HETEROGENEOUS BIG-DATA CLUSTER FOR COMPUTER VISION APPLICATIONS

By

Hazem Abdelmegeed Elsayed Abdelhafez

A Thesis Submitted to the
Faculty of Engineering at Cairo University
in Partial Fulfillment of the
Requirements for the Degree of
MASTER OF SCIENCE
in
Electronics and Communications Engineering

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2016

HETEROGENEOUS BIG-DATA CLUSTER FOR COMPUTER VISION APPLICATIONS

By
Hazem Abdelmegeed Elsayed Abdelhafez

A Thesis Submitted to the
Faculty of Engineering at Cairo University
in Partial Fulfillment of the
Requirements for the Degree of
MASTER OF SCIENCE
in
Electronics and Communications Engineering

Under the Supervision of

Prof. Dr. Hossam Ali Hassan Fahmy

………………………………..

Professor
Electronics and Communications
Engineering Department
Faculty of Engineering, Cairo
University

Prof. Dr. Ameen Mohamed Nassar

………………………………..

Professor
Electronics and Communications
Engineering Department
Faculty of Engineering, Cairo
University

Dr. Mohamed Mohamed Rehan

……………………………………

Chief Technical Officer
AvidBeam

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2016

HETEROGENEOUS BIG-DATA CLUSTER FOR COMPUTER VISION APPLICATIONS

By
Hazem Abdelmegeed Elsayed Abdelhafez

A Thesis Submitted to the
Faculty of Engineering at Cairo University
in Partial Fulfillment of the
Requirements for the Degree of
MASTER OF SCIENCE
in
Electronics and Communications Engineering

Approved by the
Examining Committee

____
Prof. Dr. Hossam Ali Hassan Fahmy, Thesis Main Advisor

____
Prof. Dr. Ameen Mohamed Nassar, Member

____
Dr. Mohamed Mohamed Rehan, Member, Chief Technical Officer, AvidBeam

____
Prof. Dr Elsayed Eissa Abdo Hemayed, Internal Examiner

____
Prof. Dr. Khaled Mostafa Elsayed, External Examiner, Faculty of Computers and Information, Cairo University

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2016

Engineer's Name: Hazem Abdelmegeed Elsayed Abdelhafez
Date of Birth: 26/09/1990
Nationality: Egyptian
E-mail: hazem.abdelmegeed@gmail.com
Phone: +2 01001128853
Address: 4, Street 100, Maadi, Cairo, Egypt
Registration Date: 01/03/2014
Awarding Date: / /2016
Degree: Master of Science
Department: Electronics and Communications Engineering

Supervisors:

Prof. Dr. Hossam Ali Hassan Fahmy
Prof. Dr. Ameen Mohamed Nassar
Dr. Mohamed Mohamed Rehan

Examiners:

Prof. Dr. Hossam Ali Hassan Fahmy
Prof. Dr. Ameen Mohamed Nassar
Dr. Mohamed Mohamed Rehan, Chief Technology Officer, AvidBeam
Prof. Dr. Elsayed Eissa Abdo Hemayed
Prof. Dr. Khaled Mostafa Elsayed, Faculty of Computers and Information, Cairo University

Title of Thesis:

HETEROGENEOUS BIG-DATA CLUSTER FOR COMPUTER VISION APPLICATIONS

Key Words:
Big-data; Parallel Computing; Graphical Processing Units; Computer Vision; Heterogeneous Computing

Summary:

A token-based scheduler is developed to enable efficient utilization of graphics processing unit in big-data clusters specifically for computer vision applications. The scheduler addresses the racing conditions that occur on the graphics processing unit due to simultaneous access by the parallel instances of the running application. The presented scheduler enables the porting of computer vision applications to big-data cluster with heterogeneous computing capabilities where multi-core central processing units exist alongside graphical processing unit.

