|  |  |  |
| --- | --- | --- |
| Property | Value | Description |
| Device | 0 | The index of device which has compute capability greater or equal to 1.0 that are available for execution |
| name | GeForce GTX 480 | ASCII string identifying device |
| Compute capability | 2.0 | Describes the features supported by a CUDA hardware. [Major compute capability] dot [Minor compute capability] |
| total global memory (KB) | 1509504 | Global memory available on device in kilo bytes |
| shared mem per block | 49152 | Shared memory available per block in bytes |
| regs per block | 32768 | 32-bit registers available per block |
| warp size | 32 | Warp size in threads |
| max threads per block | 1024 | Maximum number of threads per block |
| max thread dim | z:1024  y:1024  x:64 | Maximum size of each dimension (x, y, z) of a block |
| max grid size | z:65535  y:65535  x:65535 | Maximum size of each dimension (x, y, z) of a grid |
| clock rate (kHz) |  | Clock frequency in kilohertz |
| total constant memory (bytes) | 65536 | Constant memory available on device in bytes |
| multiprocessor count | 15 | Number of multiprocessors on device |
| memory bus width | 384 | Global memory bus width in bits |
| memory clock rate (kHz) | 1848000 | Peak memory clock frequency in kilohertz |
| L2 cache size (bytes) | 786432 | Size of L2 cache in bytes |
| max threads per SM | 1536 | Maximum size of each dimension (x, y, z) of a block |

|  |  |  |
| --- | --- | --- |
| Size of the vector(n) | GPU time/s | CPU time/s |
| 100 | 0.000023 | 0.000005 |
| 10000 | 0.000018 | 0.000038 |
| 1000000 | 0.000288 | 0.003214 |
| 10000000 | 0.001225 | 0.031610 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Case # | Case configuration | | GPU time/s | CPU time/s | Speedup () |
| Image size | Window size |
| 1 | 640640 | 33 | 0.000630 | 0.030308 | 48.123214 |
| 2 | 640640 | 55 | 0.001117 | 0.071384 | 63.895095 |
| 3 | 12801280 | 33 | 0.002675 | 0.116402 | 43.518095 |
| 4 | 12801280 | 55 | 0.004591 | 0.281064 | 61.221895 |

* + 1. When is the speedup (i.e. time of CPU version / time of CUDA version) at its lowest? And why?

Image size 12801280 and window size 33.

Cases 3, 4 has lower speedup compared to cases 1, 2. That is expected, since image size is large and device utilization should be less in cases with larger image size. When comparing case 3 and case 4, 33 kernel work has a lesser serial work that is compared to 55 kernel. It is expected to case 3 has a higher speedup than in case 4. Reason should be that 33 kernel computation has a related considerable overhead that dominates its computation time.

* + 1. When is the speedup at its highest? And why?

Image size 640640 and window size 55.

Similar explanation can be given as in i.