

ACCELERATION OF IMAGE RECONSTRUCTION PROCESS IN THE ELECTRICAL CAPACITANCE TOMOGRAPHY 3D IN HETEROGENEOUS, MULTI-GPU SYSTEM

Michał Majchrowicz, Paweł Kapusta, Lidia Jackowska-Strumiłło, Dominik Sankowski

Lodz University of Technology, Institute of Applied Computer Science

Abstract. Electrical capacitance tomography is an innovative method for visualization of industrial processes. One of its main advantages is its high time resolution that allows to the usage of ECT in systems with high volatility. In recent years there has been significant development of electrical capacitance tomography 3D, which however, has significantly reduced industrial it's applications due to the complicated process of image reconstruction. The authors propose the use of multi-node, multi-GPU system to accelerate the process of image reconstruction in ECT 3D.

Keywords: Electrical capacitance tomography, reconstruction algorithms, matrix computations, distributed systems.

PRZYSPIESZANIE PROCESU REKONSTRUKCJI OBRAZU W ELEKTRYCZNEJ TOMOGRAFII POJEMNOŚCIOWEJ 3D Z WYKORZYSTANIEM HETEROGENICZNEGO SYSTEMU MULTI-GPU

Streszczenie. Elektryczna tomografia pojemnościowa jest innowacyjną metodą wizualizacji procesów przemysłowych. Jedną z jej głównych zalet jest duża rozdzielczość czasowa pozwalająca na zastosowanie ECT w instalacjach o dużej zmienności. W ostatnich latach nastąpił znaczący rozwój elektrycznej tomografii pojemnościowej 3D, która jednakże ma znacznie ograniczone zastosowanie przemysłowe ze względu na skomplikowany proces rekonstrukcji obrazu. Autorzy artykułu proponują zastosowanie wielowzłowego systemu Multi-GPU do przyspieszenia procesu rekonstrukcji obrazu w ECT 3D.

Słowa kluczowe: elektryczna tomografia pojemnościowa, algorytmy rekonstrukcji, obliczenia macierzowe, systemy rozproszone.

Introduction

Electrical capacitance tomography is an innovative visualization method of industrial processes. One of its main advantages is its high time resolution that allows for the application of ECT in systems with high volatility. In recent years there has been significant development of electrical capacitance tomography 3D, which has significantly reduced industrial use due to the complicated process of image reconstruction. The authors propose the use of a multi-GPU system to accelerate the process of image reconstruction in 3D ECT.

1. Electrical Capacitance Tomography and GPGPU

The electrical capacitance tomography 3D approximation of the spatial distribution of electric permittivity inside the object is calculated using the knowledge capacity measurements with the use of electrodes, which are respectively arranged on the surface of the object. Since 2003 develops a large interest in the use of real, three-dimensional, volumetric ECT, both in static mode, 3D and 4D dynamic. Electrodes on the surface of the sensor provide electrical potential to the area surrounded by the sensor and then measuring its change in the boundary [1].

These measurements are then used in an electrostatic model to determine the dielectric constant of the object [2]. Today, 3D ECT becomes an important tool for visualization in the industry [5].

This imaging technique, however, has one major drawback - in the case of three-dimensional image reconstruction time can be unacceptably long (up to several hours and in some circumstances even days). To break this barrier authors propose a new method for accelerating image reconstruction in ECT using OpenCL technology and Multi-GPU (Graphics Processing Units) systems.

General-Purpose computing on Graphics Processing Units (GPGPU) is a technique of using graphic cards (GPU - Graphics Processing Unit), which normally handles graphics rendering, for computations that are usually handled by processors (CPU - Central Processing Unit). Growing interest in GPU computations started due to constant demand for higher compute capabilities, which was necessary to solve more complex algorithms. Because of this interest multi-core technology and parallel computing started to emerge.

