High Quality Cone-beam CT Reconstruction on the GPU

Gábor Jakab[1,2], Attila Rácz[1], Kálmán Nagy[1]

[1] Mediso Medical Equipment Developing and Service Ltd.
Budapest, 1022 Hungary
http://www.mediso.hu, research@mediso.hu

[2] Budapest University of Technology and Economics
Department of Control Engineering and Information Technology
Budapest, 1117 Hungary
http://www.iit.bme.hu

Abstract. The computing power of modern GPUs makes them very suitable for Computed Tomography (CT) image reconstruction. Apart from accelerating the reconstruction, their extra computing performance compared to conventional CPUs can be used to increase image quality in several ways. In this paper we present our upgraded GPU based image acquisition and reconstruction system, including gain correction, beam hardening correction, various reconstruction filters, and a new back-projection method for cone-beam geometry. We experimentally evaluated the impact of wide range of reconstruction filters on the image noise level and the resolution, and we compared the performance of the reconstruction components in phantom studies with special focus on image noise, resolution and ring artefacts.

Keywords: graphics processing unit, computer tomography, image reconstruction, CT artefacts

1   Introductions

Previously we developed a real time cone-beam helical CT (Computed Tomography) reconstruction [1, 2] that has been applied in the Mediso NanoPET/CT (Mediso Ltd.) equipment for in vivo imaging. To achieve the necessary computing performance, we decided to use a GPU (Graphical Processing Unit) for reconstruction. Using the GPU has reduced the reconstruction time dramatically. After achieving a real-time reconstruction we focused on image quality. In a post-reconstruction protocol there are less time constraints and more complex algorithms can be used for calculating the images. In this paper we go through the known fallbacks of the reconstruction method

we used (FDK-[3]) and present recent improvements. Some artefacts related to the acquisition process are also discussed.

High resolution micro-CT images are often superimposed by so called *ring artefacts*. Ring artefacts are concentric rings in the images around the centre of rotation of the CT setup, which are caused by imperfect detector elements e.g. gain inhomogenity in the detector array. We managed to suppress this artefact by an improved gain calibration and correction method.

Image noise is a major issue in in-vivo cone-beam CT. Flat panel X-ray detectors are generally not sensitive enough or have small dynamic range, and as dose and acquisition time are limited, the noise is increased. We implemented better filters before the back-projection to reduce noise, and also developed a more complex back-projection algorithm that handles the noise better.

Another challenge is the *Beam Hardening* effect. In X-ray computed tomography the decay of the incident intensity is related exponentially to the thickness of the object (Lambert-Beer law). In practice the X-ray beam is polychromatic and the Lambert-Beer law is no longer valid empirically because different energies attenuate differently (the out-scattering coefficient is wavelength dependent). As the FDK reconstruction uses only one energy, the beam hardening effect will cause the so called cupping artefact introducing gradients in homogeneous regions of the reconstructed image.

Metallic artefact is generated by objects with saliently strong attenuation in the field of view of an acquisition. Filtered back-projection algorithms are creating very shiny stars around these objects. Our improved back-projection algorithm performs better in this sense, but the artefact can only be eliminated by more complex methods, e.g. using iterative reconstruction, not dealt with it in this paper.

2   Materials and Methods

To evaluate the accuracy of a cone-beam CT system, a comprehensive test object is needed. The vmCT-610 Micro-CT Performance Evaluation Phantom (Shelley Medical Imaging Technologies) is a widely accepted tool, designed for characterizing cone-beam CT performance [4]. Sections of the phantom provide several possibilities to measure the performance of the system in many aspects like uniformity, noise, linearity and spatial resolution.

To evaluate the performance of the scanner, we measured all the modular plates with three different tube voltages (45, 55 and 65kVp), three exposure times (600, 1100 and 1600ms) and three zooms settings (1.4×, 2.5× and 5×).

