Hybrid Implementation of Expectation Maximization Algorithm using CUDA and OpenMP for Gaussian Mixture Models

Ayush Raina

Supercomputer Education and Research Centre

Indian Institute of Science, Bangalore, India

ayushraina@iisc.ac.in

Abstract—The Expectation-Maximization (EM) algorithm for Gaussian Mixture Models (GMMs) is computationally intensive, especially for large datasets with high dimensionality. This paper presents a hybrid parallel implementation that leverages both a single GPU (using CUDA) and multicore CPU (using OpenMP) to accelerate the algorithm. Our approach partitions the workload by mapping the computationally intensive E-step to the GPU's massive parallelism while utilizing OpenMP for the parameter update M-step on the CPU. We analyze the computational characteristics of both steps and design optimized CUDA kernels for probability density computations and responsibility calculations, along with efficient OpenMP parallelization strategies for parameter updates. Experimental results demonstrate that our hybrid approach achieves significant speedup compared to serial implementations, with performance gains increasing with dataset size and model complexity. We also provide a comparative analysis of pure GPU, pure multicore CPU, and hybrid implementations, showing that our hybrid approach effectively balances computation and data movement to achieve superior performance on modern heterogeneous computing platforms.

I. INTRODUCTION

This is for introduction

II. RELATED WORK

This is for literature survey. e.g. The paper by Ellsworth et al. [1] discusses...The paper by Yu et al. [2] had proposed..

III. METHODOLOGY

The third section will contain the overall methodology. You may choose to have a specific title for the section instead of "Methodology". For example, "A Divide-and-Conquer Algorithm for QR Factorization". You may also organize methodology section as multiple sections with different titles, if necessary, instead of a single section.

Try to arrange the sections into multiple subsections with clear organization, using \subsection{}.

Have figures and tables. Search in google for latex help on these.

IV. EXPERIMENTS AND RESULTS

A. Experiment Setup

B. Results

V. CONCLUSIONS

Have one para for conclusions and one para for future work.

REFERENCES

[1] D. Ellsworth, B. Green, C. Henze, P. Moran, and T. Sandstrom, "Concurrent Visualization in a Production Supercomputing Environment," *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 5, pp. 997–1004, 2006.

[2] H. Yu, K.-L. Ma, and J. Welling, "A Parallel Visualization Pipeline for Terascale Earthquake Simulations," in *SC '04: Proceedings of the 2004 ACM/IEEE conference on Supercomputing*, 2004, p. 49.