Parallel programming is not a new idea, though till only recently it was reserved for high performance clusters with many processors, the cost of such solution was extremely high [3]. This

changed with the introduction of many core processors to the mainstream market. GPUs fit well in that trend, even take it to another level. Compared to CPUs, which today have usually 2 to 16 cores, GPUs consist, of hundreds and even thousands of smaller, simpler cores designed for high-performance calculations. Thanks to that there can be much more of them on a single chip (Fig. 1), which in turn allows running many thousands of threads at once, compared to only few on CPU [3]. All this made development of new algorithms possible, using higher computing power of the GPUs. Many computations that were compute heavy and time consuming can now be made in close to real time, with small investments in hardware compared to the cost of achieving the same results using established methods.

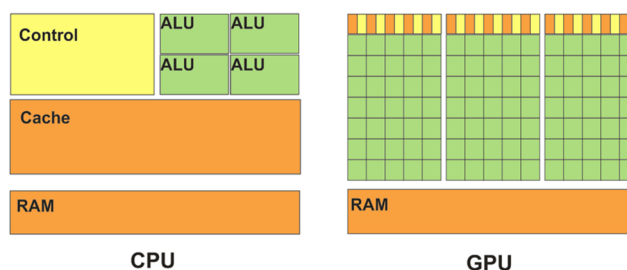


Fig. 1. Comparison CPU and GPU architectures

Moreover, thanks to advancements in computer graphics, multi-GPU systems have been developed – combining multiple graphics processors in a single computer. This allows for further speed-up of computations, provided the algorithms are properly adapted to such configuration, which is not trivial.

2. Developed Solution

The authors propose an innovative approach to 3D image reconstruction in ECT using multi-GPU solutions. Instead of sharing one job (picture frame) between multiple GPUs, each node gets its own frame of video to calculate. Such an approach does not decrease the time needed to calculate a single frame, but by application of the synchronization and load balancing algorithms results can be evenly distributed in time and it is possible to achieve a constant number of frames per second, and thus smooth image. This approach introduces a delay equal to the time required to reconstruction of single image frame by the slowest available GPU, but may be applied effectively in situations when reaction

time is not critical, for example, visualization of already collected data as well as for testing algorithms.

3. Verification

In order to verify the proposed solution authors have implemented a modified version of a Landweber iterative algorithm [6, 7], which can be run on both the CPU and GPU. It is described by the following equation:

$$\boldsymbol{\varepsilon}_{k+1} = \boldsymbol{\varepsilon}_k - \alpha \cdot \mathbf{S}^T \cdot (\mathbf{S} \cdot \boldsymbol{\varepsilon}_k - \mathbf{C}_m) \quad (3)$$

where:

$\boldsymbol{\varepsilon}_{k+1}$ – the image obtained in the current iteration,

$\boldsymbol{\varepsilon}_k$ – the image obtained in the previous iteration,

α – convergence factor,

\mathbf{S} – sensitivity matrix,

\mathbf{C}_m – capacity measurements vector.

Execution times of the algorithm using the GPU were then compare with the times for execution on a traditional processor as well as with the algorithm LBP [6], described by the equation:

$$\boldsymbol{\varepsilon} = \mathbf{S} \cdot \mathbf{C} \quad (2)$$

As can be seen from the equation, this algorithm is much less computationally complex, but does not allow for the achievement of such good quality as Landweber's algorithm, and therefore is not the optimal choice for Electrical Capacitance Tomography 3D [7].

CPU tests were conducted on an Intel i7 930 clocked at 2.8 GHz. GPU used for tests were Nvidia Tesla C2070 and server Nvidia Tesla S1070 as well as the GeForce 8600GT, which was not involved in the calculation, and served only for displaying an image. In the case of AMD Radeon HD5970 GPUs were used.

Tests were conducted on two sets of test data (sensitivity matrices). The medium one has about 2.25 million items and will be later in the article called "average mesh". Meshes of this size are often used in the visualization, because they provide satisfactory results within a reasonable time. The second mesh has approximately 11 million data points and occupies over 250 MB after writing to disk and will be later in the article called "large mesh".