In addition to using this professional phantom, we also built and measured a simpler one. Our phantom is divided into three areas: water, air, and a slanted edge (for resolution measurement). The aim of measurements with the simple phantom is to define an executable quality control routine for the NanoPET/CT system.

2.1   Resolution measurement

The common measure of resolution is the spatial frequency where the modulation transfer function (MTF) falls to 2% of the maximum value. The MTF is calculated from the edge spread function:

$$MTF(u) = \left| \int_{-\infty}^{\infty} \left(\frac{d}{dx} ESF(x) \right) \exp(-2\pi iux) dx \right|$$

where the edge spread function ($ESF(x)$) is the image of a slanted edge.

In practice the average ESF along a line perpendicular to the edge is used. The intensity of each voxel in the image is plotted against its distance from the edge. The edge is found by using the Canny edge detector. After deriving the ESF the following model function is fitted to the measured data:

$$ESF(x) = c_1 + \sum_{i=1}^{J} \frac{b_i}{2} \left(1 + erf\left(\frac{x - x_0}{\sqrt{2}a_i} \right) \right)$$

with

$$b_J = c_2 - c_1 - \sum_{i=1}^{J-1} b_i$$

The c_1 and c_2 parameters correspond to the asymptotic values of the ESF. The a_i is the spread, the b_i is the amplitude and x_0 is the position of the ith error function. We found that $J=2$ is enough to represent the ESF.

2.2   Noise measurement

The noise is measured in the water part of our image quality phantom. A VOI is selected using thresholding and the standard deviation is calculated.

3   Increasing the image quality

We tested all standard configurations of the Mediso NanoPET/CT system. After executing many data acquisitions and measurements, we identified a few major image quality issues: the ring and the beam-hardening artefacts, and image noise. We focused our development on these topics.

3.1   Ring artefact

The detector element of our CT is a CMOS active pixel sensor (Rad-icon Corp, CA, USA). This type of detector is inherently less linear than a CCD or passive pixel sensors, as the charge to voltage conversion takes place at the pixel level and not at a common readout amplifier.

The slight inhomogeneous nonlinear gain of pixels has little consequence on the linearity of the reconstructed CT image compared to e.g. the beam hardening effect,

but due to its deterministic nature it can cause severe ring artefacts. Nonlinear pixel-wise gain correction can almost completely eliminate this artefact.

The basic idea of the gain calibration is to record the response of each pixel at different known X-ray intensities and calculate the difference between the ideal and the actual signal as a function of the X-ray intensity. To produce controlled X-ray illumination, varying the tube current would be the most straightforward method. This did not work for us however, because changing the current caused small movement of the X-ray focal spot and hence changed the light distribution over the detector panel.

Another option to tune the illumination is to vary the exposure time. Unfortunately at the shortest exposure time of the detector (370 ms) the pixel response is still much higher than in the shadow of a typical sample, so we measured the behaviour at low illumination by covering the detector surface with an aluminium plate. Both "covered" and "direct" images were recorded at several exposure times. The response curve of our detector has a characteristic S-shape: it starts nonlinearly at low exposure, has a relatively linear region at intermediate exposure and saturates at strong illumination. We utilized the linear part, available in both the "covered" and the "direct" image sets to scale together the two sets. Linear region of the response curve was determined experimentally at several X-ray energies. The scaling is done for each pixel separately. We fit a fourth order polynomial to the "merged" pixel response – exposure time curve.

Correction is done by evaluating the polynomial at a recorded pixel value and replacing the pixel by the result. In figure 1 the difference between the old (quadratic) and the new (fourth order) gain calibration can be seen.
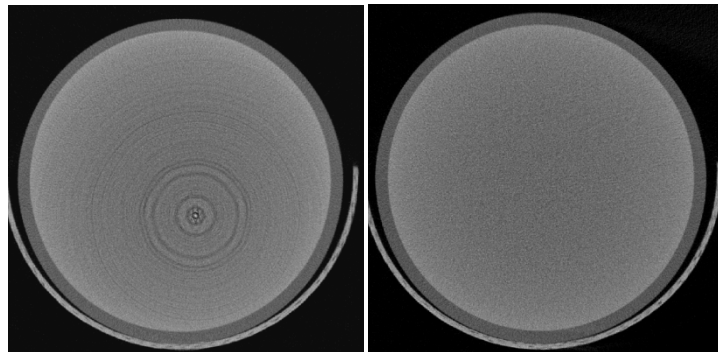

