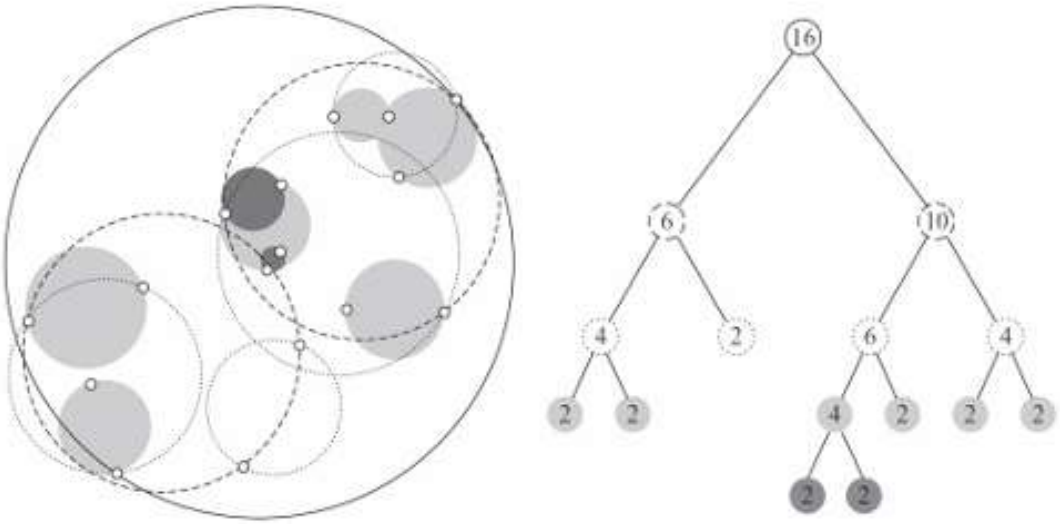
Vantage Point Tree

THMMY , 7th semester , Parallel and Distribute Systems:1st assignment

Authors :

Apostolos Moustaklis
Stavros Portokalidis

github link : https://github.com/tolism/Vantage-Point-Tree



The goal of the first assignment was to create the vantage point tree sequentially and afterwards using the libraries pthread , openmp and cilk parallelize the calculation of the distances and the computation of the two subtrees. Thresholding was crucial in order to achieve better perfomance due to the fact that assigning too little work to be done in parallel , slows the program.

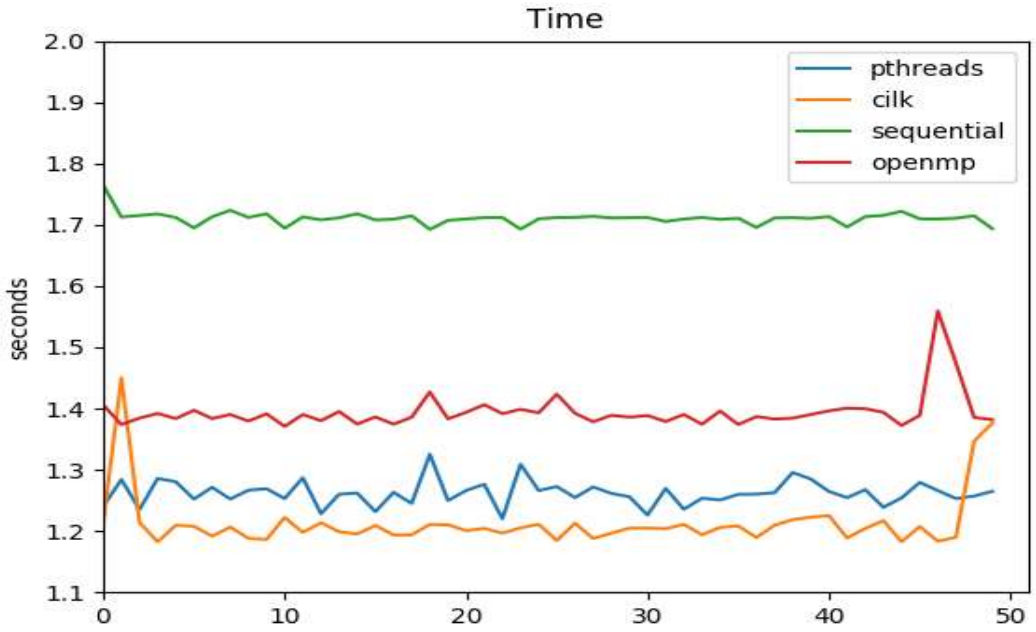
After many tests , cilk implementation was the fastest , followed by pthreads . As we increased the number of points and dimensions , we got better and better results , in comparison to the sequential implementation .However , due to the lack of memory we were unable to see extreme speed ups. In the following graph , you can see the 4 performances of the implementations at 50 tests and the average execution time.

As for thersholding , we found out that it is better to calculate the distance sequentially when the number of points is less than 250000. When it comes for great number of points such as more than 1million , we should calculate the distance in parallel at a node when it's n is less than N/4 , where N stands for the original number of points. In other words , when our tree , reaches the level 2 (level 0 is root) , it's better to go serial.

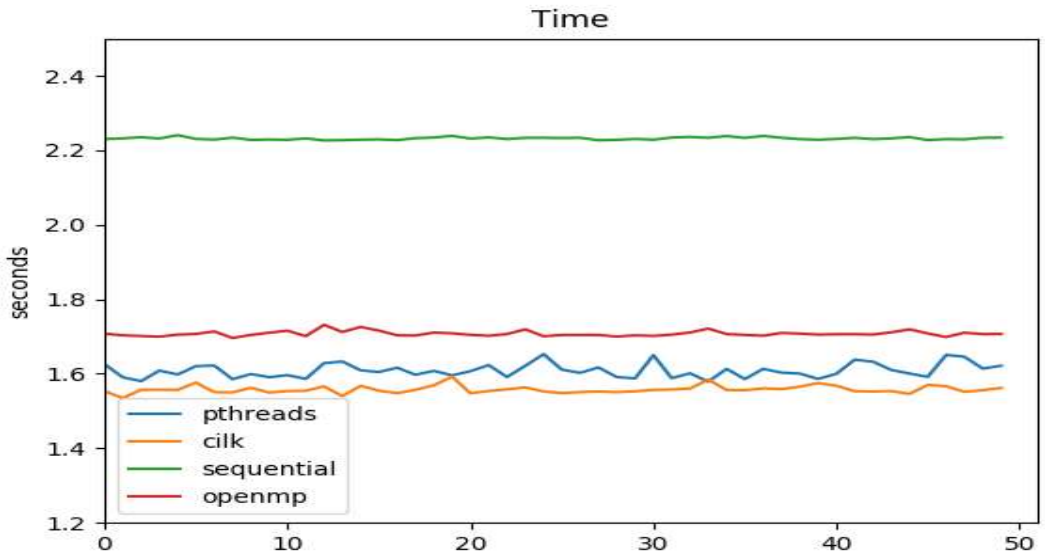
Hardware used : Intel core i7 7th gen, 8 GB RAM

Number of points **N** of the original set : 1000000

Number of dimensions **D** :20



Number of points **N** of the original set : 1000000
Number of dimensions **D** :30



Number of points **N** of the original set : 100000

Number of dimensions **D** :10