Comparative study of methods of Principal Component Analysis of automatic segmentation of functional magnetic resonance imaging (fMRI).

CPU (Central Processing Unit) -wikipedia

GPU (Graphics Processing Unit) – wikipedia

GPU (Graphic Processor Unit) – CUDA documentation

1. Motivation - how to proces huge volume of data to obtain results in a real-time application
2. In the beginning of the last decade an increase of CPU (Central Processing Unit) clock speed was generally stopped. There are some reasons for that, the essential one is because of the thermal losses.
3. Nowadays in order to maintain continuous increase of the performance processors comprise many cores (multi-cores processor).
4. This implies that a paradigm of sequentially written programs has become unable to fully utilize this architecture. To achieve that it is necesarry to develop parallel applications i.e. applications which exploit all available cores so that the total execution time of the program is lower than the one implemented sequentially using one CPU.
5. In practice there are two main approaches to develop parallel applications. The first one is about processors containing several cores (2,4,6,8,…), each one (processor) processing several „heavy” threads. This approach is common for clusters of CPU processors.

Another type of processors are those which contain many cores (hundreds, thousands) being able to proces many „light” threads. This is how GPU (Graphic Processor Unit) works.

1. Nowadays numerical applications with big computational complexity are implemented mainly on GPU. This is because of some of the features they have. Graphic procesor unit (GPU) is specialized for compute-intensive, highly parallel computation - exactly what graphics rendering is about - and therefore designed such that more transistors are devoted to data processing rather than data caching and flow control [1 cuda programming guide].
2. Objectives

The aim of this work was to implement PCA algorithm on CUDA (Compute Unified Device Architecture) platform in order to obtain the better performance (speedup) regarding to matlab version of the algorithm. The study includes the comparison of the execution time of the methods implemented in Matlab and CUDA. These methods were tested for real problems (on data sets obtaining from fMRI scanning).

1. Explanation of CUDA and Matlab computational platform

Wykład 1 RIM

Model SIMD (Single Instruction, Multiple Data)

Wydajność

przepustowość

For example about flops and bandwidth (wydajność , przepustowość – memory bandwidth) - o tym pod koniec, bo wydajność przepustowość wynika z pewnych cech CUDA, o których mamy napisać