Fig. 1. Left image: Quadratic gain correction with several ring artefacts. Right image: new fourth order gain correction without the ring artefact. The same water phantom was scanned.

3.2 Noise

The FDK algorithm is basically a filtered back-projection that applies a filter in the frequency domain before back-projecting the projection frames. The mathematical derivation prescribes the RAMP (a.k.a. RamLak) filter which suppresses the low

frequency components but raises the high frequency parts of the signal. In practice this filter gives suboptimal noise performance because the highest frequencies often do not contain useful information and the high frequency edge of the filter behaves badly during the discrete inverse Fourier transform. We can combine the RAMP filter with several window functions for filtering the high frequency components. There are some well known window functions for the back-projection operator of a CT reconstruction. The new filters named after the window function are simply the product of the original RamLak filter and a window function.

We can categorize the filters into three groups. The RamLak, Butterworth and the Shepp-Logan are high-resolution filters. In the balance between resolution and noise they are definitely on the resolution side. For the most common cases we suggest to use the Shepp-Logan filter as it reduces the noise by at least 10% while the resolution is almost 96% of the RamLak filtered image. The Butterworth filter has even less adverse effect on the resolution while still improving noise performance.

In the second group there are the Cosine, Hamming and Hann filters. The noise reduction of these filters is better than that of the high-resolution filters; on the other hand the resolution is also reduced.

We experimented with several low-resolution filters but decided to keep only the Blackman filter. Low-resolution filters cut off not only the noise but also reduce the resolution significantly. With the Blackman filter the noise is about the 60%, the resolution is proximately 75% of the RamLak filtered image.

The characteristics of the different filters are shown in Figure 2, while the resolution and noise measured in a water phantom are plotted in Figure 3. The resolution is expressed in 2% MTF and the noise is the standard deviation of water in Hounsfield units.
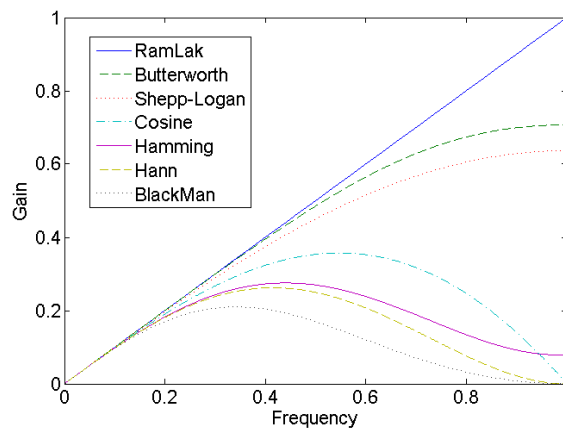

Fig. 2. Filter characteristics. Horizontal axis represents the frequencies from 0 (base frequency) to 1 (highest frequency). Vertical axis represents the gain (multiplier) for the current frequency.
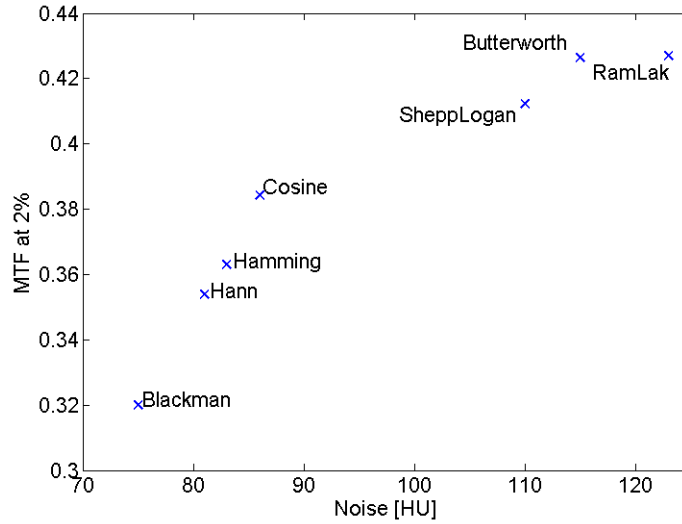
Fig. 3. Resolution versus noise evaluated with different reconstruction filters. The noise level in Hounsfield Unit and resolution in MTF.