For simplicity we assume that the image quality produced by the Landweber's algorithm is proportional to the number of performed iterations. Therefore, tests were conducted at 100, 200 and 400 iterations to test the behavior of the implementation as well as to fully exploit the available hardware computing power.

In the case of GPU always the worst of calculation times were taken into account. The current implementation of OpenCL has a certain instability, so that execution times can vary by 5-10%. Therefore, we decided to pay special attention to the worst times of execution, not the average or the best cases, since worst cases are more important from an environmental perspective for reconstruction in real time.

3.1. CPU Computations

The Authors first carried out calculations using the CPU to achieve exemplary results, which were then compared to those obtained using the GPU. Results have been collected in Table 1.

Table 1. Reconstruction time [ms] for computations using CPU

	Average mesh [ms]	Large mesh [ms]
LBP CPU	32	117
Landweber CPU – 100 iterations	4415	22371
Landweber CPU – 200 iterations	8783	44632
Landweber CPU – 400 iterations	17488	89201

As can be seen reconstruction of a tomographic image frame, using only the conventional processors, can (in so considered cases) take up to 90 seconds, which is not acceptable value in both

the visualization and adaptation and optimization of algorithms. It can be also noted that the LBP algorithm is more than an order of magnitude faster than the Landweber's algorithm, but, as we have indicated before, it does not provide as good image quality.

3.2. Single GPU Computations

To measure the acceleration of calculations that can be obtained using a single GPU authors conducted a test using cards AMD Radeon HD5970 GPU and a single component of the computing server Nvidia Tesla S1070.

Table 2. Reconstruction time [ms] for computations using single GPU

	Average mesh [ms]	Large mesh [ms]
Tesla system – 100 iterations	1100	5064
Tesla system – 200 iterations	1697	11050
Tesla system – 400 iterations	4038	22574
System AMD – 100 iterations	1793	9630
AMD system – 200 iterations	3973	17890
AMD system – 400 iterations	9879	35657

As can be seen by analyzing the results listed in Table 2 calculations using GPU Tesla are more than 4 times more efficient than using a conventional processor for 100 iterations, more than five times faster for 200 iterations, and surprisingly only four times faster for 400 iterations. AMD graphics processor was about 2.5 times faster than the CPU for 100 iterations, 2.2 times higher for 200 iterations and 1.8 times higher for 400 iterations.

For a large mesh results do not differ significantly from the previous case. A single GPU Tesla to 4.4-times, 4-times and 3.6 times faster than the CPU for 100, 200 and 400 iterations of the Landweber's algorithm. A single AMD GPU under the same conditions achieved 2.3-fold, 2.5-fold, and again 2.5 times faster with respect to the CPU.

3.3. Multi-GPU system

As shown in Table 2, even application of a single GPU allows for considerable acceleration of tomography computation. But this is not enough, so we conducted a test in configurations of multiple graphics cards (Multi-GPU). In this article we describe cases of configuration consisting of two, four, and all available (five if using Nvidia and eight using AMD) GPUs.

3.4. Landweber's Iterative Algorithm - 2 GPUs

Table 3. Reconstruction time [ms] for a Multi-GPU system – 2 GPUs

	Average mesh [ms]	Large mesh [ms]
Tesla system – 100 iterations	570	2555
Tesla system – 200 iterations	871	5553
Tesla system – 400 iterations	2040	11307
AMD system – 100 iterations	918	4836
AMD system – 200 iterations	2010	8966
AMD system – 400 iterations	4960	17849

For this configuration, the average mesh, compared to the calculation on the processor, Tesla GPUs are 7.7 times, 10 times and 8.6 times faster at 100, 200 and 400 iterations. In the case of AMD it reached 4.8-fold, 4.4-fold and 3.5-fold acceleration of the reconstruction time.

For a large mesh 8.75 times, 8 times and 7.9 times better performance was achieved than in case of single CPU for Nvidia

hardware, and 4.6 times, 5 times, and again 5 times faster calculation time on the AMD GPUs.

