imaging. Effective non-uniform attenuation correction and DDSR reduction procedures should be applied in order to improve the SPECT image quality. We have developed a novel parallel reconstruction method using the Maximum Likelihood Expectation Maximization iterative reconstruction algorithm with attenuation correction and compensation for the DDSR effect in the forward projector. In order to compensate the well-known extreme computation intensity of this reconstruction method a parallel version of the algorithm is created where the computation tasks of the algorithm are executed simultaneously on a GPU. By this reduction of the running time this accurate reconstruction algorithm become available for the use in the clinical applications. The algorithm has been verified using simulation studies.

Keywords: Medical Imaging, Nuclear Medicine, SPECT, Photon Attenuation, Attenuation Correction, Collimator Blurring, Distance Dependent Spatial Resolution, DDSR Correction, Image reconstruction, Maximum Likelihood Expectation Maximization, MLEM, High Performance Computing, GPU programming

1. INTRODUCTION

Parallel projection based Single Photon Emission Computed Tomography (SPECT) is a widely used nuclear imaging technique for functional imaging of the human body (Miles N. Wernick (2005)). During SPECT examination a compound labelled with gamma photon emitting radionuclide (an often used radioactive material is the Tc99m) is injected into the human body. The injected radionuclide is concentrated in particular parts of the body that are of medical interest for disease detection. A rotating gamma camera is then obtaining 2D views (projection images) of the 3D distribution of the radionuclide at different angular positions. After data acquisition the 3D distribution of the radioactive substance can be reconstructed using appropriate reconstruction algorithm.

Several physical effects may cause artefacts in reconstructed images. The most important distortion effects in parallel based SPECT imaging are arisen from the distance dependent spatial resolution of the detector and from the non-homogeneous attenuating medium surrounding the imaged object. The distortions caused by these effects are even more serious if the imaged activity is not in the center of the field of view and in the case of 180° data acquisition.

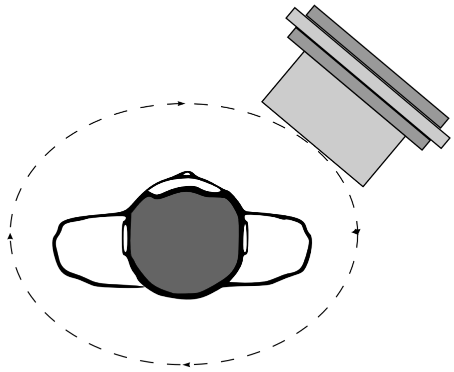


Fig. 1. During SPECT imaging a radionuclide is injected into the body. A gamma camera is rotated around the body and projection images are collected. After data acquisition the 3D distribution of the radioactive substance can be reconstructed using appropriate reconstruction algorithm.

1.1 Effect of gamma photon attenuation

Gamma photons interact with the material while travelling through the volume. As the consequence of the interactions a portion of the events are not detected. The event loss results in attenuated projections. The amount of event loss can be calculated according to the Beer-Lambert formula:

$$I = I_0 \cdot e^{-\int_l \mu(x) \cdot x dx} \tag{1}$$

Where I is the measured number of counts in a detector pixel, I_0 is the emitted number of counts on a voxel and $\mu(x)$ represents the material distribution containing total attenuation coefficients. The total attenuation coefficients represents the attenuation factor cased by all the physical phenomenons such as photoelectric absorption and Compton scattering. The line integral in equation 1 is calculated along the line l in the volume (from the voxel to the given detector pixel).

The activity in an image voxel suffer from different attenuation factors in different directions in the case of non-homogeneous attenuating medium and this effect causes serious distortion in the reconstructed image.

1.2 Effect of distance dependent response of the detector

A gamma camera is consisted of a crystal material that is absorbing the incident gamma photon and produces optical photons. The optical photons are converted to electrical signal by the photomultiplier tubes. A collimator is a block of lead with many small parallel holes. The collimator guarantees that only the perpendicularly incident gamma photons can reach the crystal material, theoretically.

