



**University of Sousse**

---

**National Engineering School of Sousse**



# **Comparison of Performance Between CPU and GPU on Point-to-Point Image Processing Operations**

**Conducted by:**

MHAMED Mohamed

Saddem Ghada

**Field:**

Industrial Electronics

# I. Abstract

This report explores the comparison of performance between CPU and GPU in point-to-point image processing operations. Image processing, a specialized area within Digital Signal Processing, requires substantial computational power due to the nature of its complex algorithms, such as convolution and matrix operations. Traditional CPU-based sequential processing often fails to meet the high computational demands of modern image processing tasks. To address this, GPUs, with their parallel computing capabilities, have emerged as a more efficient solution. Using CUDA, a platform for parallel processing created by NVIDIA, this report evaluates various image processing algorithms executed on both CPU and GPU. The results indicate that the GPU significantly outperforms the CPU in most cases, particularly in tasks involving high-resolution images.

## II. Introduction

Image processing is a subfield of Digital Signal Processing (DSP) that involves manipulating images to enhance or extract useful information. With the increasing demand for high-quality images in applications such as medical imaging, computer vision, and satellite imagery, traditional CPUs have struggled to keep up due to their inherent serial processing limitations.

Parallel processing, facilitated by GPUs, has proven to be an efficient alternative. Unlike CPUs, which process data sequentially, GPUs can perform multiple operations simultaneously due to their numerous cores. This makes them ideal for tasks such as image processing, where multiple pixels can be processed concurrently. CUDA, developed by NVIDIA, allows developers to harness the power of the GPU for computationally intensive tasks.

This report focuses on point-to-point image processing, which transforms each pixel individually without considering neighboring pixels. Operations such as brightening, darkening, and converting RGB images to grayscale are executed on both CPU and GPU to compare their performance.

## III. Important Definitions

- **Digital Image Processing:** Refers to the manipulation of digital images through computational algorithms.
- **Parallel Processing:** The simultaneous execution of multiple computations, reducing the time required for tasks involving large datasets, such as image processing.
- **CUDA (Compute Unified Device Architecture):** A parallel computing platform developed by NVIDIA, enabling developers to utilize the GPU's processing power.
- **OpenCV:** An open-source library of computer vision and machine learning algorithms, widely used for image processing tasks.
- **GPU (Graphics Processing Unit):** A hardware component designed for rapid image rendering and parallel computations, making it ideal for tasks requiring intensive mathematical operations.

- **Grayscale:** A type of image where each pixel represents a shade of gray, often used in image processing to simplify the data being analyzed.
- **Thresholding:** A method of segmenting images by converting grayscale images to binary images based on a threshold value.

## IV. Experiment and Results



The experiment involves converting an image to grayscale using both CPU and GPU processing techniques. The primary goal is to compare the performance of the two approaches. Below are the steps and results obtained from the experiment.

### Experimental Setup

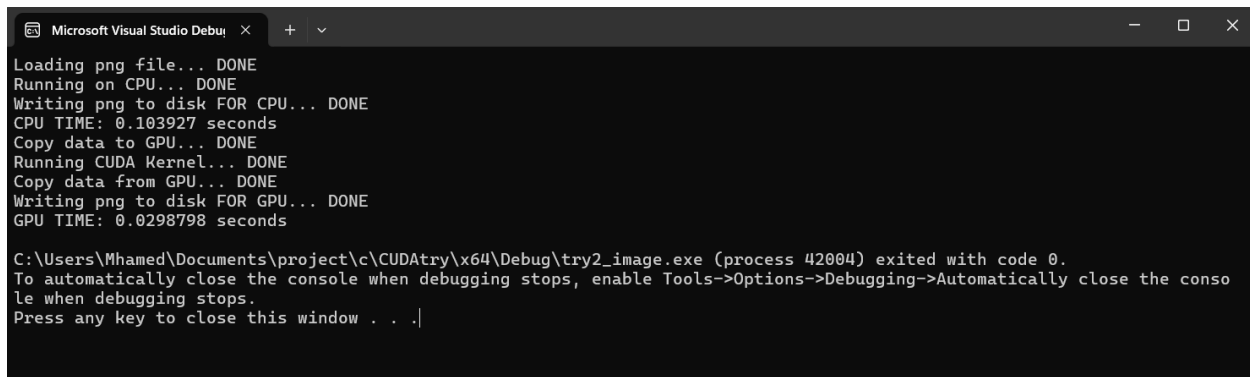
1. **Input Image:** The input image is a 4K resolution PNG file with RGBA channels, where each pixel has four components: red (r), green (g), blue (b), and alpha (a).
2. **CPU-Based Image Grayscale Conversion:** The CPU-based grayscale conversion iterates over each pixel, calculating the grayscale value using the following formula:

$$\text{pixelValue} = 0.2126 \times R + 0.7152 \times G + 0.0722 \times B$$

This pixel value is then applied uniformly to the red, green, and blue channels to produce a grayscale image.

