BRNO UNIVERSITY OF TECHNOLOGY
VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FACULTY OF INFORMATION TECHNOLOGY
FAKULTA INFORMAČNÍCH TECHNOLOGIÍ

DEPARTMENT OF COMPUTER SYSTEMS
ÚSTAV POČÍTAČOVÝCH SYSTÉMŮ

ACCELERATION OF PHOTOACOUSTIC IMAGING

AKCELERACE FOTOAKUSTICKÉHO SNÍMKOVÁNÍ

MASTER'S THESIS
DIPLOMOVÁ PRÁCE

AUTHOR **Bc. SAVA NEDELJKOVIĆ**
AUTOR PRÁCE

SUPERVISOR **Doc. Ing. JIŘÍ JAROŠ, Ph.D.**
VEDOUCÍ PRÁCE

BRNO 2020

Brno University of Technology
Faculty of Information Technology

Department of Computer Systems (DCSY) Academic year 2019/2020

Master's Thesis Specification

21922

Student: **Nedeljković Sava, Bc.**
Programme: Information Technology Field of study: Information Technology Security
Title: **Acceleration of Photoacoustic Imaging**
Category: Parallel and Distributed Computing

Assignment:

1. Get acquainted with photoacoustic imaging techniques and their physical background. Focus primarily on the back projection techniques.
2. Familiarize yourself with advanced code optimization on GPU-accelerated supercomputers.
3. Design an algorithm for photoacoustic back projection in 2 and 3 dimensions aiming for high performance and efficiency.
4. Implement the designed algorithm.
5. Create a set of benchmark tasks to evaluate the quality of reconstructed images, computational performance and memory consumption.
6. Discuss about the achieved results and their contribution to photoacoustic imaging.

Recommended literature:

- According to supervisor's advice.

Requirements for the semestral defence:

- Items 1 to 4 of the assignment.

Detailed formal requirements can be found at https://www.fit.vut.cz/study/theses/

Supervisor: **Jaroš Jiří, doc. Ing., Ph.D.**
Head of Department: Sekanina Lukáš, prof. Ing., Ph.D.
Beginning of work: November 1, 2019
Submission deadline: June 3, 2020
Approval date: May 25, 2020

Abstract

The goal of this thesis is to provide a new method of image reconstruction out of data generated using Photo-Acoustic imaging. Photo-Acoustic imaging is a very popular biomedical in-vivo imaging modality based on the non-invasive laser-induced generation of ultrasound waves recorded by the acoustic sensors, during which very large amounts of data are generated. The amount of data makes the image reconstruction process very time-consuming. This thesis demonstrates image reconstruction using Back-Projection, an algorithm that is simple enough to be optimized for execution on modern accelerated processor architectures. Two versions of this algorithm are designed: from the perspective of the pixel and from the perspective of the sensor. Both versions are implemented using 3 different execution acceleration methods: vector-level parallelism, thread-level parallelism, and parallelism on the Graphical Processing Unit (GPU). All 3 implementations of both algorithm versions are tested and their results are compared to the much slower but more accurate Time-Reversal reconstruction method. The results have shown that the GPU parallelism implementation offers the fastest execution, which is faster more than 200 times on average compared to the Time-Reversal method. This possibly makes it suitable even for real-time applications.

Abstrakt

Hlavním cílem této práce je navrhnout novu metodu rekonstrukce obrazu z dat fotoakustického snímkování. Fotoakustické snímkování je velmi populární neinvazivní metoda snímkování založená na detekování ultrazvukových vln vyvolaných laserovým paprskem. Proces snímkování generuje velké množství dat, a kvůli tomu je proces rekonstrukce obrazu velmi časově náročný. Táto práce demonstruje proces rekonstrukce obrazu pomocí zpětné projekce, algoritmu který je dostatečně jednoduchý na přizpůsobení moderním architekturám procesorů umožňující různé způsoby optimalizovaného výpočtu. Dvě různé variantu algoritmu byly navrženy: z pohledu pixelu a z pohledu senzoru, který detekuje ultrazvukové vlny. Obě varianty byly implementovány třemi různými způsoby: pomocí vektorového paralelismu, vláknového paralelismu a paralelismu na grafické karetě (GPU). Všechny 3 implementace obou variant algoritmu byly testovány a výsledky byly srovnány s výsledkem rekonstrukce algoritmu reverzního času, přesnějšího ale mnohokrát pomalejšího algoritmu. Výsledky ukázaly, že GPU paralelismus nabízí nejrychlejší výpočet, cca. 200 krát rychlejší než u algoritmu reverzního času, a proto se dá použit i v aplikacích pracující v reálném čase.

Keywords

Photo-Acoustic Imaging, Back-Projection, GPU, CUDA, OpenMP, Vectorization, Multi-threading

Klíčová slova

fotoakustické snímkování, zpětná projekce, GPU, CUDA, OpenMP, vektorizace, vláknový paralelismus

Reference

NEDELJKOVIĆ, Sava. *Acceleration of Photoacoustic Imaging.* Brno, 2020. Master's thesis. Brno University of Technology, Faculty of Information Technology. Supervisor Doc. Ing. Jiří Jaroš, Ph.D.

Rozšířený abstrakt

Fotoakustické snímkování je velmi populární metoda snímkování používaná v biomedicíně. Jedná se o neinvazivní metodu, což znamená, že se dá použit přímo na snímkování živé tkání. Tato metoda je založená na detekování ultrazvukových vln, které vygeneruje snímkovaný objekt. Ultrazvukové vlny jsou vyvolány laserovým paprskem. Laserový světelný paprsek musí mít přesně nastavenou vlnovou délku tak, aby byl propuštěn skrz snímkovanou tkáň ale zároveň absorbován potenciálním objektem uvnitř tkáně, který je cílem detekce. Při absorbování světelného paprsku nastává tzv. fotoakustický jev, tj. vznik akustických (ultrazvukových) vln. Tyto vlny trvají jen krátkou dobu a proto musejí byt detekovaný pomocí ultrazvukových senzorů s krátkou časovou odezvou. Senzory jsou umístění kolem snímkované tkáně. Čím je větší počet senzorů, tím je větší rozlišení výsledného zrekonstruovaného obrazu, ale zároveň je také větší velkost nasnímaných dat. Aby se zrekonstruoval obraz rozumného rozlišení právě je zapotřebí mí velké množství dat, což znamená, že je samotný proces rekonstrukce pomalejší.

Proces fotoakustického snímkovaní vypadá následovně. Nejdříve se snímkovaná tkáň položí do misky s tekutinou, která dobře přenáší ultrazvukové vlny. Kolem misky jsou umístěny senzory. Tkáň je pak osvícená laserovým paprskem s přesně zvolenou vlnovou délku. V snímkované tkání se potenciálně nachází objekt, který předmětem detekce. Tento objekt absorbuje světelný paprsek a zareaguje tak, že se fyzický zdeformuje a vygeneruje ultrazvukové vlny. Tyto vlny jsou nakonec detekovány senzory kolem misky.