In fact the geometrical properties of the collimator such as the finite hole and septa size, the gamma photon penetration between the parallel holes of the collimator and the crystal material intrinsic resolution results in a distance dependent point spread function (PSF) of the gamma camera (Fig. 2.).

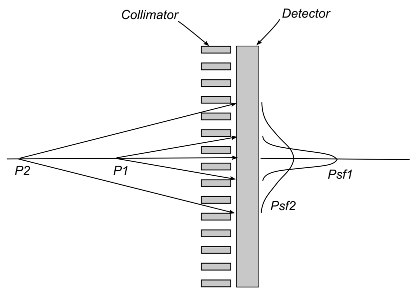


Fig. 2. The gamma camera produce a distance dependent point spread function resulting in a Gaussian blurring.

The distance dependent response function of the detector creates a blurring in the reconstructed image.

1.3 Goal of the research

The effects mentioned above result in distortions in the reconstructed images. Therefore, effective non-uniform attenuation correction (Tsui et al. (1989)) and DDSR reduction procedures (Tsui et al. (1994)) should be applied in order to improve the SPECT image quality (Zeng et al. (1991)).

We used the Maximum Likelihood Expectation Maximization (MLEM) algorithm (Shepp and Vardi (1982) and Lange and Carson (1984)) with ordered subsets (OSEM) of projection data (Hudson and Larkin (1994)) in order to reconstruct the 3D distribution of the radionuclide. We incorporated correction method for photon attenuation and the DDSR effect in the forward projection of the MLEM algorithm. However, the appropriate modified MLEM algorithm has become extremely computational intensive. The main goal of the work presented in this paper was to create a GPU based parallel version of the MLEM algorithm with the above mentioned extensions, with attenuation correction and correction for the DDSR effect. The final goal of the execution time reduction is to make possible the clinical use of the resulted algorithm.

2. METHODS

2.1 Maximum Likelihood Expectation Maximization

The goal of the reconstruction algorithm used in SPECT is the reconstruction of the distribution of radioactive tracer in the human body based on the projection images. The most widely used iterative reconstruction methods for emission tomography is the Maximum Likelihood (ML) Reconstruction using the Expectation Maximization (EM) algorithm.

In the MLEM algorithm the Poisson probabilistic phenomena of the radioactive decay is taken into account. The algorithm calculates the activity distribution which has produced the measured values with the highest likelihood according to the following equation:

$$f_j^{k+1} = f_j^k \cdot \frac{1}{\sum_{i=1}^{I} a_{ij}} \cdot \sum_{i=1}^{I} \frac{b_i}{\sum_{n=1}^{J} f_n^k \cdot a_{ni}} \cdot a_{ij} \tag{2}$$

There are J different image voxels and I detector locations. a_{ij} is the probability that a photon emitted in voxel j is detected in detector location i. In parallel SPECT imaging a_{ij} is proportional to the crossing length between the image pixel and the the line toward the detector pixel location perpendicularly.

The MLEM algorithm can be seen as a set of successive projections and backprojections where the ML-EM estimate for the $(k+1)$ iteration is defined by the equation 2. The denominator part of the sum in equation 2 is the forward projection while the sum over I is representing the backprojection process.

The main advantage of using an iterative reconstruction algorithm such as the MLEM is that physical effects and the geometrical properties of the SPECT camera can be incorporated into the forward projection. The attenuation factor can be incorporated into the model by modifying the a_{ij} values with the calculated line integrals according to equation 1. The detector response can be incorporated into the forward projection in the way that by calculating the added intensity of a voxel to a detector pixel an additional Gaussian blurring have to be performed.

A widely used convergence acceleration technique of the MLEM algorithm is dividing the projection images into subsets and applying to them the MLEM process separately (Hudson and Larkin (1994)). This is called as the Ordered Subset Expectation Maximization (OSEM).