The micro-CT systems are intended for creating high resolution images so low resolution filters are not an option. With high resolution filters we achieved noticeable but not satisfactory noise reduction.

Besides filtering, improving the back-projection operator (BP) is the way to enhance image quality. Due to limited computing power, the exact BP is only approximated in contemporary algorithms. Currently there are three different approaches for BP: pixel-, voxel-, distance-driven.

In the first case rays are generated from the detector pixels. They go through the whole volume and the contribution of the current pixel to the voxels among the actual ray is calculated [5, 6].
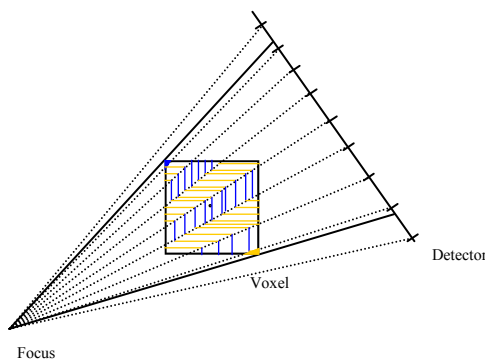

Fig. 4. Pixel-driven back-projection schema.

The pixels of the projection image are normalized by function:

$$P'_{\theta,i} = -\ln(\frac{I_{\theta,i}}{I_0})$$

and the image is filtered ($P_{\theta,i}$) by the selected filter (see previous section). In Figure 4 a single voxel is shown. The incremental voxel value for the current projection (at θ angle) is

$$V_{\theta,j} = \sum_{i=0}^{i=N}(P_{\theta,i} \cdot \frac{V_i}{V_v})$$

where V_i is the volume of the section of the voxel and the "ray-pyramid" and V_v is the volume of the voxel. In 3D the base of the "ray-pyramid" is the current detector pixel, and the zenith is the focus point of the X-ray source. In Figure 4 the two dimensional case is presented for simplicity; blue and yellow areas represent sections of the "ray-pyramids". In practice the sections are usually approximated by starting several rays from each pixel. The calculations are still complex and have to be repeated for each projection image, which makes the algorithm slow. Another disadvantage of this method is that its implementation for GPU is not efficient, because there can be write collision in the GPU memory when writing back the calculated (incremental) value of the voxel. The implementation is possible by using atomic operations (e.g. atomic add) but that significantly slows down the reconstruction speed.

The distance-driven (DD) projector [7, 8] is a current state-of-the-art method. It maps the horizontal and vertical boundaries of the image voxels and detector cells onto a common plane such as xz or yz plane, approximating their shapes by rectangles. It calculates the lengths of overlap along the x (or y) direction and along the z direction, and then multiplies them to get the area of overlap.

The voxel-driven (a.k.a. ray-driven [9, 10]) method generates rays from the focal point of the X-ray source through the voxels. The projection is independent for each voxel, making a parallel implementation that suits for the GPU architecture straightforward. The difficult part of the algorithm is to calculate the footprint of the voxels (see Figure 5). This is a function generated by projecting the 3D voxel onto the detector surface at a given angle. More information about the footprint function can be found in [11].
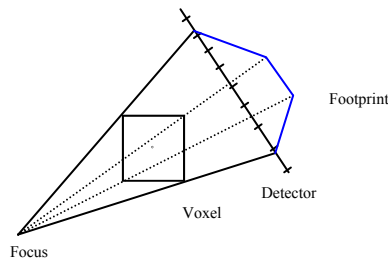

Fig. 5. Voxel-driven back-project schema with voxel footprint.