Hlavním cílem této práce je navrhnout novu rychlou metodu rekonstrukce obrazu z dat fotoakustického snímkování. Z existujících algoritmů rekonstrukce byl zvolen algoritmus zpětné projekce. Tento algoritmus byl zvolen jelikož se jedná o relativně jednoduchou metodu rekonstrukce, která se dá přizpůsobit moderním akcelerovaným architekturám procesorů. Tento algoritmus spočívá v tom, že se, na základě časového zpoždění detekované vlny, její amplitudě a znalosti rychlostí šíření vlny skrz tkáň, dá vzdálenost bodu vzniku vlny, umístěného na kružnici kolem senzoru. Byly navrženy 2 varianty algoritmu zpětné projekce. První varianta provádí zpětnou projekci z pohledu senzorů, tak že se pro každý senzor zvlášť, na základě jeho nasnímaných dat, vykreslí všechny kružnice kolem jeho. Druhá varianta provádí zpětnou projekci z pohledu pixelů, tak že se pro každý pixel zvlášť vypočítá jeho hodnota na základě vzdálenosti od každého senzoru a nasnímaných dat pro odpovídající vzdálenost.

Tři nejpoužívanější metody akcelerovaného výpočtu v dnešní době jsou vektorový paralelismus a vláknový paralelismus na CPU a paralelismus na GPU. Všechny tři metody byly použity k implementaci navržených variant zpětné projekce. Implementace pro výpočet na CPU jsou určeny k spuštění na procesorech podporovaných kompilátorem společnosti Intel. Implementace pro výpočet na GPU je určená pro grafické karty s technologií CUDA. Vektorový paralelismus spočívá v tom, že je procesor schopen provádět stejnou operaci nad několika operandy současně. Jedná se o tzv. datový paralelismus (podle Flynovy klasifikace patří do kategorie SIMD). Vláknový paralelismus spočívá v tom, že se procesor skládá z více jader. Procesor je pak schopen na každém jádru provádět jednu spuštěnou instanci programu. Pro každou instanci programu může jádro provádět vektorový paralelismus. GPU paralelismus je kombinací předchozích dvou popsaných metod. Grafická karta (GPU) nabízí velký počet samotných jader (mnohem větší než u CPU), ale každé jádro je schopno provádět pouze jednoduší operace (vzhledem k vláknům CPU) efektivně. Na rozdíl od vláken CPU, všechna GPU vlákna by měla provádět stejný kód aby šlo dosáhnout efektivnímu výpočtu.

Výsledky ukázaly, že nejrychlejší výpočet nabízí GPU paralelismus a to pro zpětnou projekci z pohledu pixelů. Tato implementace je cca 200 krát rychlejší ve srovnání s neoptimalizovaným kódem algoritmu reverzního času. Algoritmus zpětné projekce je obecně rychlejší i beze žádné optimalizace za cenu nižší přesnosti (která je ale stale dostačující).

Výsledky této práce považuji za úspěšné, jelikož dosažené zrychlení se zachováním dostačující přesností umožňují, aby se výsledná implementace používala v aplikacích pracující v reálném čase.

Acceleration of Photoacoustic Imaging

Declaration

I hereby declare that this Master's thesis was prepared as an original work by the author under the supervision of Doc. Ing. Jiří Jaroš Ph.D. I have listed all the literary sources, publications and other sources, which were used during the preparation of this thesis.

. . . .
Sava Nedeljković
June 3, 2020

Acknowledgements

I would like to thank Doc. Ing. Jiří Jaroš Ph.D. for his guidance, helpful advices and providing me all the necessary materials and information.

Contents

Chapter 1

Introduction

Photo-Acoustic imaging (sometimes called Opto-Acoustic imaging or Photo-Acoustic tomography) is a new and rapidly expanding biomedical imaging method based on the detection of light-induced ultrasound waves[16]. The main advantage of this method is that it uses non-invasive laser light pulses. These non-destructive properties allow it to be used on living biological structures. It offers a very high optical contrast and high ultrasonic resolution[8]. This means that it can be used for visualizing almost all kinds of biomolecules and displaying them very clearly in high-resolution images. The price of the high-resolution images is the amount of the generated data during the scanning process, which can be in orders of tens or hundreds of GB, and large amounts of data cause the image reconstruction algorithm to be very slow[2]. Even small images could take many hours to develop. The main goal of this thesis is to reduce the computational time of the image reconstruction algorithm and optimizing it for the modern processor acceleration techniques. It would be of great importance for the biomedical field if the reconstruction could be done in near real-time. Because of its high resolution and high contrast capabilities, Photo-Acoustic imaging has a great potential for replacing other invasive imaging methods such as X-ray Computed Tomography, Positron Emission Tomography (PET), or even non-invasive regular ultrasound imaging. Even small improvements may have a great impact on the biomedical industry.

The next chapter (2) describes Photo-Acoustic imaging and the fundamental physical principles behind it in more detail. It also shows several usage examples and comparison with other imaging methods. Lastly, it explains how the whole scanning and imaging process is performed.

Chapter 3 is devoted to the Back-Projection image reconstruction algorithm. It describes how the general algorithm works and how it is modified to work with Photo-Acoustic imaging. Two different version of the algorithm are developed for this purpose.

Chapter 4 explains three possible acceleration techniques that can be used to optimize code for modern processors. It also discusses the positives and negatives of each of them, both in general and in the context of using them for the 2 developed versions of Back-Projection algorithm.

The actual implementation of the programs is explained in the chapter 5. This chapter also gives information on the requirements for compiling and running the programs.

The results and performance of the implemented programs and are discussed and compared in the chapter 6.

The final chapter (7) evaluates the results of this thesis and discusses several ideas for future extension of the programs.

Chapter 2

Photo-Acoustic Imaging

The following chapter describes the Photo-Acoustic imaging and its usage in biomedicine in more detail.

In the scientific literature, Photo-Acoustic imaging can be known by several different names, such as Photo-Acoustic topography, Opto-Acoustic imaging, Ppto-Acoustic topography, etc. Throughout the document, the term Photo-Acoustic imaging or its abbreviated form PAI will be used. Its most general definition is, as its name suggests: visualizing structures based on observing sound waves (*acoustic*) induced by the laser light (*photo*).

This hybrid imaging modality has been broadly studied in biomedicine. In recent years, it has been evolving rapidly, leading to a variety of exciting discoveries and applications, with clinical applications on the way.[9] Several attractive characteristics such as the use of non-ionizing and non-invasive electromagnetic waves, very good resolution and contrast, portable instrumentation, etc. have made Photo-Acoustic useful for imaging cancer cells, wound healing, gene expression and many other both human and animal kinds of tissue[15].