2.2 Determination of the attenuation map

The inclusion of attenuation in the transfer matrix requires apriori knowledge of the attenuation map. In clinical and research applications adequate methods must be performed to generate the attenuation map. This can be done for example through transmission scanning, segmented MRI data, or appropriately scaled CT scans acquired either independently on separate or simultaneously on multimodality imaging systems (Zaidi and Hasegawa (2003)).

In this work we have used a CT scan to determine the attenuation map used by the attenuation correction. The attenuation map has been derived by segmenting the CT scan after biomaterials which HU (Hounsfield Unit) values is known (See Table 1). In the segments we have used the total attenuation coefficients from the xcom database created by National Institute of Standards and Technology (NIST (2010)).

Table 1. Cutting edge values for segmenting the materials in the human body

Material	HU
Air	-1000
Adipose	-120
Water	0
Muscle	+40
Bone	+400

2.3 Calibration procedure for DDSR correction

In order to use DDSR compensation in the MLEM algorithm the distance dependent point spread function of the detector has to be determined. For this reason we have defined a calibration procedure as follows.

A small capillary is placed in the front of the detector in different, increasing distances and planar images are collected (Fig. 3.). On each planar image the sigma parameter of a fitted Gaussian function is determined so that for all line profile of the capillaries the fitting is applied and the average value is calculated in order to gain the σ parameter.

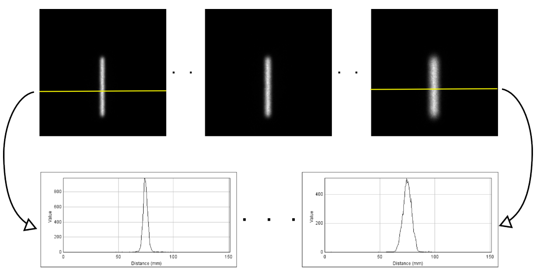


Fig. 3. Calibration procedure for DDSR compensation.

The σ parameter of the Gaussian function is the linear function of the distance from the detector surface. A linear regression is used to determine the parameters of the line (Fig. 4.).

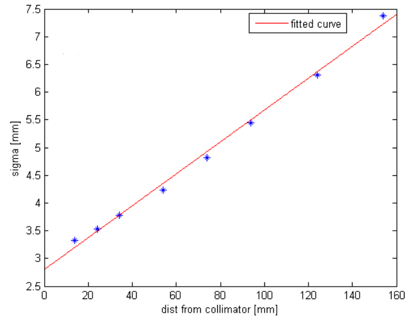


Fig. 4. Linear fitting on σ.

The result of the calibration procedure is the $psfa$ and $psfb$ parameters. Then the point spread function can be calculated during reconstruction according to the following formulas:

$$y = a \cdot e^{-2 \cdot \frac{x^2}{\sigma^2}} \tag{3}$$

$$\sigma = psfa + psfb \cdot d \tag{4}$$

where d is the distance from the detector.

2.4 High performance computing using GPU

We used the OSEM algorithm with the correction methods mentioned above. However, the appropriate modified OSEM algorithm has become computational extremely intensive. One iteration of the CPU based implementation was running for approximately 30 minutes on a Core2Duo based computer. Consequently, the application of high performance computing (HPC) technique was necessary.

In the recent years GPUs have been successfully applied in a wide range of general computing applications due to the increasing level of programmability and flexibility. The use of GPUs for general-purpose computing has become widely known as General Purpose GPU. GPGPU has enabled the acceleration of computations in domains such as signal processing, database processing, computer vision, image processing, and also medical imaging (GPGPU (2011)).

Recent GPUs (Fig. 5. shows an nVidia GPU architecture) are generally programmable, have full IEEE single-precision floating point arithmetic (since 2008 also double precision), their memory capacity is up to to 1.5 gigabytes, and due to the PCI-Express bus also have fast data transfer rates from main memory to GPU, where the data transfer may be overlapped with ongoing computations.

There are two available interfaces available for programming GPUs. The first is the Shader Model that is especially designed for graphics applications. This is supported by both of the main manufacturers, NVIDIA and AMD (ATI). On the other hand, the graphics hardware can be threated as general stream processor. To achieve this goal, NVIDIA introduced the CUDA (Compute Unified Device Architecture) and AMD designed the CMT (Close-to-Metal). These interfaces allow users to write high-performance programs for any computational intensive task in the standard C language.