For a given θ angle we define the $f_{\theta,j}(t)$ footprint function, where j is the index of a voxel and t is a point on the detector. The incremental voxel value for the θ angle is:

$$V_{\theta,j} = \frac{\int (P_\theta(t) \cdot f_{\theta,j}(t)) dt}{\int f_{\theta,j}(t) dt}$$

where $P_\theta(t)$ function is the logarithmically normalized and filtered value of the detector at point t, the integrals are calculated for the full detector surface. The value of a voxel is the sum of all the incremental values of the voxel calculated for all detector angles (similarly to the pixel-driven method).

Unfortunately this quasi exact solution gives a rather slow algorithm, although recently Long and Fessler [11] suggested a new method. Their paper describes a separable approximation of the footprints for the integration that accelerates the calculation considerably. We will try to incorporate this method in the future.

This integration can be approximated by sampling if we choose the distribution of the sampling points proportional to the current integrand, the $f_{\theta,j}(t)$ footprint function. Furthermore if we are sampling a voxel with a homogeneous distribution, the projections of these points will represent the necessary distribution (see Figure 6). Calculation of the incremental voxel value now simplifies to the following equation:

$$V_{\theta,j} = \frac{\sum_{k=0}^{k=S} P_{\theta,k}}{S},$$

where is S the number of sampling points and $P_{\theta,k}$ is the normalized filtered pixel value for the sample point k.

How can we choose the best sampling method? We can use a regular three dimensional grid inside the voxel. It is a very simple method but if we use some low-discrepancy sequence we can reach more uniform distribution [12]. There are several low- discrepancy sequences like Sobol's, Halton's or Poisson Disk (PD) [13, 14]. We choose the Poisson Disk [15] distribution. With Poisson Disk sampling we have achieved lower noise levels with only half of the points we used for regular sampling. Moreover with quasi-random sampling we can avoid the Moiré effect which is the most common artefact of the voxel-driven BP. Generating the PD samples is very computation intensive, but it can be done offline. So we have generated a couple of sample point sets offline and later simply choose one for a selected projection. With this multi-sampled voxel-driven BP we can significantly suppress the noise without lowering the resolution.
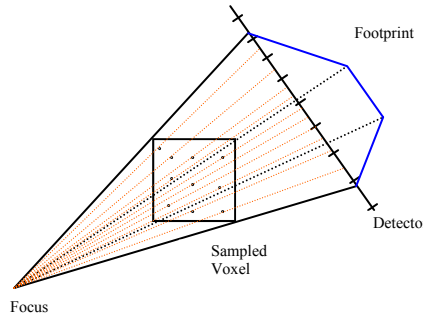

Fig. 6. Multi-sampled voxel-driven back-project schema with voxel footprint.

In Figure 7 we compare image noise using the normal (3D regular) sampling and the PD sampling for reconstruction. Beside the type of the sampling and the number of samples all other parameters of the reconstruction are the same.
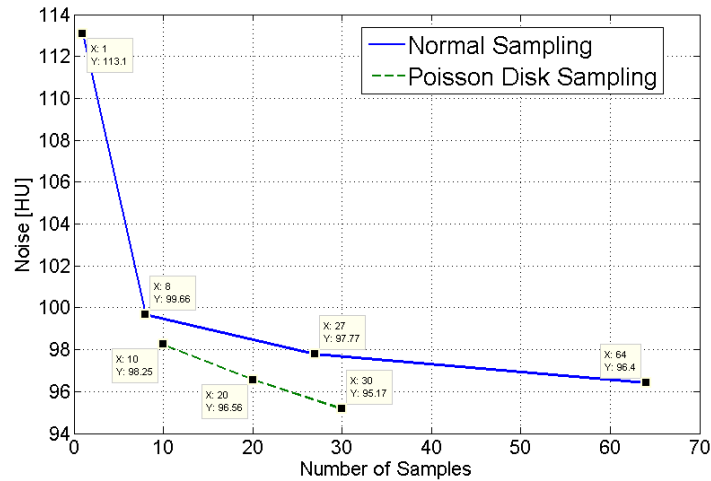