2.1 Photo-Acoustic Effect

More than a century ago, Alexander G. Bell first observed the Photo-Acoustic effect. He found that absorption of electromagnetic waveforms, such as radio-frequency (rf) or optical waves, can generate transient acoustic signals in media. Such absorption leads to local heating and thermoelastic expansion, which can produce megahertz ultrasonic waves in materials.[15] The principle is depicted in Figure 2.1. However, research on the PA effect made little progress for about 80 years after its discovery, primarily due to the lack of appropriate light sources. It was not until the 1970s that Photo-Acoustics regained interest and since then Photo-Acoustics has since been widely implemented in physics, chemistry, biology, engineering, and medicine. Progress was slow until the last decade of the 20th century when many pioneering works demonstrated the PA effect in optically scattering media and biological tissue.[9]

2.2 Working Principle of the Imaging Process

As mentioned previously, recorded ultrasound signals originate from optical absorption. The process of Photo-Acoustic signal generation can be described in three steps: (1) an object absorbs light, (2) the absorbed optical energy is converted into heat and generates a temperature rise, and (3) thermoelastic expansion takes place, resulting in the emission

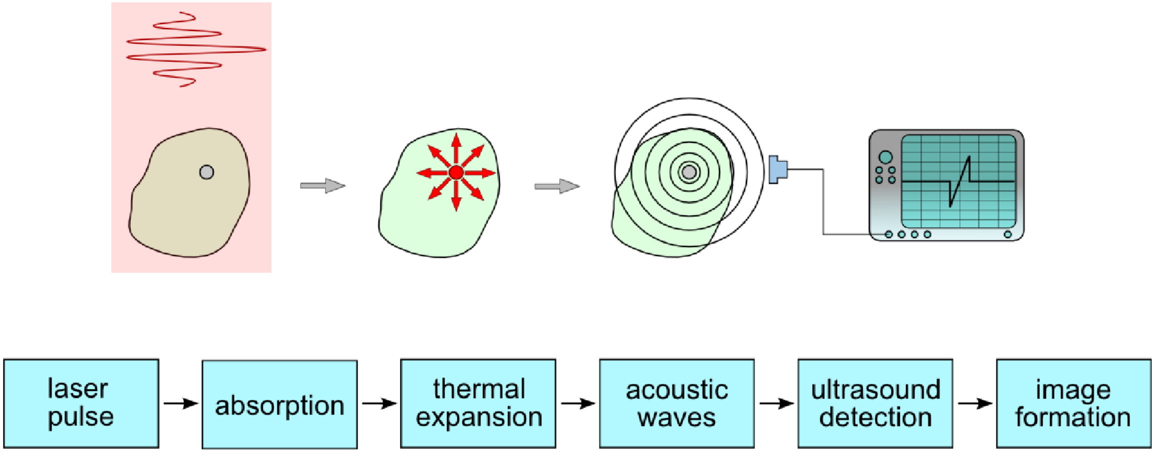

Figure 2.1: Principle of Photo-Acoustic effect (obtained from [12])

of acoustic waves. To generate acoustic waves, the thermal expansion needs to be time-variant. This requirement can be achieved by using a pulsed laser. Pulsed laser excitations are the most widely used because they provide high signal to noise ratio than other methods. For an effective PAI signal generation, the laser pulse duration is normally within several nanoseconds, which is less than both the thermal and stress confinement times. The generated ultrasound pressure signal propagates through the sample and is detected by an ultrasonic transducer or transducer array. The goal of Photo-Acoustic image reconstruction is to recover the distribution of ultrasound pressure from the time-resolved received signals.[14] An example of PAI setup is shown in Figure 2.2.

Since different biological tissues have different absorption coefficient, by measuring the acoustic signals with ultrasonic transducers, one can rebuild the distribution of optical energy deposition and ultimately obtain images of the biological tissues. The combination of high ultrasonic resolution with good image contrast due to optical absorption is quite advantageous for imaging purposes. When compared with optical imaging, in which the scattering in tissues limits the spatial resolution with increasing depth, PAI has higher spatial resolution and deeper imaging depth since the scattering of the ultrasonic signal in tissue is much weaker. When compared with ultrasound imaging, in which the contrast is limited due to the mechanical properties of biological tissues, PAI has better tissue contrast which is related to the optical properties of different tissues. Also, the absence of ionizing radiation makes PAI safer than other imaging techniques such as computed tomography and radionuclide-based imaging techniques.[15]

2.3 Photo-Acoustic Imaging Scenarios

An ideal scenario for Photo-Acoustic imaging would be that light absorption of normal tissue should be low for deeper signal penetration, while the absorption for the object of interest should be high for optimal image contrast.[15] The color of the scanned tissue can be estimated based on the absorption wavelength. If the same object is scanned multiple times each time with pulses of different wavelength fine-tuned to target different inner structures the whole 3D image of the object's inside structure could be obtained.

There are scenarios where the absorption of the object of interest is not high enough. In these cases, a special material called contrast agent could be used. The purpose of such

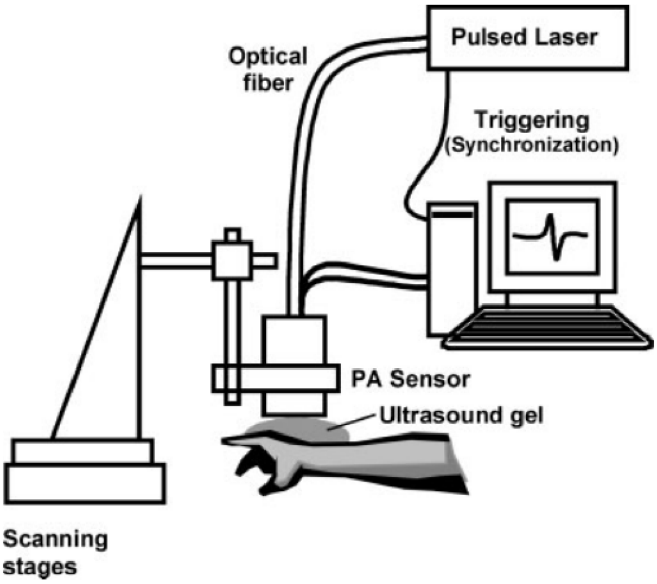

Figure 2.2: An example of PAI setup (obtained from [6])

material is to be bounded to the object of interest. The application of contrast agents can increase the sensitivity and specificity during PAI. Both organic dyes and inorganic nanoparticles are good candidates as Photo-Acoustic contrast agents.[4] Several contrast agents are shown in Figure 2.3.