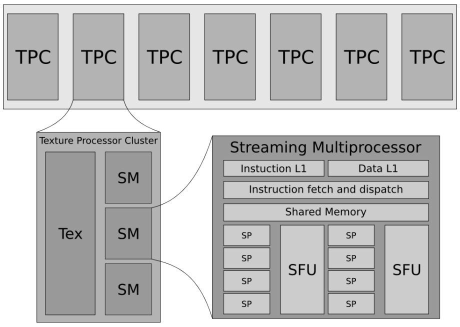


Fig. 5. nVidia GPU architecture.

3. RESULT AND DISCUSSION

3.1 Implementation details

As CUDA is more widely used and therefor better supported we decided to use nVidia GPUs. Therefore the implementation has been carried out in the CUDA programming language.

In a CPU based parallel implementation the forward projection step of the reconstruction algorithm would be calculated in voxel driven way. This means for all voxel the added value to the detector pixels are calculated simultaneously. In this case one thread precesses the calculations of one pixel and at the end of the calculation the value of the detector pixel is incremented. This requires synchronization of data accesses to the detector pixel values in order to provide data consistency.

In order to avoid thread synchronization and provide data consistency the so called "collecting like" approach of the calculation has been developed for our GPU based implementation.

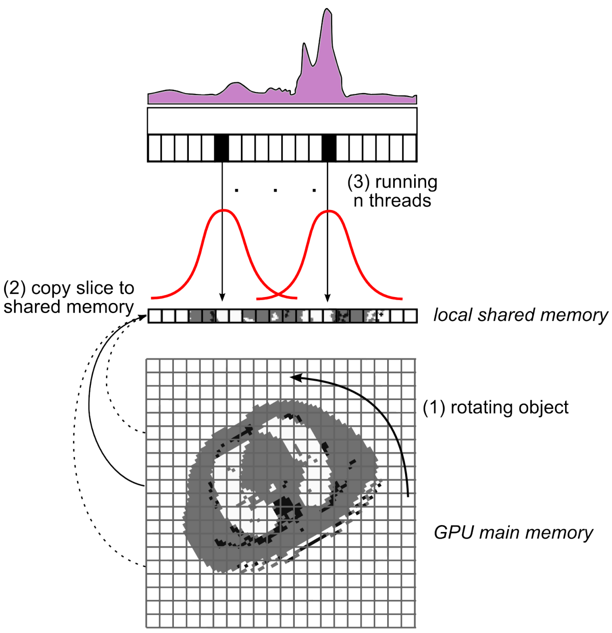


Fig. 6. Forward projection on GPU.

The "collecting like" approach in the case of the forward projection means that one independently running thread collects all the values from the whole image space for one detector pixel (See Fig. 6.).

In GPUs the main source of bottlenecks is the access to memory. In nVidia GPUs there are different types of memories with different access times: registers, shared memory and main memory. However, faster memory types are available in limited quantity. Registers are the fastest memories while the access to the main memory of the GPU is the most time consuming operation. For each multiprocessor 64 kB shared memory is available. The access to the shared memory is approximately 100 times faster than to access to the main memory but only the threads of one multiprocessor can access them. In order to use the fast shared memory the forward projection has been implemented in a special way. Rather than "rotating" the camera by processing detector pixels in the projection images the image volume has been rotated (see Fig. 6.) and the slices of the image has been cached into the shared memory. This solution resulted in significant increase in performance.

Thus the forward projection can be divided into the following steps:

(1) A projection image is selected for processing.
(2) The object is rotated in the appropriated position.
(3) Slice is cached in the shared memory.
(4) Execute the threads, one thread is collecting all the data from the cached slice.
(5) Repeat steps (3)-(4) until all slices are processed.
(6) Repeat steps (1)-(5) until all projection images are processed.

