Speeding up Generalized Fuzzy *k*-Means Clustering Algorithm by GPUs

Dinh Trung Vu

Department of Computer Science and Engineering

National Taiwan Ocean University

Keelung, Taiwan 202

R. O. C.

ABSTRACT

The graphics hardware is becoming increasingly more powerful and programmable with the introduction of Graphics Processing Units (GPU) like the NVidia GeForce series. The GPU’s exceed the ordinary general purpose CPU’s ability to do ﬂoating point operations due to the massively parallel architecture in the GPU’s.

With the newest GPU’s one actually have enough programmable freedom to do other computations than computer graphics processing. This project will take advantage of this in order to get high performance implementations of image analysis algorithms.

In this project we will implement an image analysis algorithm, which is Generalized Fuzzy k-Means Clustering Using m nearest Cluster Centers (GFKM) [1], on a GPU. We will also make comparisons with CPU based implementations and analysis the pros and cons of using GPU’s in image analysis.

PLAN

1. Learn the fundamentals of parallel computing with the GPU and the CUDA (Compute Unified Device Architecture) programming environment.
2. Design a GPU-based parallel GFKM algorithm:

* Review GFKM algorithm.
* Analysis steps of the algorithm can be parallel implemented on GPUs: (1) Finding closest centroid, (2) Computing new centroids.
* Design GPU-based parallel algorithm for each section:
* Step (1): For low-dimensional data sets, we utilize the GPU on-chip registers to minimize the latency of data access; for high-dimensional data sets, we use both registers and shared memory and apply a very efficient reduction algorithm that treats the most time-consuming part of k-Mean as matrix multiplication [2].
* Step (2): It is difficult to be fully parallelized due to write conflict, so we use GPU to speed up part of the task that is appropriate for parallelization, and leave the remaining part for CPU execute. We will design an algorithm which adopts “divide and conquer” strategy [2].

1. Implement GPU-based parallel GFKM algorithm and make comparisons with CPU based implementations.

REFERENCES

[1] Franklin J. C. Lai, Eric Y. T. Juan, and Jim Z. C. Lai, Generalized Fuzzy k-Means Clustering Using m nearest Cluster Centers, 2013.

[2] You Li, Kaiyong Zhao, Xiaowen Chu, and Jiming Liu, Speeding up K-Means Algorithm by GPUs, 2010.

**Algorithm 1:** CPU-based GFKM

//

//

//



**Algorithm 2:** Computing initial and initializing based on CPU

//

//

//

//



**Algorithm 3:** Computing initial and initializing based on GPU

//

//

//

//

//

//

//

//

//

//

//



**Algorithm 4:** Computing distance based on shared memory of the GPU

//

//

//

//

//

//

//

//

//

//

//

//

//

//



**Algorithm 5:** Computing distance and updating membership and based on CPU

//

//

//

//



**Algorithm 6:** Computing distance and updating membership and based on GPU

//

//

//

//

//

//

//

//

//

//

//



**Algorithm 7:** Computing distance based on shared memory of the GPU

//

//

//

//

//

//

//

//

//

//

//

//

//

//



**Algorithm 8:** Computing the new center for each clusters based on CPU

//

//

//

//

//



**Algorithm 9:** Computing the new center for each clusters based on GPU