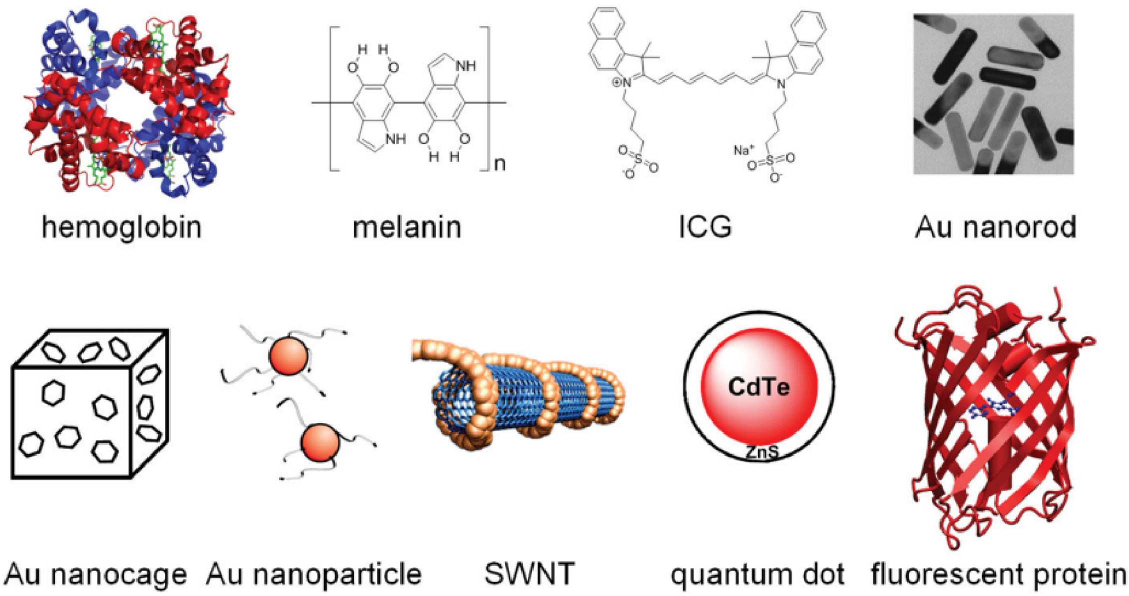

Figure 2.3: A wide variety of contrast agents (obtained from [15])

Chapter 3

Back-Projection Image Reconstruction

After the Photo-Acoustic imaging process, many simulations need to performed on the recorded ultrasound data to recreate the initial pressure distribution and thus visualize the scanned object. During the simulation, some sort of reconstruction method must be used. In this thesis, Back-Projection was chosen as the reconstruction method.

Back-Projection (not to be confused with Back-Propagation) is one of the oldest and most-known reconstruction method. It originates from X-ray Computed Tomography. Even though there are more precise algorithms such as Time-Reversal algorithm, Back-Projection is relatively easy to implement and produces satisfactory results. It consumes the input in one go and directly produces the output, making it suitable for the real-time applications[2].

This chapter explains the original Back-Projection algorithm and how it is adapted for Photo-Acoustic imaging.

3.1 Original Back-projection Algorithm

The well known usage of the Back-Projection algorithm was in X-ray Computed Topography. During the process of X-ray scanning, the only thing that is being recorded is the amount of absorbed radiation. Thus the recorded value is the highest in places where the scanned object cast a shadow on a detector array. Figure 3.1 shows how the absorbance values are recorded for 3 different angles of scanning.

At any point in the recorded projection, the absorbance value is the sum of all absorption along the way. The simplest assumption about this pattern of absorption is that it is uniform along the trajectory. That is is equivalent to assuming that the object is completely homogeneous and that the attenuation arises equally from all points along the way. The summed value is then back-projected into a pixel grid. Using the known values of the projection angle and the distance along the projection, the measured attenuation is divided up equally among the pixels along the measurement beam path, as it is demonstrated in Figure 3.2.[11]

This procedure is then repeated for each obtained projection. This method is sometimes called the summation method, as it sums up all projections at every pixel position. As shown in Figure 3.3, the dense point in the object will now be represented as the intersection of stripes obtained from all projections. After a few projections, the reconstructed layout of the object will have the appearance of a star with the highest value in the center of

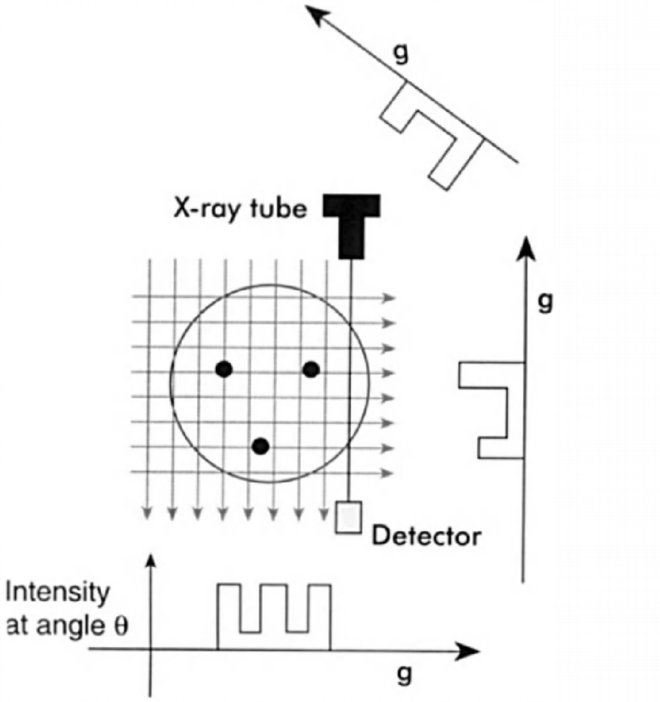

Figure 3.1: Example of X-ray scanning for 3 different angles (obtained from [11])

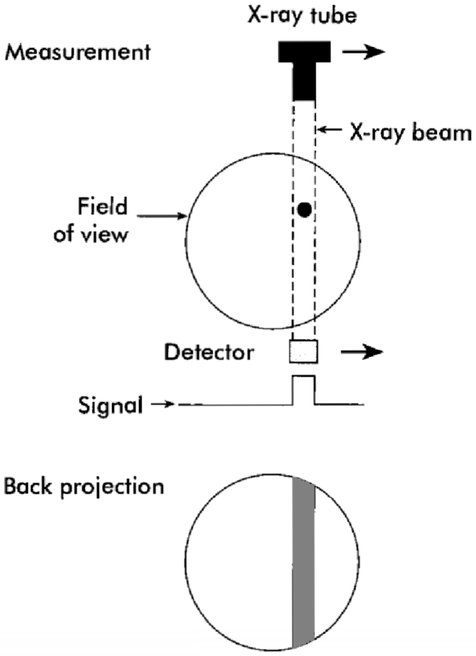

Figure 3.2: Example of a single back-projection. Thick point recorded along a single projection shows up as a strip through the pixel grid. (obtained from [11])

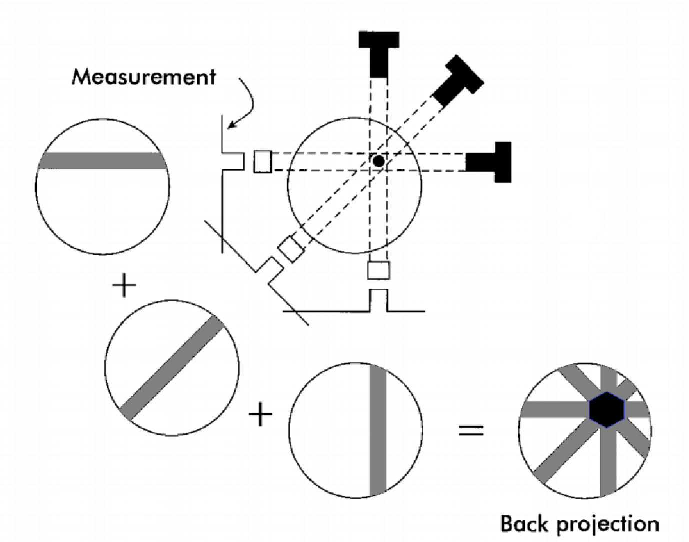

Figure 3.3: Sum of several back-projections. The most dense part is the intersection of all projections. (obtained from [11])

all intersections. After many projections of the back-projected object will have a circular symmetrical shape with noticeable artifacts around it.[11]

