Using Multiple Graphics Cards as a General Purpose Parallel Computer : Applications to Computer Vision

James Fung and Steve Mann*
University of Toronto
Department of Electrical and Computer Engineering
10 King's College Road, Toronto, Ontario, Canada
{fungja,mann}@eecg.toronto.edu

Abstract

Pattern recognition and computer vision tasks are computationally intensive, repetitive, and often exceed the capabilities of the CPU, leaving little time for higher level tasks. We present a novel computer architecture which uses multiple, commodity computer graphics devices to perform pattern recognition and computer vision tasks many times faster than the CPU. This is a parallel computing architecture that is quickly and easily constructed from readily available hardware. It is based on parallel processing done on multiple Graphics Processing Units (GPUs). An eigenspace image recognition approach is implemented on this parallel graphics architecture. This paper discusses methods of mapping computer vision algorithms to run efficiently on multiple graphics devices to maximally utilize the underlying graphics hardware. The additional memory and memory bandwidth provided by the graphics hardware provided for significant speedup of the eigenspace approach. We show that graphics devices parallelize well and provide significant speedup over a CPU implementation, providing an immediately constructible low cost architecture well suited for pattern recognition and computer vision.

1. Introduction

In this paper, we present results from a system which uses multiple computer graphics cards to operate as a parallel computer architecture which carries out pattern recognition and computer vision computations much faster than any single processor system. Many tasks in pattern recognition and computer vision are computationally intensive and repetitive (and therefore good candidates for parallel architecture). The requirements of pattern recognition and computer vision often exceed the CPUs capabilities. Moreover, freeing up the processor from the pattern recognition tasks allows it to perform other, higher level tasks.

*Thanks to NSERC, SSHRC, Canada Council for the Arts, Ontario Arts Council, Toronto Arts Council, and Ontario Graduate Scholarships/Lewfam Foundation Scholarships in Science and Technology agency for support. Thanks to nVIDIA, ATI, and Viewcast for equipment donations.

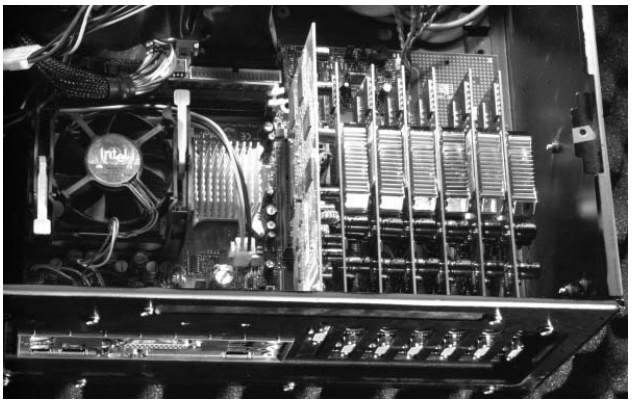

Figure 1: A computer vision machine with 6 PCI graphics cards, and 1 AGP graphics card. Each PCI cards has a GeForce FX 5200 GPU which runs pattern recognition and computer vision tasks in parallel, creating a cheap, powerful, and easily constructible parallel architecture well suited for pattern recognition and computer vision.

Modern computers now incorporate a Graphics Processing Unit (GPU), which carries out the computation required for graphics applications efficiently.

In fact, graphics cards themselves are now becoming a significant part of the overall compute power of any modern system. For instance, modern GPUs now have more transistors than modern CPUs. Consider that the Intel Xeon processor has 108 million transistors, the Radeon R300 GPU has 110 million transistors, and the GeForce FX GPU has 125 million transistors Furthermore, the transistors on the Xeon are largely (two thirds) used for implementing cache memory, whereas the majority of the graphics cards transistors are used in the implementation of multiple floating point SIMD units. Additionally, the graphics cards still boast a very high memory access rate using DDR and DDR2 memory.

It has often been said that computer graphics and com-