

Reporte técnico

César Armando Valladares Martínez

Agosto 2018

Resumen

El siguiente documento es un reporte de la primera tarea de la materia de programación multinúcleo donde se realiza una comparación entre el tiempo transcurrido de una operación de matrices, la comparación se realiza entre usando una CPU, una CPU con threads y una GPU modificando la cantidad de Threads y Blocks.

Introducción

La programación paralela se utiliza actualmente para realizar procesos de manera mas rápida utilizando GPU en lugar de CPU, la ventaja de las GPU es que pueden realizar procesos de manera simultánea, es decir en paralelo, utilizando los hilos (Threads) que se dividen por bloques, específicamente hablando de las GPU de NVIDIA.

Desarrollo

Para estas pruebas y medir los tiempos de cada algoritmo se realizaron unas 30 repeticiones de cada uno, así obteniendo un promedio del tiempo transcurrido.

Los resultados de estas pruebas se pueden apreciar en la siguiente tabla:

30 repeticiones	1000 y , 1000 x
CPU	9172.441162 ms
CPUopenmp	5334.867399 ms
GPU 16/16	149 ms
GPU 16/32	148 ms

En la segunda tabla vemos los speed up's es decir, la diferencia de tiempo entre los algoritmos y donde se ejecutaron, el speed up se mide dividiendo el tiempo del secuencial entre el tiempo del paralelo.

1	CPU	CPUopenmp	1.72
2	CPU	GPU 16/16	61.56
3	CPU	GPU 16/32	62

Conclusión

Con base en los anteriores resultados podemos concluir que es mucho mas eficiente ejecutar los algoritmos en la GPU en lugar de la CPU, con la razón principal de que podemos usar muchos mas hilos que en un procesador, la cantidad de hilos posibles en un procesador es igual al numero de núcleos que éste tenga