Similarly to the above described forward projection method in the backward projection step also one thread collects data for one image voxel from all detector pixel (see Fig. 7.).

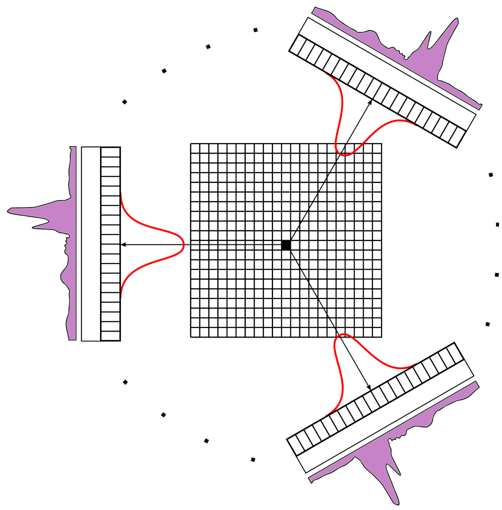


Fig. 7. Backward projection on GPU.

Due to the GPU based parallel implementation the running time of the algorithm has been significantly reduced. In the case of reconstructing 64 number of 128x128 projection images in a 128^3 volume discretization with 3 mm pixel size the running time of 25 iterations of the MLEM algorithm with attenuation correction and compensation for the DDSR effect lasted 205 seconds on an nVidia GTX

480 GPU. This is a high performance gain compared to the CPU based version where the running time of one iteration was approximately 30 minutes (on a Core2Duo based computer).

3.2 Verification

The reconstruction algorithm has been verificated using the GATE simulation toolkit (Jan and et al (2004)).

We have defined a regularly used SPECT system with Low Energy High Resolution (LEHR) collimator. The psf parameters has been determined according to the calibration procedure described above. The resulted psf parameters are:

- psfa = 2.33 mm
- psfb = 0.033

An attenuation map has been created by taking the voxelized phantom of the GATE simulation and using the total attenuation coefficients determined by the National Institute of Standard and Technologies (NIST).

As the first study a ring phantom with non-homogenous attenuating medium has been defined (Fig. 8.).

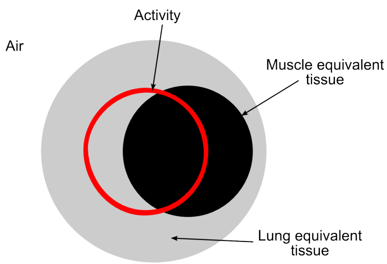


Fig. 8. Ring phantom. Attenuation map and activity.

The GATE simulation was executed and single events were collected. During the simulation the radius of rotation was 281 mm. After running the simulation the Compton scattered events were removed. In this way an "ideal" scatter correction was applied. Then from the single events projection images in 128 views over 360 degree were generated with 128x128 pixels and with a pixel size of 3mm.

Reconstruction was made using the Ordered Subset of Expectation Maximization (OSEM) algorithm and the new GPU based OSEM algorithm with attenuation correction and compensation for the DDSR effect. 8 subsets and 20 iterations was applied during reconstructions. Reconstructed slices are shown on Fig. 9.

With a Cardiac Stress/Rest software package (Mediso (2010)) standard bullseyes were generated. They can be seen on Fig. 10.

On Fig. 9. and Fig. 10. can be seen that the algorithm corrected the effect of photon attenuation while the full ring is recovered. Also the effect of the distance dependent resolution of the detector is compensated while the ring has the same thickness as the reference.

