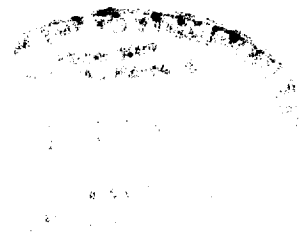
EXACT PARALLEL PATTERN MATCHING ALGORITHMS ON NVIDIA CUDA GRAPHICS PROCESSORS

A Thesis

Submitted in Partial Fulfilment of the Requirements for the Degree of
Master of Philosophy in Computer Science

of
The University of the West Indies



Avery Joel Stephen
2011

Department of Computer Science, Maths & Physics
Faculty of Pure & Applied Science
Cave Hill Campus

Abstract

Exact Parallel Pattern Matching on NVIDIA CUDA Graphics Processors

Avery Stephen

Exact pattern matching is one of many inherently sequential algorithms, whose performance increased with successive Central Processor Unit (CPU) generations. However, today's mainstream CPUs are multiple-core parallel processors that decrease the performance of serial algorithms without explicit parallelism. This happens because serial algorithms only use one processor core, and single-core CPUs generally operate at lower clock frequencies, as compared to multiple-core CPUs. Therefore, we explicitly redesigned some matching algorithms for parallel processing.

Instead of using the multiple-core CPU that has a predicted performance plateau at eight cores, we used the graphics processor's ability to process general-purpose instructions using NVIDIA CUDA (Compute Unified Device Architecture). We parallelised the Brute-Force, Shift-Or, Horspool and SunKim1 algorithms using a data parallelism technique. This involved splitting the text into segments of a fixed maximum size having an overlap of characters and searching each segment in parallel on the GPU.

We compared algorithm performance on two types of match results. To find the number of matches in a text, we used an algorithm based on the CUDA Atomic Add function and a Parallel Count algorithm. To find the positions of matches in the text, we used a similar algorithm based on the CUDA Atomic Add function and our modified Parallel Stream Compact algorithm.

Using simulations with ten matches in forty megabytes of text, we observed algorithm speedup over single-core CPU on both types of match results.

Keywords: Parallel Pattern Matching, Parallel Searching, NVIDIA CUDA, Prefix Sum, Stream Compact, Graphics Processor