Ideally, each point in the actual image would be represented only as a single point in the reconstructed image, but using the back-projection, the point in the output image is blurred by the star pattern. This the biggest drawback of the back-projection and it is the main reason why the alternative filtered back-projection is used more often in X-ray imaging.[11]

3.2 Back-Projection Algorithm Adaptation for Photo-Acoustic Imaging

There are a few key differences between Photo-Acoustic and X-ray scanning. The first is that, as it was described in the previous chapter (2), during the Photo-Acoustic scanning the recorded value is time-varying ultrasound pressure wave. Here, 2 information are known (value of a pressure wave and the time of its recording), whereas in X-ray scanning only single information (the amount of absorbed radiation) is known. The other difference is that the ultrasound wave is spreading in a form of an expanding circle (or a sphere in case of 3D) from its source, whereas X-ray travels in a straight path. The final key difference is that in Photo-Acoustic imaging the source of a signal is the unknown observed object, and in X-ray scanning, there is a separate source and the unknown observed object is the signal absorber. These differences are the reason why the back-projection method needs to be adapted in order to work with Photo-Acoustic imaging.

Back-projection in Photo-Acoustic imaging can be simplified as the triangulation problem. The time and pressure values are known. If the recorded time is multiplied with the speed of sound in an observed tissue, the result is the distance of a wave source from the sensor. Since the recorded value is a wave, it could have come from every direction around the sensor, and thus the source is somewhere on a circle with the radius equal to the previously calculated distance. If there enough sensors are used, the intersection of concentric circles

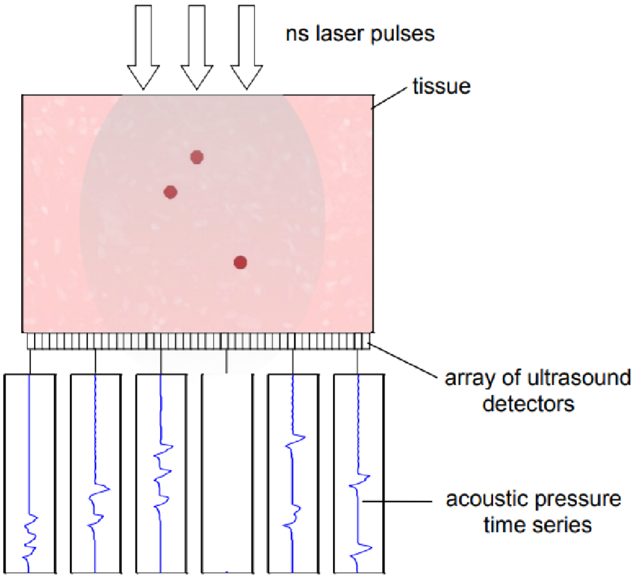

Figure 3.4: Example of signal recording during Photo-Acoustic imaging (obtained from [11])

rasterized by them give an accurate representation of the wave sources. In other words, the recorded wave pressure is Back-Projected to the pixel grid and it is equally spread along all the pixels on a calculated circle around the sensor. The Back-Projection is performed for every sensor on the same grid. The resulting image is a sum of all projected values. The grid is then usually normalized, ie. all the values on it are divided by the number of sensors.[2] Figure 3.4 shows how pressure values are recorded and Figure 3.5 how the back-projection works.

Equation 3.1 for Photo-Acoustic imaging back-projection is: [2]

$$p_{PB}(x) = \frac{1}{N} \sum_{i=1}^{N} \frac{p(\tau_i(x))}{2\pi |x_{0j} - x|^2} \tag{3.1}$$

where x is the current pixel position, N is the number of sensors, $p(\tau_i(x))$ is the recorded pressure value at time delay $\tau_i(x)$, and finally $|x_{0j} - x|^2$ is the calculated distance from the sensor (radius of a circle).

3.3 Proposed Versions of Back-Projection Algorithms

Back-Projection can be performed in two different ways, from the perspective of each sensor and from the perspective of each pixel.

Sensor's Perspective Back-Projection

This version of the algorithm reconstructs the image by rasterizing concentric circles of all possible radius values around the center of the sensors for each sensor. The radius of the circle of each recorded value is determined by the time delay of the recording of that specific value and the speed of sound propagation through the scanned medium. The benefit of

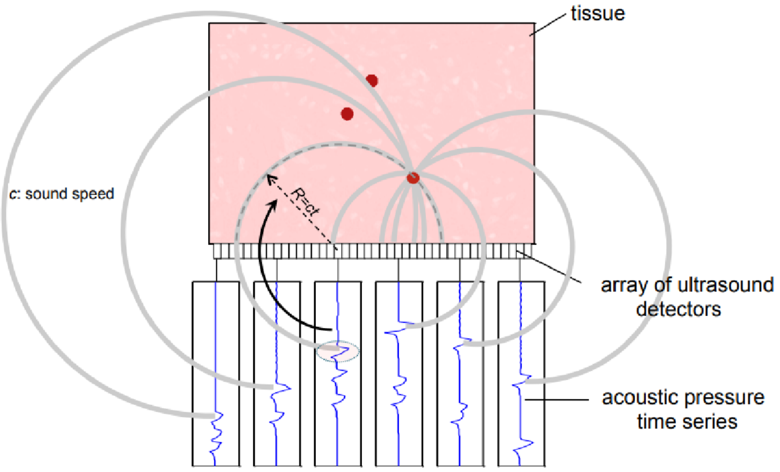

Figure 3.5: Example of Photo-Acoustic imaging back-projection. (obtained from [11])

this version is that there aren't any computationally heavy mathematical operations, the only essential operation is the multiplication of time and speed to calculate the distance (radius). The negative property of this version is the need to perform circle rasterization, which (if implement correctly) has more of a non-deterministic behavior. Accurate circle rasterization is implemented with the loop of unknown number of iterations (while loop) with several selection constructs (if else branching) inside the loop body.

Pixel's Perspective Back-Projection

The second version of the Back-Projection algorithm calculates the value for each pixel separately. It is somewhat the opposite in principle form the first one. As opposed to calculating the values of all pixels around each sensor, it calculates the value of 1 specific pixel by accumulating corresponding recorded values for that pixel from all sensors. The corresponding value is determined by the distance from the pixel to each of the sensors and speed of sound propagation converted to time delay - the opposite compared to the previous version. The positive property of this version is that its behavior is very deterministic, i.e. each pixel performs the exact same operations. The drawback of this version is that the calculation of the distance between the pixel and each sensor - Cartesian distance equation:

$$\sqrt{|x_{pixel} - x_{sensor}|^2 + |y_{pixel} - y_{sensor}|^2}$$

Chapter 4

Code Acceleration Methods

The 3 most commonly used techniques for accelerated execution are vector-level parallelism and thread-level on the CPU and the GPU parallelism. All three methods were used in this thesis for implementing the Back-Projection algorithm.

