Computer Science Department
Faculty of Computer & Information Sciences
Ain Shams University

Performance Analysis of GPU-Accelerated Augmented Reality Marker-less real-time Object Detection and Tracking

A thesis submitted as a partial fulfillment of the requirements for the degree of Master of Science in Computer and Information Sciences.

By

Yara Medhat Mohamed Abdelaal

B.Sc. in Computer & Information Sciences
Demonstrator at Computer Systems Department
Faculty of Computer and Information Sciences, Ain Shams University

Under supervision of

Prof. Dr. Hossam Faheem

Head of the Computer Systems Department
Faculty of Computer and Information Sciences, Ain Shams University

Assoc. Prof. Dr. Samy Ghoneimy

British university in Egypt

Cairo, 2014

Acknowledgement

First and foremost, I would like to thank God the almighty, for giving me the strength and courage to do this work. I would also like to thank my research supervisors, Dr Sami Ghoneimy for all his help and continuous support and Professor Hossam Faheem, not only a supportive professor but also a father. Without their assistance and dedicated involvement in every step throughout the process, this thesis would have never been accomplished. I would like to thank you very much for your support and understanding over these past years. I would also like to thank Dr Mahmoud Fayez and Dr Heba Khaled, Salah Nahl, Mariam Asaad and Nada Sherif, who helped me all along the thesis, and where of constant support.

Most importantly, none of this could have happened without my family. My grandmother, god bless her soul. To my parents who have always supported me since I was a little kid. A wonderful mother always pushing me forward, and lifting me up when I am down. A father who cares about my future, and a brother; a brother any girl would wish to have, a loving and supporting brother, who supported me despite his commitments and the fact that he is across the oceans – it would be an understatement to say that, as a family, we experienced some difficulties in the past years. Every time I was ready to quit, you did not let me and I am forever grateful. This dissertation stands as a testament to your unconditional love and encouragement.

Abstract

Augmented reality (AR) is an increasingly recognized paradigm of information visualization that merges a computer-generated virtual object onto a real scene to visually augment reality. With recent and continued increases in computing power, and advances in the field of computer graphics, realistic augmented reality environments can now offer inexpensive and powerful solutions in a whole range of training, simulation and leisure applications. One key challenge to maintaining convincing augmentation, and therefore tracking real objects from natural features, is decoupling the rendering process of the augmented information from the displaying system, by which we can augment the scene using computationally expensive rendering techniques that requires high computational power.

As such, this thesis focuses on potential areas of high performance computing utilization within one category of augmented reality, Marker less systems, for improving the visual quality of an augmented reality system by considering enhanced implementations face feature extraction, detection and tracking algorithms that run on the graphics processing unit (GPU) and is suitable for video analysis in real-time vision systems. While significant acceleration over standard CPU implementations is obtained by exploiting parallelism provided by modern programmable graphics hardware, the CPU is freed up to run other computations in parallel.

A series of practical experiments have been done using different GPU systems and carried out under varying environment settings. The results of these experiments suggest that the proposed enhanced algorithms and their parallel implementation framework is found to be reasonably accurate for the face recognition and tracking purposes in real time augmented reality systems.

Finally, we evaluated the results of our proposed methods by comparing the achieved quality with results of the latest published conventional augmented reality methods which showed that our GPU-based implementation successfully tracks features in real-time and is approximately 33 times improvement over the CPU-based implementation.

Keywords: Augmented Reality, Virtual Reality, Feature Extraction, Texture Extraction, Face Detection, GPU, PCI Express

Table of Contents

List of Tables

List of Figures

Chapter 1
Introduction

Introduction

1.1 Overview

Augmented Reality (AR) is a combination of real world and computer-generated objects, which refers to a multi-axis spectrum of areas such as Virtual Reality (VR), telepresence, and other related technologies. In Virtual reality the environment is completely synthetic and in telepresence it is completely real[1], and both are combined in real time by adding computer vision and object recognition. The design of such systems requires taking into consideration some aspects such as; (1) Detecting and tracking the desired real world objects; (2) Mixing computer generated objects with the tracked objects; (3) Interactivity and rendering the computer generated 3D Model in real time.

As such, augmented reality is a real world condition modified by computer into a new condition in real time. The AR generated object is a result that is based on real time detection and tracking methods, either marker-less or marker-based tracking. The generated objects are the results of object tracking, hand tracking, or face tracking. Recently, various methods and algorithms have been used for generating more accurate and better AR in real time. The major problem in these applications is the processing time.