Fig. 7. Noise level (in Hounsfield Unit) of different sampling methods, and various number of samples.

By increasing the size of the voxels the noise is reducing but the PD sampling always gives better results. If we use only one sample, the curve is flattening faster than in the PD sampling case (Figure 8). The necessary number of sampling points is increasing for larger voxels but we can calculate it before the reconstruction (larger voxels generate larger footprint):

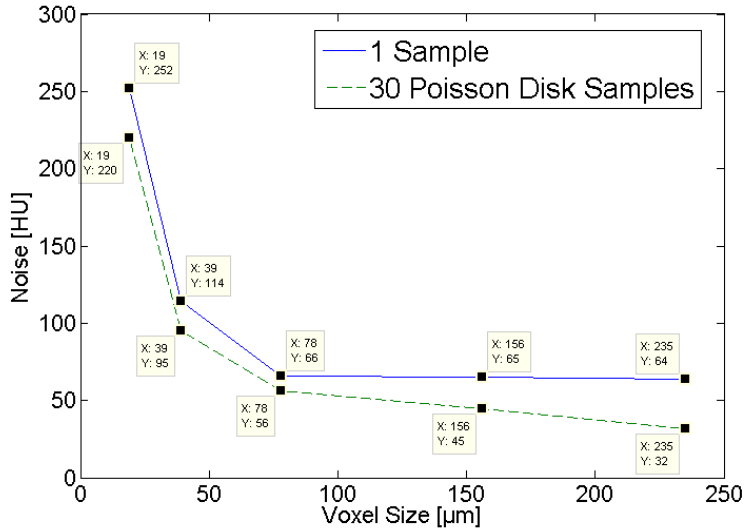| voxel size:	19 um	39 um	78 um	156 um	235 um
Number of samples needed for optimal footprint coverage:	10	20	30	40	50

Fig. 8. Noise level (in Hounsfield Unit) for different voxel size and sampling method.

Multisampling makes the reconstruction time longer, but not dramatically because we do not have to do all calculations as many times as sampling points we have inside the voxel. A lot of optimization can be done. The projection image is stored in the texture memory of the GPU so it is spatially cached.

Number of samples	10	20	30	64
Recon. time multiplier	2.1	4.2	5.8	11

According to our measurements, 30 PD sample points are enough in most cases.

With this upgraded BP operator the noise level can be reduced by about 25-30% without lowering the spatial resolution of the reconstructed image. We can combine this new BP operator with the Shepp-Logan filter and we can get a 40% noise reduction with high resolution.

3.3 Beam Hardening

In CT projection the beam hardening causes a nonlinear relation between the thickness and the attenuation (see Figure 9). If we use only one energy model for reconstruction (like the traditional FDK), cupping artefacts can be seen in reconstructed images.
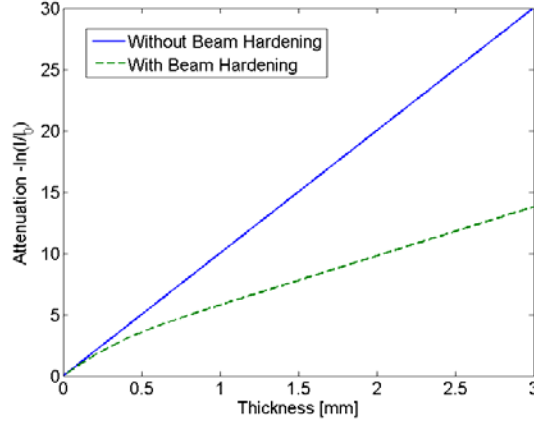
Fig. 9. The beam hardening effect for a homogenous object.