Vector-level parallelism is based on the fact that the CPU is capable of executing 1 operation on several operands at the same time. It also sometimes referred as data parallelism. According to the Flyn's classifications it belongs to Single Instruction Multiple Data (SIMD) category. Thread-level parallelism is based on the fact that the processor is consisted of several cores, each of which is capable of executing its own instance of running program. The running programs may be executing the same code and the individual cores can perform vector-level parallelism on their own. GPU parallelism is the combination of previous two techniques. GPU constructs of many cores, but each core can only effectively perform relatively simple operation, as opposed to the CPU cores, which are smaller in count but are computationally more powerful. Another difference from the CPU cores is that, if the GPU is supposed to run code effectively in parallel, each of the core must execute the same code. More detailed explanations of these 3 techniques is explained in the following sections.

4.1 CPU Vectorization

Vectorization is the process of converting an algorithm from operating on a single value at a time to operating on a set of values at one time. Modern CPUs provide direct support for vector operations where a single instruction is applied to multiple data (SIMD).[5] SIMD stands for Single Instruction Multiple Data; the same instruction is used on multiple data elements simultaneously. [10].

For example some processors provide registers that are 512 bits wide. One of these registers may be filled with 8 double precision floating point values, or 16 single precision floating point values, or 16 integers. When the register is fully populated a single instruction is applied that operates on 8 to 16 values, like shown in the Figure 4.1.

In order for code to be successfully vectorized, the arrays upon which the operations are performed should be aligned in memory on the value of the width of vector registers. The data can then easily be copied from an array to the vector register. If an array is not properly aligned, time consuming shift operations would need to be performed and will thus reduce the effect of the vectorization.

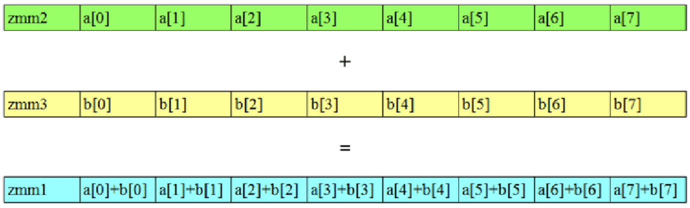

Figure 4.1: Example of SIMD operations on 512 registers (obtained from [10])

4.2 CPU Multi-threaded Parallelism

A thread is a run time entity that is able to independently execute a stream of instructions. The operating system creates a process to execute a program: it will allocate some resources to that process, including pages of memory and registers for holding values of objects. If multiple threads collaborate to execute a program, they will share the resources,including the address space, of the corresponding process. The individual threads need just a few resources of their own: a program counter and an area in memory to save variables that are specific to it (including registers and a stack). Multiple threads may be executed on a single processor or core via context switches; they maybe interleaved via simultaneous multi-threading. Threads running simultaneously on multiple processors or cores may work concurrently to execute a parallel program. Multi-threaded programs can be written in various ways, some of which permit complex interactions between threads.[13]

The main advantage of multi-threading parallelism is that multiple threads are executed separately, thus if one threads in a blocking state and/or waiting for data IO, the other threads can effectively use the hardware resources. One problem which may occur in a multi-threading parallelism is that the threads can interfere with each other while sharing hardware resources.

4.3 GPU CUDA Parallelism

CUDA is a hardware platform for parallel computing created and supported by NVIDIA Corporation to promote access to high-performance parallel computing. The hardware aspect of CUDA involves graphics cards equipped with one or more CUDA-enabled graphics processing units (GPUs).[3] The most essential property of a GPU that enables parallelization is that the device contains not one or several computing units (like a modern multicore CPU) but hundreds or thousands of computing units. CUDA employs the single instruction multiple thread (SIMT) model of paralleliza-tion. CUDA GPUs contain numerous fundamental computing units called cores, and each core includes an arithmetic logic unit (ALU) and a floating-point unit (FPU). Cores are collected into groups called streaming multiprocessors. Executing the same instructions is not just an exercise in redundancy, because each thread performs distinct computations using unique index values that are provided by CUDA [3]. The SIMT approach is scalable because computational throughput can be increased by providing more SMs to share the computing load. Figure 4.2 illustrates the contrast between the architecture of a CPU and a GPU. The CPU has a few computing cores occupying a small portion of the chip and a large area dedicated to control and cache to help those few cores run fast. It is a general rule (and a recurring CUDA theme) that the time required to access data increases with the distance between the computing core and the memory location where the data is stored.[3]

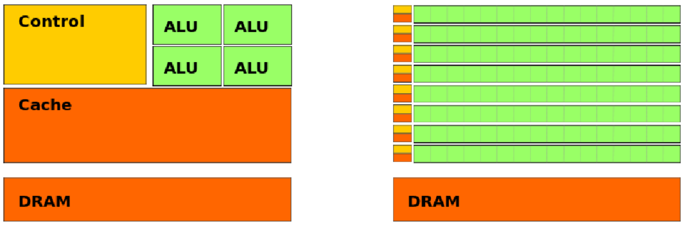

Figure 4.2: Illustration of architecture difference between CPU and GPU. (obtained from [1])

Chapter 5

Implementation

Instead of performing Photo-Acoustic scanning in real life, it was decided for the purposes of this thesis that the simulation could be performed instead. A special script was created, which performs a simulation of Photo-Acoustic scanning within a specified way. This is done by using the k-Wave toolbox. K-Wave is an open-source MATLAB toolbox designed for the time-domain simulation of propagating acoustic waves[7]. Other then simulating PAI scanning, the script also performs a very precise but slow method of image reconstruction based on time-reversal. This reconstructed image will serve as a comparison for the Back-Projection reconstructed image.

There are 3 implementations of the Back-Projection program. Each implementations offers both versions of Back-Projection algorithm.

Vector-level parallelism code is implemented with the help of Intel C/C++ compiler and requires Intel processor in order to run. Multi-threaded parallelism code is implemented as well with the help of Intel C/C++ compiler and requires Intel processor in order to run. It also requires OpenMP library. GPU parallelism is implemented with the help of CUDA Compiler.

Software Requirements

- Matlab
- Intel C/C++ Compiler
- CUDA C/C++ Compiler

Hardware Requirements

- Intel processor
- CUDA GPU

Chapter 6

Results

This chapter briefly describes the results of the implemented programs. The output reconstructed images of all 3 implementations are the same, the only difference is the computational time.