3. **GPU-Based Image Grayscale Conversion:** A CUDA kernel performs a similar grayscale conversion, where each thread processes a single pixel in parallel. The threads are arranged in a grid of blocks, each block containing 32x32 threads, processing a chunk of the image at a time.

## Results



```
Microsoft Visual Studio Debug Console
Loading png file... DONE
Running on CPU... DONE
Writing png to disk FOR CPU... DONE
CPU TIME: 0.103927 seconds
Copy data to GPU... DONE
Running CUDA Kernel... DONE
Copy data from GPU... DONE
Writing png to disk FOR GPU... DONE
GPU TIME: 0.0298798 seconds

C:\Users\Mhamed\Documents\project\c\CUDAtry\x64\Debug\try2_image.exe (process 42004) exited with code 0.
To automatically close the console when debugging stops, enable Tools->Options->Debugging->Automatically close the console when debugging stops.
Press any key to close this window . . .|
```

- **CPU Time:** 0.103927 seconds
- **GPU Time:** 0.0298798 seconds

The GPU-based approach was approximately **3.48 times faster** than the CPU-based method in this experiment, showcasing the advantage of parallel processing for tasks like image manipulation.

## V. Comparison

In terms of performance, GPU-based image processing significantly outperforms CPU-based methods due to its parallel architecture. For large image sizes, like the 4K image used in this experiment, the difference in execution time becomes highly noticeable.

- **CPU Processing:** While straightforward to implement, CPU image processing scales poorly with image size due to its sequential nature. Each pixel is processed one by one, which makes it time-consuming for high-resolution images.
- **GPU Processing:** CUDA kernels allow each pixel to be processed in parallel, greatly improving the execution time, especially for large datasets like 4K images. By leveraging thousands of GPU cores, the task can be done much faster than on a typical CPU with fewer cores.

## VI. Conclusion

This experiment highlights the superior performance of GPUs over CPUs in point-to-point image processing operations. By leveraging parallel processing, the GPU can handle complex tasks more efficiently, particularly when dealing with large images. The results suggest that for real-time or large-scale image processing tasks, GPUs are the preferred choice due to their significant speed advantage. This experiment also demonstrates the utility of CUDA in maximizing the processing potential of GPUs, making it an essential tool for modern image processing applications.

Future work may involve extending the experiment to more complex image processing tasks such as convolution and filtering, which may reveal even greater performance disparities between CPUs and GPUs.

## CUDA\_Grayscale

The CUDA Image Grayscale Conversion Project is designed to demonstrate the power and efficiency of NVIDIA CUDA in image processing tasks. The primary objective of this project is to convert color images into grayscale, utilizing the parallel processing capabilities of CUDA to achieve significant performance improvements over traditional CPU-based methods.

This project includes multiple implementations, showcasing different approaches to the grayscale conversion process. The resulting grayscale images are generated quickly and efficiently, making this project a valuable resource for anyone interested in leveraging GPU acceleration for image processing tasks.

The project also features a detailed report that documents the implementation process, performance evaluations, and results, offering insights into the effectiveness of CUDA for real-time image processing applications.

## Includes and Structs

```
#include "cuda_runtime.h"
#include "device_launch_parameters.h"

#include <iostream>
#include <string>
#include <cassert>
#include <chrono>

#include "stb_image.h"
#include "stb_image_write.h"

struct Pixel
{
    unsigned char r, g, b, a; // RGBA pixel structure
};
```

## CPU Grayscale Conversion Function

```
void ConvertImageToGrayCpu(unsigned char* imageRGBA, int width, int height)
{
    for (int y = 0; y < height; y++)
    {
        for (int x = 0; x < width; x++)
        {
            Pixel* ptrPixel = (Pixel*)&imageRGBA[y * width * 4 + 4 * x];
            unsigned char pixelValue = (unsigned char)(ptrPixel->r * 0.2126f + ptrPixel->g * 0.7152f + ptrPixel->b * 0.0722f);
            ptrPixel->r = pixelValue;
            ptrPixel->g = pixelValue;
            ptrPixel->b = pixelValue;
            ptrPixel->a = 255; // Set alpha to fully opaque
        }
    }
}
```

```
}
```

ConvertImageToGrayCpu: A function that processes the image pixel by pixel.