In the second step of the verification process the NCAT torso phantom (Segars et al. (1999)) was used for verification (See Fig. 11.). The radius of rotation was 281 mm and

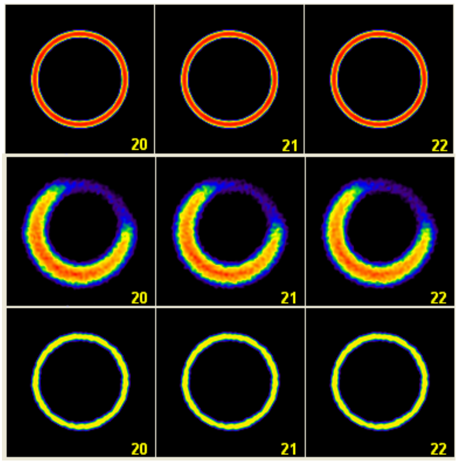
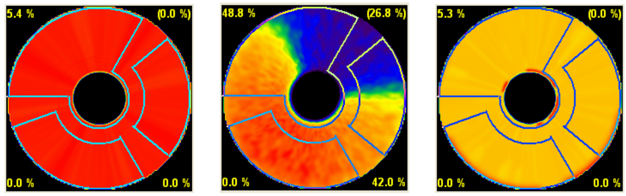


Fig. 9. Ring phantom three adjacent transaxial slices. Top row: reference, middle row: reconstructed with 2D-OSEM, bottom row: reconstructed with GPU based OSEM with correction for attenuation and the DDSR effect.

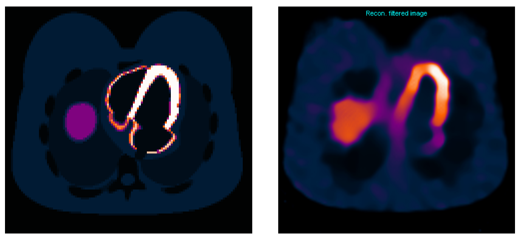


(a) Reference (b) OSEM 2D (c) OSEM with AC and DDSR compensation

Fig. 10. Ring phantom bullseye.

the Compton scattered events were removed in this case as well. Then from the single events projection images in 128 views over 360 degree were generated with 128x128 pixels and with a pixel size of 3mm.

Reconstruction was made using the Ordered Subset of Expectation Maximization (OSEM) algorithm and the new GPU based OSEM algorithm with attenuation correction and compensation for the DDSR effect. 8 subsets and 20 iterations were applied during reconstructions. One slice of the reference and the corresponding reconstructed slice that was reconstructed with the GPU based algorithm is shown on Fig. 11.



(a) Reference (b) OSEM with AC and DDSR compensation

Fig. 11. Simulated NCAT phantom transaxial slice.

The two reconstructions has been compared with the Cardiac Stress/Rest software package. The orientations can be seen on Fig. 12. while the bullseyes are shown in

Fig. 13. A significant gain in image contrast can be seen on the images.

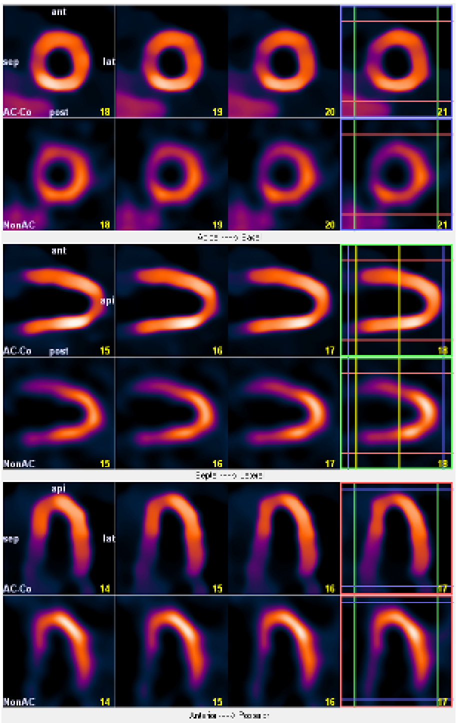


Fig. 12. Simulated NCAT phantom orientations. Top row (marked as AC-Corr): reconstructed with GPU based OSEM. Bottom row (marked as NonAC-Corr): reconstructed with 2D-OSEM.