Test Case 1

Execution Times:

- Reference Time-Reversal: 330.46s
- Vectorized Back-Projection: 65.32s
- OpenMP Multi-threaded Back-Projection: 12.75s
- CUDA Back-Projection: 2.46s

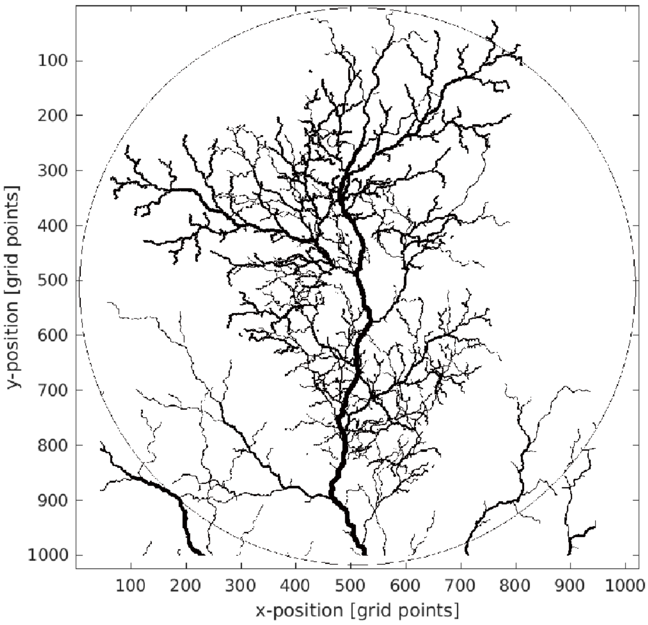

Figure 6.1: Original sensors (circle) and signal sources visualization, 1024x1024 px, 30x30mm

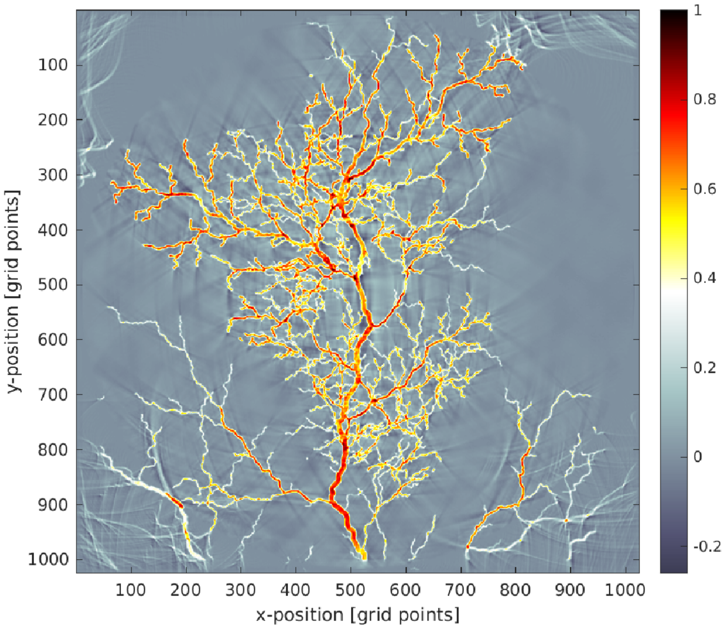

Figure 6.2: Refrence Time-Reversal reconstructed image

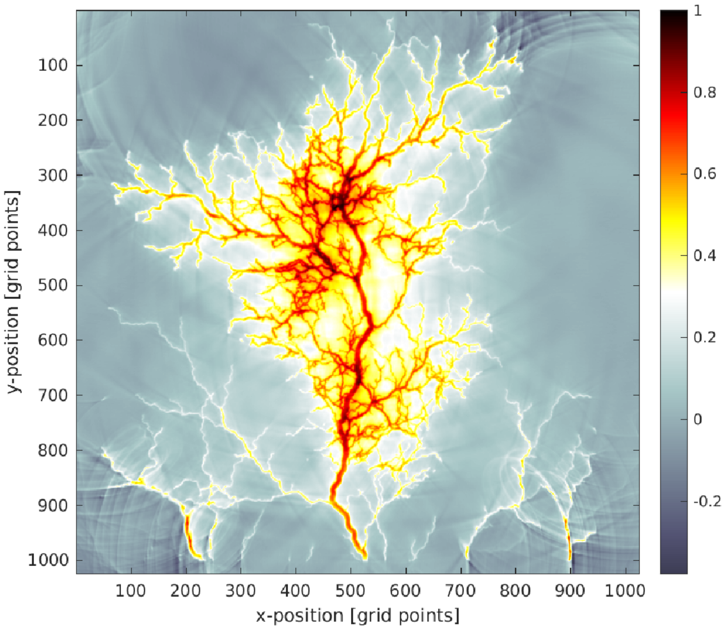

Figure 6.3: Back-Projection reconstructed image

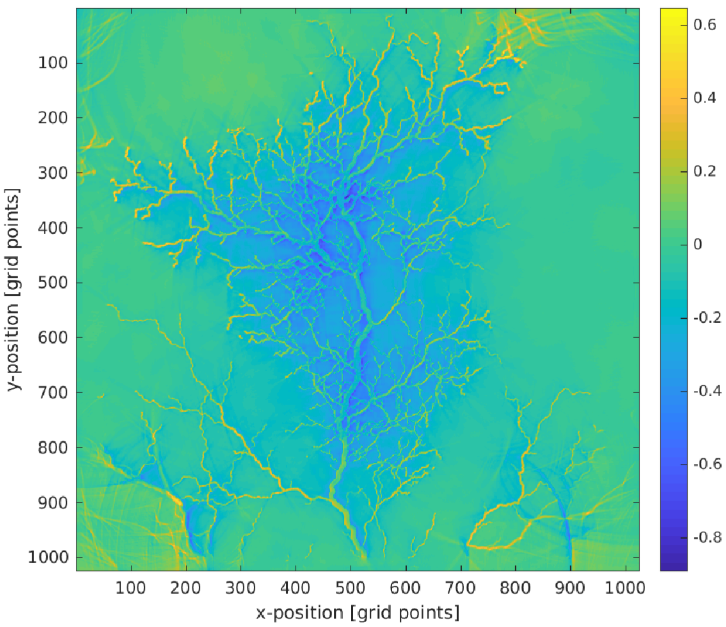

Figure 6.4: Absolute difference

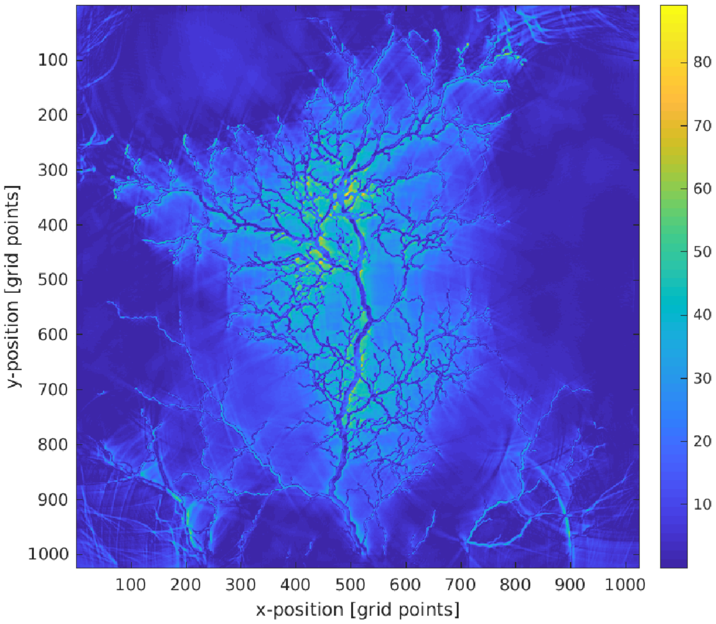

Figure 6.5: Relative difference (in percentage)

