Analysis of algorithms for the exact pattern-matching problem using OpenCL

Andrii Rozumnyi, Dmytro Chasovskyi

University of Tartu

The exact pattern-matching problem which means to find all occurrences of pattern inside a given text, nowadays has many different applications such as parsers, word processors, spam filters, DNA applications in computational molecular biology etc. In some cases, when the string length is relatively small, this can be efficiently solved using classical algorithms with linear time complexity. However, in some areas such as Bioinformatics this task still be a problem as due to huge length of the genomic data (for instance human genome is around three billion of characters) and big amounts of patterns (it can be millions of patterns) processing remains to be time consuming. As the time complexity cannot be faster as linear in such case, there is a need to use another approach for increasing efficiency. Thus, parallel computing is able significantly speed up time for solving exact pattern-matching problem.

As OpenCL allows using wide range of devices in order to do parallel computing, it is a good idea to find such algorithm and configuration of particular devices, which give the best results. In this work we implement the most widely used algorithms for the exact pattern-matching problem and compare them with the same algorithms but adopted for concurrent processing using the power of OpenCL. In addition, performance on different hardware configurations will be measured.

Here is the link `https://github.com/JaakTree/pattern_matching` to main code sources for the current project.