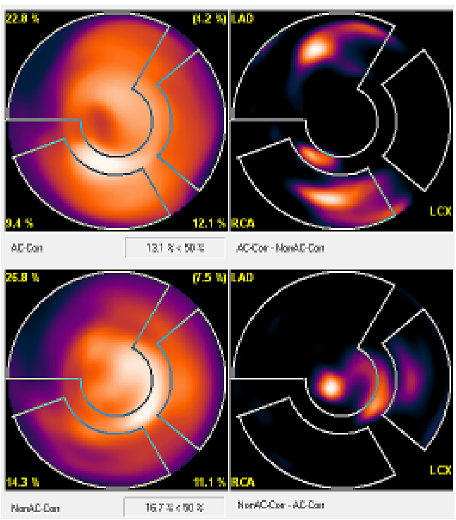


Fig. 13. Simulated NCAT phantom bullseye. Top row (marked as AC-Corr): reconstructed with GPU based OSEM. Bottom row (marked as NonAC-Corr): reconstructed with 2D-OSEM.

4. CONCLUSION

We have developed a GPU based parallel version of the OSEM algorithm with compensation for gamma photon attenuation and the DDSR effect for SPECT parallel imaging. The simultaneous execution of the steps of the reconstruction algorithm successfully compensates the extreme computation intensity of the original algorithm. In this way the running time has been significantly reduced so that the algorithm can now be used in clinical applications. The algorithm has been verified with mathematical phantoms using GATE simulations.

ACKNOWLEDGEMENTS

This work was supported by Mediso Ltd and by the National Office for Research and Technology (NKTH) Grant No. TECH_08_A2-TeraTomo and by the Hungarian National Research Found (OTKA) Grants No. T69055, CK80316 and K82066. This work is connected to the scientific program of the "Development of quality-oriented and harmonized R+D+I strategy and functional model at BME" project, supported by the New Hungary Development Plan (Project ID: TÁMOP-4.2.1/B-09/1/KMR-2010-0002).

REFERENCES

GPGPU (2011). General-purpose computation on graphics processing units. `http://www.gpgpu.org/`.

Hudson, H. and Larkin, R. (1994). Accelerated image reconstruction using ordered subsets of projection data. *Medical Imaging, IEEE Trans. on*, 13(4), 601 –609.

Jan, S. and et al (2004). Gate: a simulation toolkit for pet and spect. *Physics in Medicine and Biology*, 49(19), 4543.

Lange, K. and Carson, R. (1984). EM reconstruction algorithms for emission and transmission tomography. *Journal of Computer Assisted Tomography*, 8(2), 306–316.

Mediso (2010). Interview xp software package. `http://www.mediso.hu/`.

Miles N. Wernick, J.N.A. (2005). *Emission Tomography: The fundamentals of SPECT and PET*. Springer-Verlag London.

NIST (2010). Xcom: Photon cross sections database. `http://www.nist.gov/index.html`.

Segars, W., Lalush, D., and Tsui, B. (1999). A realistic spline-based dynamic heart phantom. *Nuclear Science, IEEE Transactions on*, 46(3), 503 –506.

Shepp, L.A. and Vardi, Y. (1982). Maximum likelihood reconstruction for emission tomography. *Medical Imaging, IEEE Transactions on*, 1(2), 113 –122.

Tsui, B.M.W., Frey, E.C., Zhao, X., Lalush, D.S., Johnston, R.E., and McCartney, W.H. (1994). The importance and implementation of accurate 3d compensation methods for quantitative spect. *Physics in Medicine and Biology*, 39(3), 509.

Tsui, B.M.W., Gullberg, G.T., Edgerton, E.R., Ballard, J.G., Perry, J.R., McCartney, W.H., and Berg, J. (1989). Correction of Nonuniform Attenuation in Cardiac SPECT Imaging. *J Nucl Med*, 30(4), 497–507.

Zaidi, H. and Hasegawa, B. (2003). Determination of the Attenuation Map in Emission Tomography. *J Nucl Med*, 44(2), 291–315.

Zeng, G., Gullberg, G., Tsui, B., and Terry, J. (1991). Three-dimensional iterative reconstruction algorithms with attenuation and geometric point response correction. *Nuclear Science, IEEE Transactions on*, 38(2), 693 –702.