Herein lies the problem we are trying to attack in this thesis which is related to how fast can we detect, track, generate, render and interact with the generated model specially face. Indeed, face detection is one of the most difficult parts in AR due to the variability of appearance of this deformable moving object, which makes it difficult to model and does not easily cope with real time implementation using conventional processing techniques.

1.2 Previous Work

Previous studies were directed to optimizing the integral image calculation on the General Purpose Graphics Processing Unit (GPU) by dividing the integral image into sub-regions and a local Integral Image is calculated (LII), the Global Integral Image (GII) is calculated by merging the LIIs. The input image is divided into sub-regions such that the data dependency is limited[2].

Other attempts to accelerate Object detection have been made by OpenCV implementations that reached 1.78 Frame per Second (fps) on VGA image sizes.

Endang Setyati proposed improving face detection; the main interest was not to detect human-alike faces but to detect human faces. They did so by processing a two dimensional (2D) face data into a three dimensional (3D) space aiming to improve the weakness of 2D detection. 3D face detection was able to improve the benefits of face detection by increasing the precision and accuracy of the process. The researcher used Active Shape Model algorithm due to its low error rate but this was at the cost of the performance and processing power consumption, along with a limitation of the image environment as it has significant impact on the output[3].

Agrawal proposed a design, which will have a great impact on video surveillance systems, whereas a segmentation is used after background subtraction and background estimation, this was used to reduce noise and locate a moving target in a frame. For multiple moving objects detection in poor lighting conditions, wavelet-based contrast change detector was integrated with locally adaptive thresholding scheme for initial frames. For later frames latest Change detector mechanism was used. This has significant results for surveillance systems, where the camera is mounted in a stable location, but for handheld devices this will not be accurate for our study[4].

Daniel Hefenbrock and Jason Oberg proposed a design by executing a sequence of kernels, functions that run under Single Instruction, Multiple Threads (SIMT) model. They reached a performance of 3.8fps on a single GPU but on the cost of GPU utilization[5].

Bilgic proposed using Histograms of Oriented Gradients (HoG) features on a GPU, by using Cascade-of-rejectors algorithm to make fast rejections for blocks that do not contain a person in early stages.Viola & Jones using HAAR-like features first introduced this, although it has higher error rate, but it has better performance for a real-time application such as Augmented Reality[6].

Edmund Weng lately proposed an improved version of the SURF algorithm, which was used for features extraction from live mobile camera image and recognition of real world objects.

Homography techniques were used to determine the pose matrix from extracted features. Different characteristics of virtual objects such as rotation, scaling and translation were controlled by calculated pose matrix[7].

Our research was based on enhancing a proposed design, which was based on pre-scaling the HAAR-like features based on all possible scale factors, meaning that the GPU will no longer need to scale neither the features nor the image itself. The GPU then will only be used to evaluate features and scales in a cascading manner. Each thread accessed its own scaled feature according to its window size, andits start and end position (x,y position) of the window and the size it will be working on. This implementation targeted VGA images of size [640*480] with a smallest window size of [20*20].This approach reaches34fps[8].

1.3 Motivation

The image processing and computer vision communities, owing to the multiplicity of applications, have studied real time face detection and tracking in video sequences for more than ten years: teleconferencing, human machine interaction, and robotics. However, despite the ongoing progress in image processing and video tracking algorithms, the increasing computation speed of digital multi-core processors and Graphics processing units (GPUs) is still the object of active research in the area of real time processing. Indeed, face detection is one of the most difficult parts in AR due to the variability of appearance of this deformable moving object, which make it difficult to model and do not easily cope with real time implementation using conventional processing techniques.

Based on the above, one of the major motivations in this thesis is how we can utilize the capabilities of GPUs for solving augmented reality complex real-time problems specially face detection and tracking in order to achieve higher user perception in real time.

Rapid development realms of parallel processing using multicore architectures and specifically GPUs could have been more efficiently utilized in solving Augmented Reality complex real-time computations in order to achieve higher user perception.

AR is a real time application, and research tends to perform marker-less object detection and tracking, which adds extra computational overhead to perform feature detection and tracking on each frame. This problem drovethe researchers to take benefit of GPUs, as GPUs where originally designed to handle real-time graphics for gaming industry. Khronos group provided a set of APIs and standard programming language to interact with GPUs, which is OpenCL.

"*The goal for developers is to utilize all of the system's resources*" Arnon Peleg, Software Products Marketing Manager, Intel.

The concept, referred to as "General Purpose Graphics Processing Unit or GPGPU", essentially make the GPU available for computing chores other than graphics, and the ability to offload over-taxed CPUs can boost application performance.