

UNsharp Filter Report

B00235610 2017

This reports sets out the development and techniques used to optimize a C++ program that sharpened an input image. GPGPU was used in tandem with various optimizations to provide a version that was significantly faster but more accurate as well.

Table of Contents

[Introduction 1](#_Toc478585898)

[Problem 1](#_Toc478585899)

[Hardware 1](#_Toc478585900)

[Objectives 2](#_Toc478585901)

[Development 3](#_Toc478585902)

[OpenCL Development 3](#_Toc478585903)

[C++ Bindings 3](#_Toc478585904)

[Gaussian Blur 3](#_Toc478585905)

[Image Objects 3](#_Toc478585906)

[OpenCL Optimisations 4](#_Toc478585907)

[Filter Mask 4](#_Toc478585908)

[Constants 4](#_Toc478585909)

[1D Blur 5](#_Toc478585910)

[Viewer Development 5](#_Toc478585911)

[Visualiser development 5](#_Toc478585912)

[Performance 6](#_Toc478585913)

[Method 6](#_Toc478585914)

[Small Image – Lena 6](#_Toc478585915)

[Computation Time 6](#_Toc478585916)

[Program Time 6](#_Toc478585917)

[Large Image – Ghost Town 6](#_Toc478585918)

[Computation Time 6](#_Toc478585919)

[Program Time 6](#_Toc478585920)

[Conclusion 7](#_Toc478585921)

[Positives 7](#_Toc478585922)

[Negatives 7](#_Toc478585923)

[Summary 7](#_Toc478585924)

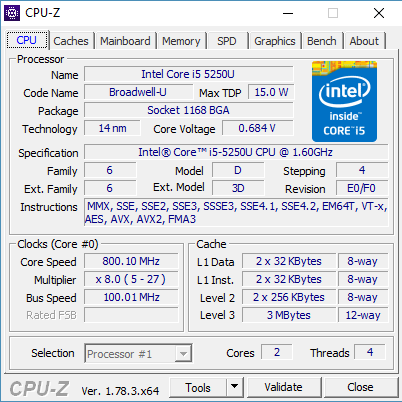
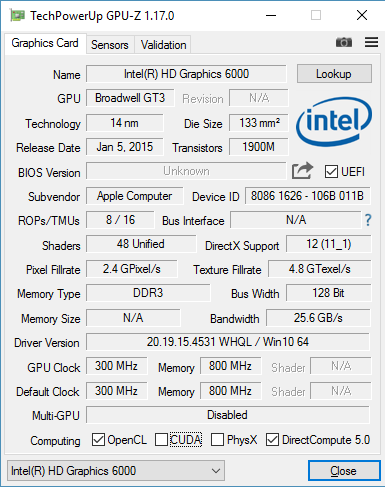
# Introduction

## Problem

C++ code was given which takes an PPM input image and performs an unsharp mask and writes the result to disk. The unsharp mask is implemented as a box blur whose result is subtracted from the input image however, the code is slow and could be made faster by using parallel compute and this is what is set out in this report where I describe how this application was made faster using OpenCL.

## Hardware

As we will be comparing performance I decided to give an overview of the hardware which the timings were recorded on. On the CPU side the code was run on a i5 dual core processor with hyper threading allowing for 4 logical cores. On the GPU, I am using an integrated chip the Intel HD 6000 rather than a dedicated graphics card. It was necessary to dispense this information I feel as the performance increase shown may have been even larger had I run the kernels on a dedicated graphics card which has more horsepower at its disposal.



## Objectives

There are several objectives for this coursework. First is that the optimized version should be significantly faster than the original by factors not percent. Additionally, the optimized version should produce a more accurate result by not taking shortcuts as the given method does. Finally, utilities should be developed to help and test the program. This includes a visualizer which renders the generated output in real time with keys allowing to see the effect of increasing or decreasing of the blur radius. Additionally, there should be a Viewer program that loads and displays a PPM file quickly and accurately.

# Development

## OpenCL Development

When porting the code over to OpenCL serval implementation decisions were taken.

### C++ Bindings

When writing the OpenCL code the C++ bindings were used to do as much as possible with little code. It also enabled easy clean up and resource management thanks to resource acquisition is initialization (RAII). The benefits of this approach can be seen in the given source code and in the Figure 1.



Figure 1: C++ vs C example with OpenCL

### Gaussian Blur

The box blur was dropped in favour of a full Gaussian blur that would be optimized later. Continuous use of a box blur is an optimisation technique for a Gaussian blur, where repeated use as stated by the central limit theorem results in an accurate approximation of a Gaussian blur. By using a Gaussian blur, we will produce a result that is more accurate and by implementing a 2D Gaussian blur meant the sharpen and blur operation to be integrated into the one OpenCL kernel which was nice.