Acknowledgments

I would like to thank my mother and my family for their support and understanding during the time working on the MScs, without their support and love I wouldn't have crossed this far in my academic life and career. I would like to thank Dr. Hossam and Dr. Rehan for their endless support, patience, sincerity and dedication during my journey in this thesis. I would like also to thank all my colleagues at Avidbeam, Mohamed Fouad, Ahmed Sobhi and Ahmed Abdelsamad for helping me prepare the hardware setup I used in my thesis. Also I would like to thank Menna Ghoneim, Dina Helal and Rana Morsi for helping me understand computer vision algorithms better and for reviewing my thesis to make sure it is in perfect shape. I would like also to thank the management at Avidbeam, Dr. Hani El-Gebaly and Hossam Samy for providing all the help I needed and supporting me during my MScs journey. Finally, I am also thankful for the free, Open-Source community that provided freely most of the tools I used in this work and we gratefully acknowledge the support of NVIDIA Corporation with the donation of the Tesla K40 GPU used for this research.

Dedication

Dedicated to my mother and my family...

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AdaBoost	Adaptive Boost
AM	Application Master
APIs	Application Programming Interfaces
ASF	Apache Software Foundation
AVX	Advanced Vector Extension
BSD	Berkeley Software Distribution
Cgroups	Linux Control Groups
CUDA	Compute Unified Device Architecture
CUDASA	Computer Unified Device and Systems Architecture
CUs	Compute Units
DAGs	Directed Acyclic Graph
DSL	Domain-Specific Language
DSP	Digital Signal Processing
FPGA	Field Programmable Gate Array
FPS	Frames per Second
GFS	Google File System
GPGPU	General Purpose Computing on GPU
IGP	Integrated Graphics Processor
IPP	Intel Performance Primitives
JAR	Java Archive
JSON	JavaScript Object Notation
JVMs	Java Virtual Machines
LPR	License Plate Recognition
MPS	Nvidia's Multi-process Service
ms	milliseconds
NM	Node Manager
OCR	Optical Character Recognition
OpenACC	Open Accelerators
OpenCL	Open Computing Language
OpenCV	Open Computer Vision
OpenGL	Open Graphics Library
PCI	Peripheral Component Interconnect
PEs	Processing Elements
PMGMR	Pipelined Mutli-GPU Map-Reduce
RDD	Resilient Distributed Dataset
RM	Resource Manager
SDK	Software Development Kit

SIMD	Single Instruction Multiple Data
SM	Streaming Multiprocessor
SSE	Streaming SIMD Extensions
SSSE3	Supplemental Streaming SIMD Extensions
TOM	Task-oriented Modules
VPE	Virtual Processing Elements
YARN	Yet Another Resource Negotiator

Abstract

Big-data technology in recent years has become increasingly utilized to process huge amount of data in a timely manner. The growth in the amount of visual data - videos and images - generated nowadays raises the need for porting computer vision applications to big-data frameworks in order to increase the processing throughput of these applications.

On the other hand, developers and the scientific community have already ported many computer vision algorithms to the Graphics Processing Unit (GPU) that successfully accelerates these algorithms thanks to its data-parallel architecture.

Combining big-data with GPUs to scale computer vision applications both horizontally and vertically yields a promising architecture for processing the huge amounts of visual data and to fulfill the urging need to mine these data for underlying patterns and information. Unfortunately, the number of GPUs available on a typical processing node in a big-data cluster is limited and most of the time there is only one GPU on such nodes. Therefore, multiple instances of the same computer vision application running on top of any big-data framework leads to competition between these instances on the scarce GPU resource.

In this thesis, we address the challenge of combining GPUs with big-data technology in order to accelerate the processing of computer vision applications. We introduce a GPU scheduler that is responsible for assigning the GPU to multiple instances of the computer vision application efficiently with minimal competition and best performance compared to using either the GPU or the Central Processing Unit (CPU) solely. In order to achieve this we propose a token based scheduler that guarantees that no competition occurs on the GPU. The evaluation shows increased processing throughput up to 2.3x compared to CPU-only big-data processing with 24 cores, 2.1x compared to CPU-GPU big-data processing and up to 32x compared to the sequential processing on a single CPU core.