3.5. Landweber’s Iterative Algorithm - 4 GPU

Table 4. Reconstruction time [ms] for a Multi-GPU system – 4 GPUs

	Average mesh [ms]	Large mesh [ms]
Tesla system – 100 iterations	296	1287
Tesla system – 200 iterations	449	2787
Tesla system – 400 iterations	1033	5664
AMD system – 100 iterations	470	2430
AMD system – 200 iterations	1017	4496
AMD system – 400 iterations	2492	8937

For a configuration with four GPUs and the average mesh gave the following results:

- in the case of Tesla GPU - 14.9 times faster for 100 iterations 20 times faster for 200 iterations and 17 times faster for 400 iterations;
- in the case of the Radeon GPUs - 9.4 times faster for 100 iterations, 8.6 times faster for 200 iterations and 7.2-fold acceleration for 400 iterations.

For a large mesh results are as follows:

- in the case of GPU Tesla - 17.4 - 100 times faster for a 16-fold for 200, and 15.75 times faster for 400 iterations with respect to the CPU;
- The GPU Radeon - 9.20 times faster at 100, 9.9-fold for 200 and 10 - 400 times faster for iteration, compared with the CPU.

3.6. Landweber’s Iterative Algorithm - All Available GPUs

In this case the tests were conducted using all available GPUs in the system, which meant 5 for Nvidia and 8 for AMD. The calculation results are shown in Table 5.

Table 5. Reconstruction time [ms] for a Multi-GPU system – all available GPUs

	Average mesh [ms]	Large mesh [ms]
Tesla system – 100 iterations	245	1038
Tesla system – 200 iterations	364	2237
Tesla system – 400 iterations	837	4543
AMD system – 100 iterations	250	1231
AMD system – 200 iterations	524	2263
AMD system – 400 iterations	1266	4487

For average mesh:

- In case of Tesla GPUs calculations were 18.2-fold, 24-fold and 21 times faster than on CPU for respectively 100, 200 and 400 iterations;
- AMD achieved 17.8-fold, 16.8-fold and 13.9-fold acceleration calculation for 100, 200 and 400 iterations.

For a large mesh:

- For Tesla GPU a 21.5 times acceleration time speedup for 100 iterations, 20 times for 200 and 19.65 times for 400 iterations.
- Configuration of AMD was respectively 18.21 - times, 19.7 times and 19.9-times faster for 100, 200 and 400 iterations of the Landweber’s algorithm than on CPU.

3.7. GPU Results Comparison

To better visualize the results, and show differences between GPU different manufacturers results were placed on the collective chart for both the case of medium and large nets.

3.8. Average mesh

Using a single GPU solution from Nvidia turned out to be more powerful than the Radeon AMD in each test, 1.6 times, 2.3 times and 2.45 times for the 100, 200 and 400 iterations.

Two Tesla cards were also faster than two AMD cards - 1.5 times for 100 iterations. The difference has widened even for 200 and 400 iterations, where Nvidia cards turned out to be respectively 2.30 times and 2.4 times faster than the Radeon.