### Image Objects

From the start of the project OpenCL image objects were used from the start for added performance. Not only does this allow the program to utilise automatic clamping if we access a pixel outside of the image but it allows us to choose how we sample pixels in the image. This is a great functional feature but there is an added performance benefit as well as GPUs have specific hardware to enable fast image manipulation and by using it we will likely see a performance increase of traditional buffers.

## OpenCL Optimisations

The first implementation was a 2D Gaussian blur single pass program. This was simple however it had several key performance faults. First was that blur weights were calculated on the fly and were not pre-calculated which resulted in a lot of redundant calculations. Second was that it was a 2D single pass blur which resulted in a significant number of additional operations as well.

### Filter Mask

The first thing to solve was to pre-calculate the Gaussian blur mask values. This was done on the CPU end then stored in a buffer which could then be read from inside the kernel. This reduced a lot of redundant operations found inside the kernel and made it a lot faster but there was much more to do.

### Constants

Several parameters such as blur radius are passed as constants to the program via command line arguments when the program is built. This means that when the size of the mask doesn’t exceed constant memory size limits it will be marked as constant allowing for faster and cached reads. It also saves many parameters having to be passed in via the Kernel object as can be seen in Figure 2.



Figure 2: Constants example, if not defined defaults are used

### 1D Blur

With pre-calculated mask values, I looked at ways to optimise the 2D Gaussian blur. Thankfully the 2D Gaussian blur can be calculated with two passes a horizontal blur pass and a vertical blur pass computations. While this means, the program would need to launch two kernels instead of one it would lead to so many computations being removed that it was an optimisation that was impossible not to implement. This also meant a small blur mask size as the mask generated would now be a 1D mask instead of a 2D mask meaning in almost all cases it can be stored in constant memory on the GPU.

## Viewer Development

To aid in the development of this optimised version a simple C# viewer application was written to view PPM files. The provided source code reads and outputs a PPM file a file type which isn’t support by most image viewing applications. So, I wrote this one myself which simply loads and shows a PPM. It was optimised to enable the fast loading of these files as in some of our test cases we outputted an 8k image which some other software I used previously for displaying PPM files didn’t always work well.

## Visualiser development

For extra marks a visualizer was developed. This uses OpenCL to fill an OpenGL texture that is then rendered on screen in real time. This wasn’t too complex as the standard Image2D objects were just replaced by the ImageGL type.



Figure 3

Add in some additional parameters when creating an OpenCL context and OpenCL is able and ready to be used in tandem with OpenGL. When creating the context and windowing code GLFW was used. We use it to open a window and setup the OpenGL context. I then subscribe to keyboard events which handle increasing the blur radius. All the parameters are put in the windows title so you know what the current value of the radius is.

# Performance

## Method

For measuring performance, varying blur radii was tested as was image size. Two timings were taken the program time and computation time which are explained in each individual section.

## Lena Image

The first image tested was a relatively small image with a width and height of 512 pixels.

### Computation Time

Computation time is the time needed to calculate the output and store it in system ram. It does not take into consideration any OpenCL initialisation costs but does record the time taken to transfer the time from the GPU to the CPU. As we can see in the below graph the OpenCL version is significantly faster being 100x times faster with a blur radius of 5 and a 4000x speed increase when using a blur radius of 57.

### Program Time

Program time takes into consideration the total program time but excludes the time needed to write and read the image from disk so this time doesn’t corrupt out performance comparison. This means for the OpenCL version initialisation time is considered as it the additional time to compile the kernels setup the buffers etc. As you can see in the chart below recording this time results in a performance hit for the OpenCL version.

The result is it is beaten by the CPU version because the initialisation cost wipes out any computation gain. However, as we can see that with a blur radius of 35 it is still 30 times faster. What is interesting is that the GPU performs better overall with a larger radius size. This has been speculated to be because the GPU is better at transferring larger amounts of memory rather small chunks.

## Ghost Town Image

The second image tested is an 8k image with a width of 3840 pixels and a height of 2160. The reason for testing the two images was to see how image size would mutate the performance comparison.

### Computation Time

### Program Time

# Conclusion

This was a project that was bigger than expected and much can be said about it.

## Positives

There were many positives to the project. The objectives were met with a large increase in performance recorded as well as a more accurate output image. What this effectively demonstrates is that by using the GPU we need not take traditional shortcuts and sacrifice image quality when implementing traditional algorithms. Because not only is the image more accurate but it is generated faster making the traditional method redundant if you want a faster version with perfect results.

## Negatives

There were some negatives to this project as well. First a box blur was not implemented at all on the GPU which means a comparison was unable to be made. Additionally, the CPU was not sped up at all using threads and SIMD instructions meaning it’s an unfair comparison between the two however the project does effective demonstrate how much faster operations can be when GPUs or parallel compute are leveraged.

## Summary

So, to summarise this project was a great success in my view with a large increase in performance and a more accurate result with various utilities developed to show the result and effects when varying blur radii. While more work could have been done by experimenting with a wider range of blur techniques and CPU multithreading the result in my opinion is still a great project which meets its objectives hole heartedly.