The nonlinearity of the attenuation can be approximated by using the bimodal energy model [16]:

$$-\ln\frac{I}{I_0} = \mu_2 d + \ln\left(\frac{1+\alpha}{1+\alpha\exp(-(\mu_1-\mu_2)d)}\right)$$

With

$$\mu_1 = \mu(E_1)$$
$$\mu_2 = \mu(E_2)$$
$$\alpha = \frac{f(E_1)\gamma(E_1)}{f(E_2)\gamma(E_2)}$$

where $f(E)$ is the power spectrum of the source and $\gamma(E)$ is the detector efficiency.

At the calibration process we get the constant values for μ_1 and μ_2 and a pixel dependent α matrix. Using these parameters we can linearize the thickness dependency by correcting the intensities of the pixels. The thickness can be derived from the measured data, and then the thickness should be used for calculating the corrected pixel intensity

$$-\ln\frac{I'}{I_0} = \frac{\mu_1\alpha + \mu_2}{1+\alpha}d$$

where I' is the corrected intensity value.

4   Results and conclusion

Putting all proposed techniques together, we decided to acquire a real mouse to see the image quality improvement. We are interested in mouse lung because in these images there are both low and high contrast areas, there are bones and soft tissues, and we can see the small parts inside the lung.

In Figure 10 we compare our results. The ring artefact is completely removed, and thanks to our noise reduction methods the resolution can be increased. The Moiré pattern is also suppressed and less metallic artefact is noticeable in the new image.
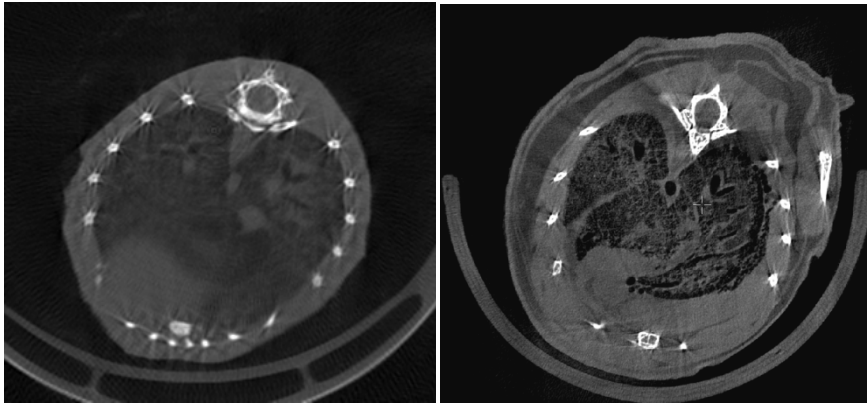

Fig. 10. Mouse lung slices. The left image is generated by the classic acquisition and reconstruction process, the right image presents the new methods.

The new method applied multi-sampled voxel driven BP with 30pcs Poisson Disk samples, fourth order gain correction, and Shepp-Logan filtering. The classic method used as a reference executed voxel-driven back-projection with one sample for one voxel, quadratic gain correction, and Hamming filtering.

The common acquisition parameters are as follows:

- Helical scan with 540 projection for each rotation,
- Normal projection size (1:4 binned frames, resolution: 512×1798 pixels),
- Reconstructed voxel size: $39\mu m$.