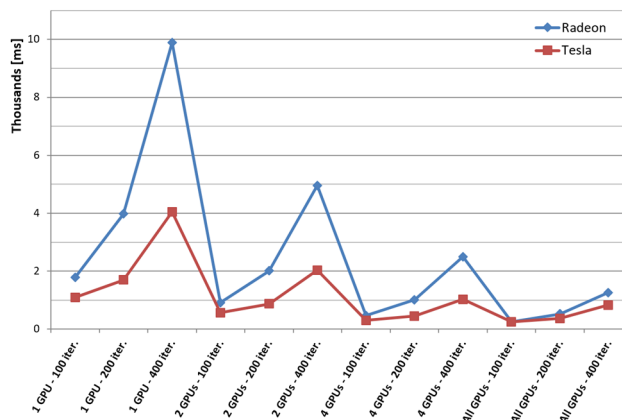


Fig. 2. Comparison of computation times on GPU – average mesh

In the case of four Tesla GPUs were 1.6 times, 2.3 times and 2.4 times (for respectively 100, 200 and 400 iterations) faster than AMD cards. Using all GPU authors found out that 5 Tesla GPUs configuration is only marginally faster than 8 AMD cards - 1.02 times for 100 iterations. In the case of 200 and 400 iterations Nvidia cards again turned out to be 1.4 times and 1.51 times faster. It seems that in the case of the Tesla GPU configurations GeForce 8600 GT which was dedicated only to display an image, played an important role since it looks that with the increase in computational complexity AMD cards most likely were not be able to cope with simultaneous calculation and displaying of a 3D images.

As can be seen in the chart smallest difference in execution time between Nvidia and AMD cards, the average mesh, occurs at 100 iterations, and the largest at 400. Additionally, you it can be also noted that with increasing the number of graphics cards gap is getting smaller. Moreover, adding new cards minimizes the impact on the overall number of iterations computing time.

3.9. Large mesh

For a large mesh single Tesla GPU was 1.90 times faster than AMD Radeon for 100 iterations, 1.61 times faster for 200 iterations and 1.57 times faster for 400 iterations. In the case of multi-GPU configurations for a large mesh difference is even greater. Two Tesla GPUs are 1.82 times faster than the two AMD GPUs for 100 iterations, 1.61 times faster for 200 iterations and 1.57 times faster for 400 iterations. The configuration of four Tesla GPUs was 1.88 times faster than the AMD for 100 iterations, 1.61 times faster for 200 iterations and 1.57 times faster for 400 iterations. In addition, 5 Nvidia Tesla GPU were 1.18 times faster than 8 AMD GPUs for 100 iterations. However, for 200 iteration of AMD configuration it is almost as fast as Tesla and for 400 iterations even little faster.

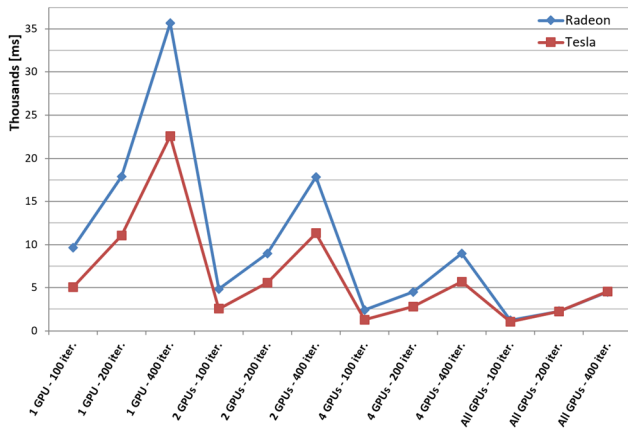


Fig. 3. Comparison of computation times on GPU – large mesh

You will notice that the mesh size has a huge impact on the differences between the Nvidia and AMD GPUs. Even with single GPU, we can see that the difference between the results for 400 iterations is less than in the previous case. In addition, increasing the number of graphics processors further reduces the difference, which completely disappears in the latter case. It is also worth noting that for all the GPU and 400 iteration AMD configuration was slightly faster, which proves that the data size has a huge impact on the performance of the devices.

These results clearly show that for small data sets and a large number of iterations NVIDIA products are a better choice, but when working with meshes of significant sizes AMD GPUs can be just as or even more powerful than Nvidia products.

4. Time-shifted reconstruction

The authors have concentrated their work on improving the speed on-line image reconstruction in 3D Electrical Capacitance Tomography. All the developed algorithm can be however used to test and verify a different approach - where system response time is not as important, as maximum achieved capacity. In this case the authors have prepared a database of 56320 measured capacitance vectors and tried to perform image reconstruction using 100 iterations of Landweber algorithm in shortest time possible. For this the authors have used the approach of vector consolidation [2], that combines input data in bigger packets and allows using faster reconstruction algorithms (Fig. 5). By combining many capacitance vectors into packets of 32 – 128 it converts all the matrix-vector operations into matrix-matrix, which are much more optimal for performing computations on graphic processors.