Chapter 7

Conclusion

Photo-Acoustic imaging is a very popular imaging modality in the biomedical industry, with the potential of being the dominant one as it has many advantages over its competition like X-ray Computed Tomography (X-ray CT), Magnetic Resonance Imaging (MRI), etc. However, current image reconstruction methods for this imaging modality are very slow (which is one of the setbacks for its progress). The reason for that is the amount of data (generated during imaging) that needs to be processed. The main goal of this thesis was to develop an accelerated implementation of the image reconstruction process. The Back-Projection algorithm was chosen as it is one of the most well-known ones, and it (in its general form) is not very complicated for adaptation for accelerated execution offered by modern processors.

The Back-Projection algorithm is firstly analyzed and adapted to Photo-Acoustic imaging. Two versions of the algorithm specific to the Photo-Acoustic imaging were developed. The first one performs Back-Projection from the perspective of the sensor by rasterizing concentric circles around it according to measured values corresponding to recorded time. The optimization of this algorithm was achieved by making each parallel execution unit performing the task for one sensor. The midpoint algorithm was chosen for circle rasterization. The second version of the algorithm performs the Back-Projection from the perspective of the pixel. It was optimized by making each execution unit performing a task for one pixel. The value of the pixel is computed by first calculating the distance to each sensor, and then taking its data recorded at a specific time corresponding to the calculated distance.

Three different accelerated execution techniques were used to implement both versions of the algorithm: vector-level parallelism, thread-level parallelism, and parallelism on the GPU. The version of Back-Projection from the perspective of the sensor performed worse for all 3 techniques, as it has less deterministic behavior because of the circle rasterization algorithm. The GPU acceleration technique gave the best results. It was more than 200 times faster than the referenced unoptimized Time-Reversal algorithm. All other implementations were much faster than the referenced algorithm.

There is room for future improvements. Only 2D image reconstruction is supported by the current implementation. The programs could be extended to support 3D reconstruction as well. The version of the Back-Projection algorithm from the perspective of the pixel is very easy to adapt in this regard. The only necessary modification is to calculate the 3D Euclidean distance for each voxel (3D pixel) from each sensor. The version from the perspective of the sensor is however harder to adapt. The main reason for this is that instead of rasterizing all possible circles, the program would need to rasterize all possible spheres with the sensor in its center. Rasterization of a sphere could be achieved by rasterizing many circles of various radii separated by 1 pixel in the 3rd dimension.

Another way to extend the current implementation is to modify it in a way, so that program can read live data in real-time. The current implementation of the Back-Projection reconstruction is fast enough, so the computing time will only be bound by the speed of the incoming data. The reconstruction implementation would need to be slightly adjusted, as the current implementation expects that all the data is present (from the whole duration of the Photo-Acoustic imaging) at the time of the beginning of image reconstruction.

Program can also be improved by performing the filtered version of Back-Projection algorithm, which first filters recorded data in the frequency domain and thus removes unwanted noise.

Overall, I consider the goal of the thesis to be accomplished, as all the implementations were very fast while still producing images of acceptable quality.

Bibliography

[1] *CUDA Toolkit Documentation v10.2.89* [online]. [cit. 2020-05-30]. Available at: `https://docs.nvidia.com/cuda/`.

[2] Bordovský, G. *Distributed Photoacoustic Imaging*. Brno, 2019. PhD Thesis Report. Brno University of Technology, Faculty of Information Technology. Supervisor Doc. Ing. Jiří Jaroš, Ph.D.

[3] Duane, S. and Mete, Y. *CUDA for Engineers - An Introduction to High-Performance Parallel Computing*. Addison-Wesly, 2016. ISBN 978-0-13-417741-0.

[4] Jiang, H. *Photoacoustic Tomography*. 1st ed. CRC Press, 2015. ISBN 9781482261042.

[5] Khartchenko, E. *Vectorization: A Key Tool To Improve Performance On Modern CPUs* [online]. January 2018 [cit. 2020-05-30]. Available at: `https://software.intel.com/content/www/us/en/develop/articles/vectorization-a-key-tool-to-improve-performance-on-modern-cpus.html`.

[6] Kolkman, R. G., Bosschaart, N., Kok, B., Leeuwen, T. G. van and Steenbergen, W. Photoacoustic Imaging of Valves in Superficial Veins. *Lasers in Surgery and Medicine*. John Wiley and Sons. 2006, vol. 38, p. 740–744. ISSN 0196-8092,1096-9101.

[7] Kolkman, R. G., Bosschaart, N., Kok, B., Leeuwen, T. G. van and Steenbergen, W. K-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave-fields. *J. Biomed. Opt.* John Wiley and Sons. 2010, vol. 15, p. 021314.

[8] Ku, X., Xie, X., Stoica, G. and Wang, L. V. Imaging of Tumor Angiogenesis in Rat Brains In Vivo by Photoacoustic Tomography. *Applied Optics*. 2005, vol. 44, no. 5, p. 770. ISSN 0003-6935,1539-4522.

[9] Li, C. and Wang, L. V. Photoacoustic Tomography and Sensing in Biomedicine. *Physics in Medicine and Biology*. Institute of Physics. 2009, vol. 54. ISSN 0031-9155,1361-6560.

[10] Mackay, D. *Improve Performance with Vectorization* [online]. September 2016 [cit. 2020-05-30]. Available at: `https://software.intel.com/content/www/us/en/develop/articles/improve-performance-with-vectorization.html`.

[11] Orrison, W. W., Lewine, J., Sanders, J. and Hartshorne, M. F. *Functional Brain Imaging*. 1st ed. Mosby, 1995. ISBN 978-0-8151-6509-5.

[12] Peter, B. *Photoacoustics* [online]. [cit. 2020-01-12]. Available at: `https://www.recendt.at/en/PA.html`.

[13] Schneider, A. and Feussner, H. *Using OpenMPPortable Shared Memory Parallel Programming*. The MIT Press, 2007. ISBN 978-0-262-53302.

[14] Xia, J., Yao, J. and Wang, L. H. V. Photoacoustic Tomography: Principles and Advances. *Progress In Electromagnetics Research*. 2014, vol. 147, p. 1–22. ISSN 1559-8985.

[15] Zhang, Y., Hong, H. and Cai, W. Photoacoustic Imaging. *Cold Spring Harbor protocols*. 2011, 2011,9, no. 9. ISSN 1559-8985.

[16] Zhao, T., Desjardins, A. E., Ourselin, S., Vercauteren, T. and Xia, W. Minimally Invasive Photoacoustic Imaging: Current Status and Future Perspectives. *Photoacoustics*. 2019, vol. 16, p. 100146.