With the power of the GPUs, we obtained high quality reconstruction in reasonable time. High resolution images can be generated with less artefact and noise. A whole body mouse reconstruction time is about half an hour with $39\mu m$ voxels (Shepp-Logan filter, 30 PD samples/voxel, all corrections are enabled). That means 768×768×2560 voxels for the volume (almost 1.5 Giga voxels).
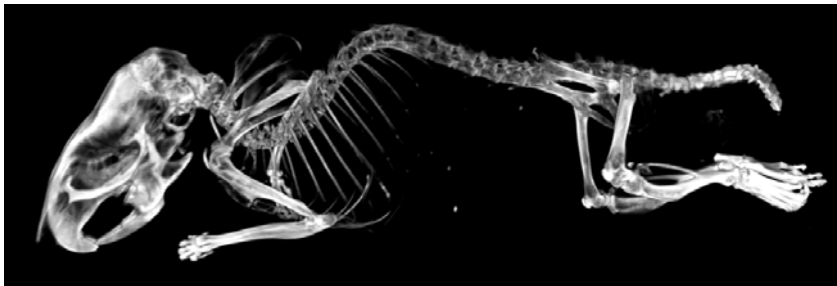

Fig. 11. Full CT scan of the mouse.

5   Acknowledgement

References

1. Jakab, G. and Domonkos, B.: Valósidejű helikális cone-beam CT rekonstrukció a Mediso NanoPET/CT készülékben. Képfeldolgozók és Alakfelismerők (KÉPAF) 7. Konferenciája. Budapest, Hungary (2009)
2. Jakab, G., Rácz, A., Németh, G., and Bükki, T.: Fully GPU Based Real Time Corrections and Reconstruction for Cone Beam Micro CT. IEEE - Science Symposium Conference Record (NSS/MIC), pp. 4068--4071. Orlando, FL, USA (2009)
3. Feldkamp, L. A., Davis, L. C., and Kress, J. W.: Practical cone-beam algorithm. Journal of the Optical Society of America, 612-619. (June 1984)
4. Du, L. Y., Umoh, J., Nikolov, H. N., Pollmann, S. I., Lee, T. Y., and Holdsworth, D.W.: A quality assurance phatom for the performance evaluation of volumetric micro-CT systems. Phys. Med. Biol. 52 7087-7108. (2007)
5. Peters, T.: Algorithms for fast back- and re-projection in computed tomography. IEEE Trans. Nucl. Sci. 28 3641–3647. (1981)
6. Herman, G.: Image Reconstruction from Projections. Orlando: Academic p 63. (1980)
7. De Man, B. and Basu, S.: Distance-driven projection and backprojection. IEEE Nuclear Science Symp. Medical Imaging Conference. Norfolk. (2002)
8. De Man, B. and Basu, S.: Distance-driven projection and backprojection in three dimensions. Phys. Med. Biol., vol. 49, no. 11, pp. 2463–2475. (June 2004)
9. Siddon, R.: Fast calculation of the exact radiological path length for a three-dimensional CT array. Med. Phys. 12 252–255. (1985)
10. Zhuang, W., Gopal, S., and Hebert, T.: Numerical evaluation of methods for computing tomographic projections. IEEE Trans. Nucl. Sci. 41 1660–1665. (1994)
11. Long, Y. and Fessler, J. A.: 3D Forward and Back-Projection for X-Ray CT using Separable Footprints. IEEE Transactions on Medical Imaging, Pages 1839-1850. (2010)
12. Szirmay-Kalos, L. and Szécsi, L.: Deterministic Importance Sampling with Error Diffusion. In Computer Graphics Forum (EG Symposium on Rendering) (2009), vol. 28, pp. 1056–1064.
13. Kocis, L. and Whiten, W. J.: Computational Investigations of Low-Discrepancy Sequences. ACM Transactions on Mathematical Software, Vol. 23, No. 2, Pages 266–294. (June 1997)
14. Szirmay-Kalos, L.: Monte-Carlo Methods in Global Illumination - Photo-realistic Rendering with Randomization. VDM, Verlag Dr. Müller, Saarbrücken, 2008.
15. Jones, T. R.: Efficient Generation of Poisson-Disk Sampling Patterns. (April 24, 2006.)
16. Van de Casteele, E.: Model-based approach for Beam Hardening Correction and Resolution Measurements in Microtomography Phd. Thesis, Antwerpen (2004)