This approach, however, has one drawback as it introduces a delay to computations as the input data for image reconstruction algorithms is much bigger than usual. However, it also allows for much higher throughput.

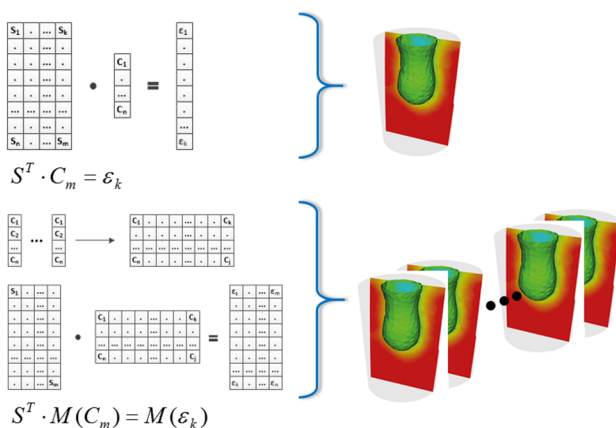


Fig. 4. Vector consolidation approach in 3D ECT image reconstruction

All the results for this test have been gathered in Table 6, as well as presented graphically on Figure 5. It is important to note, that in order to show the results properly, the axis for the CPU results have been scaled compared to GPU.

Table 6. Time-shifted (off-line) reconstruction time [s]

Image vector size	Intel i7 930 CPU [s]	Radeon HD 5970 1GPU [s]	Radeon HD 5970 2 GPUs [s]	Radeon HD 5970 4 GPUs [s]
8488	21270,0	937,2	496,7	257,7
20499	50533,6	1843,6	1004,8	511,6
60896	149841,4	5231,6	2877,4	1464,8
87152	219739,8	6617,6	3672,8	1886,0
157264	403546,2	11831,6	6625,7	3342,4

As can be seen from the data in Table 6 the proposed algorithm, when using a single GPU, is already 20 to 30 times faster than computations on a quad-core CPU. This advantage further increases with adding multiple GPUs to the system. Dual GPU system is up to 60 times faster and quad GPUs are up to 120 times faster than computations on CPU.

This speed advantage using a developed algorithm and Multi-GPU system makes it possible to reconstruct big sets of measurement data in a relatively short time. For example, in case of a resulting image vector, that consists of 157264 elements, performing image reconstruction for the whole test set takes more than four days. Performing the same operation using quad-GPU system this time can be shortened to just one hour. This makes it possible to perform any image analysis on reconstructed images much faster, than was possible before.

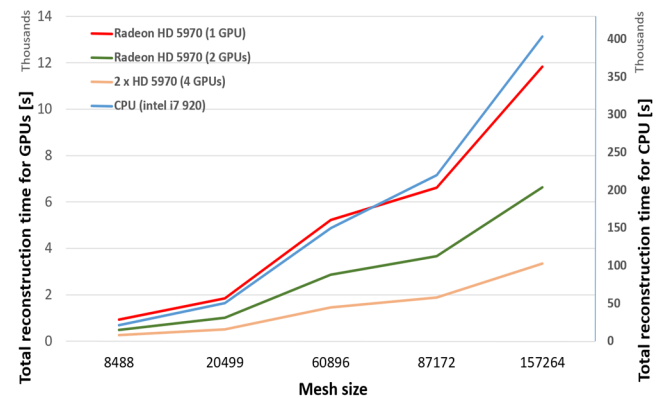


Fig. 5. Total computation time for time-shifted image reconstruction

5. Conclusion

Obtained results confirm the validity of the assumptions made by the authors that by using the proposed approach significant acceleration of the image reconstruction time in Electrical Capacitance Tomography 3D was achieved. In addition, according to the authors, after further work on the visualization algorithms on the GPU and the application of faster graphics processors even better results can be achieved.

