Cuda lab 1

Dries Van de Velde, Behzad Iqbal

February 2024

1 Introduction

The goal of this project is to compare the speed of GPU and CPU on a basic task like reversing an array. This is realised using c++ using the CUDA cores of the NVIDIA's graphics card.

2 Coding difficulties

When it comes to the thread matching, we currently use 256 threads per block. And use just as many blocks needed to handle all items. In the CUDA function we then recalculate the index by multiplying the block size with the block index and adding the thread index to it.

The operation is only executed when the calculated index is smaller then the amount of items which need to be handled. If the index is larger it is neglected since this would otherwise go outside of the used array memory space.

Another difficulty we encountered was the memory allocation on the GPU. This is not handled automatically, we thus needed to allocate the CUDA function parameters in the GPU memory before calling the CUDA function.

3 Conclusion

For small size integer arrays, the CPU is faster compared to the GPU. However as the array length gets larger the benefit of using the GPU becomes more and more notable.

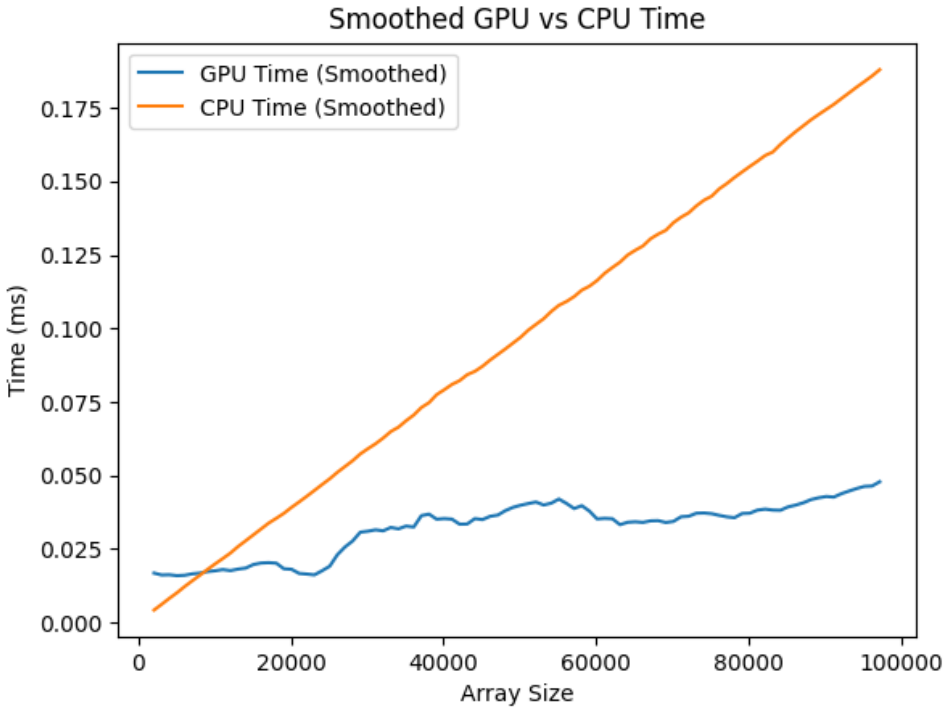

Figure 1: Execution times CPU vs GPU