GPU Assignment 02

CSE 560 - GPU Computing

Winter 2019

Maximum Marks: 50

1 Tasks

Pattern matching in textual chunks involves looking for certain keywords in a long stream of text. This could have lots of applications in the humanities and social sciences, and is also possibly relevant to some bioinformatics applications such as gene sequencing. Other applications which require fast pattern matching are antivirus engines, web search engines, text editors etc.

For this assignment a CPU code is provided that implements naive pattern matching based on a fixed number of keywords. In our case, we match pattern in a long text stream with 32 keywords. These keywords have 4 characters each which are held in a single 32-bit unsigned integer so that a single integer comparison checks for a match with a corresponding integer holding 4 characters of the input text. Because the beginning of each word in the text is not necessarily aligned with the beginning of an integer, the CPU code has to also consider 1, 2 and 3-byte offsets.

At the end of matching all the keywords, you need to report frequency of all the keywords in the given text. You are given 4 text files of different sizes in "data/ " folder and this folder also contains the 32 keywords. Understand the given CPU code to generate the keyword frequencies from the given texts in "data/ " folder.

2 Deliverables

As part of this assignment you are provided with the serial implementation, you are required to perform the following:

1. Generate the keyword frequency list (as depicted by CPU code) on the GPU. [25 marks]

2. Generate timing data for your code (GPU time (kernel and kernel + memory transfer) vs. CPU time) against the different data files. [18]

3. Write a report containing details of the strategy and speedups received. [7 marks]

Note: Ensure that the output of the GPU is identical to the CPU by implementing a function to check the equality of the results.

3 Upload Instructions

Upload all the files as a .zip file with the following naming convention - assignment01_⟨roll_no⟩.zip

1. Code for the assignment along with a Makefile
2. A report in PDF format written with Latex/MS Word. Use the acmlarge option (single column) (see sample-acmlarge.tex if writing with Latex) refer here

Note: Your code should be written by you and be easy to read. You are NOT permitted to use any code that is not written by you. (Any code provided by the TA can be used with proper credits within your program)