The results also show that the great advantage of OpenCL framework, which is the possibility to execute the unmodified code on a variety of devices can also be a limiting factor. The carried tests showed that the developed algorithm prefers Nvidia cards over AMD products, although the latter have a much higher theoretical maximum power (single GPU AMD reaches 2320 GFLOP / s while the Nvidia GPU only 1088 GFLOP / s).

Furthermore, the authors have tested their algorithm not only as a platform for performing on-line reconstruction, but also to radically decrease the time necessary to visualize large sets of measurement data.

Studies have also shown that the use of multiple computers connected together to create a computing network may be a viable option. All the necessary information is stored in GPU RAM, and only measurement data and results (which are relatively small sets of data) have to be transmitted through the network, so its speed should not be a limiting factor.

Bibliography

- [1] Banasiak R., Wajman R., Soleimani M.: An efficient nodal Jacobian method for 3d electrical capacitance image reconstruction. *Insight Non-Destructive Testing and Condition Monitoring* 51(1)/2009, 36–38.
- [2] Kapusta P.: Algorytmy równoległego przetwarzania danych pomiarowych i obrazów w wybranych systemach tomografii przemysłowej, PhD Thesis, 2014, 47–50.
- [3] Kirk D., Hwu W.: *Programming Massively Parallel Processors, Second Edition: A Hands-on Approach*. Morgan Kaufmann Publishers, 2012.
- [4] Lionheart B.: Reconstruction algorithms for permittivity and conductivity imaging, *Proceedings of 2nd World Congress on Industrial Process Tomography*, 2001, 4–11.
- [5] Soleimani M.: Three-dimensional electrical capacitance tomography imaging. *Insight Non-Destructive Testing and Condition Monitoring*, 48(10)/2006, 613–617.
- [6] Wajman R., Banasiak R., Mazurkiewicz L., Dyakowski T., Sankowski D.: Spatial imaging with 3D capacitance measurements. *Measurement Science and Technology*, 17(8)/2006, 2113–2118.
- [7] Yang W.Q., Peng L.: Image reconstruction algorithms for electrical capacitance tomography. *Measurement Science and Technology IOP Journal* 14/2003, R1–R13.

Prof. Lidia Jackowska-Strumillo
e-mail: lidia_js@kis.p.lodz.pl

Professor at Lodz University of Technology (TUL), Poland and Vice-Director for didactics in the Institute of Applied Computer Science at TUL. She received the M.Sc., the Ph.D. and the DSc. degrees in electrical engineering from TUL in 1986, 1994 and 2010, respectively. In 1990/91 she has been staying in the Industrial Control Unit at the University of Strathclyde in Scotland where she worked on her Ph.D. project. From 1986 to 1998 she worked in the Institute of Textile Machines and Devices TUL. Her research interests include computer engineering, modelling of industrial objects and processes, artificial intelligence, computer measurement systems, identification methods, text processing, image processing and analysis.



Prof. Dominik Sankowski
e-mail: dsan@iis.p.lodz.pl

Head of the Institute of Applied Computer Science, at the Lodz University of Technology. He achieved a PhD with distinction (1979), DSc (habilitated doctor - 1989), and granted the title of professor (1998). Since 2001 he has held the position of full Professor. Member of: IEEE Computer Society, Electro-thermal Section of the Polish Academy of Sciences, Group of Experts in the Computer Science specialisation of National Accreditation Committee (PKA). Communication Director of International Society of Process Tomography (ISIPT). His research interests include Image Processing and Analysis, Pattern Recognition, Artificial Intelligence, Software Engineering, Data Processing, Real-Time Systems, IT Systems in Management, On-line Identification of Industrial Objects, Automation of the Thermal Processes.



otrzymano/received: 20.09.2016

przyjęto do druku/accepted: 15.02.2017