Conversion Logic: It calculates the grayscale value using the formula:

$$\text{Gray} = 0.2126 \times R + 0.7152 \times G + 0.0722 \times B$$

Setting Values: The calculated grayscale value replaces the red, green, and blue channels, while the alpha channel is set to 255 (fully opaque).

## GPU Grayscale Conversion Kernel

```
__global__ void ConvertImageToGrayGpu(unsigned char* imageRGBA)
{
    uint32_t x = blockIdx.x * blockDim.x + threadIdx.x;
    uint32_t y = blockIdx.y * blockDim.y + threadIdx.y;
    uint32_t idx = y * blockDim.x * gridDim.x + x;

    Pixel* ptrPixel = (Pixel*)&imageRGBA[idx * 4];
    unsigned char pixelValue = (unsigned char)
        (ptrPixel->r * 0.2126f + ptrPixel->g * 0.7152f + ptrPixel->b * 0.0722f);
    ptrPixel->r = pixelValue;
    ptrPixel->g = pixelValue;
    ptrPixel->b = pixelValue;
    ptrPixel->a = 255; // Set alpha to fully opaque
}
```

ConvertImageToGrayGpu: A CUDA kernel function to perform the grayscale conversion on the GPU.

Thread Indexing: Each thread computes its x and y position based on the block and thread indices.

Pixel Processing: Similar grayscale conversion logic is applied to the pixels, but this happens in parallel across many threads.

## Main Function

```
int main(int argc, char** argv)
{
    // Check argument count
    if (argc < 2)
    {
        std::cout << "Usage: 02_ImageToGray <filename>";
        return -1;
    }

    // Open image
```

```

int width, height, componentCount;
std::cout << "Loading png file...";
unsigned char* imageData = stbi_load(argv[1], &width, &height, &componentCount, 4);
if (!imageData)
{
    std::cout << std::endl << "Failed to open \"" << argv[1] << "\"";
    return -1;
}
std::cout << " DONE" << std::endl;

// Validate image sizes
if (width % 32 || height % 32)
{
    std::cout << "Width and/or Height is not divisible by 32!";
    return -1;
}

// CPU processing
std::cout << "Running on CPU...";
auto startCpu = std::chrono::high_resolution_clock::now();
ConvertImageToGrayCpu(imageData, width, height);
auto endCpu = std::chrono::high_resolution_clock::now();
std::chrono::duration<double> cpuDuration = endCpu - startCpu;
std::cout << " DONE" << std::endl;

// Save the CPU-processed image
std::string cpuFileName = "ship_4k_rgba_gray_cpu.png";
std::cout << "Writing png to disk FOR CPU...";
stbi_write_png(cpuFileName.c_str(), width, height, 4, imageData, 4 * width);
std::cout << " DONE" << std::endl;
std::cout << "CPU TIME: " << cpuDuration.count() << " seconds" << std::endl;

// Reload the original image to avoid processing already processed data
stbi_image_free(imageData);
imageData = stbi_load(argv[1], &width, &height, &componentCount, 4);
if (!imageData)
{
    std::cout << std::endl << "Failed to reopen \"" << argv[1] << "\"";
    return -1;
}

// GPU processing
std::cout << "Copy data to GPU...";
unsigned char* ptrImageDataGpu = nullptr;
assert(cudaMalloc(&ptrImageDataGpu, width * height * 4) == cudaSuccess);
assert(cudaMemcpy(ptrImageDataGpu, imageData, width * height * 4, cudaMemcpyHostToDevice) ==
cudaSuccess);
std::cout << " DONE" << std::endl;

// Run the kernel
std::cout << "Running CUDA Kernel...";
dim3 blockSize(32, 32);
dim3 gridSize(width / blockSize.x, height / blockSize.y);
auto startGpu = std::chrono::high_resolution_clock::now();
ConvertImageToGrayGpu << <gridSize, blockSize >> > (ptrImageDataGpu);
cudaDeviceSynchronize();

```

```

auto endGpu = std::chrono::high_resolution_clock::now();
std::chrono::duration<double> gpuDuration = endGpu - startGpu;
std::cout << " DONE" << std::endl;

// Copy the processed data back from the GPU
std::cout << "Copy data from GPU...";
assert(cudaMemcpy(imageData, ptrImageDataGpu, width * height * 4, cudaMemcpyDeviceToHost) ==
cudaSuccess);
std::cout << " DONE" << std::endl;

// Save the GPU-processed image
std::string gpuFileName = "ship_4k_rgba_gray_gpu.png";
std::cout << "Writing png to disk FOR GPU...";
stbi_write_png(gpuFileName.c_str(), width, height, 4, imageData, 4 * width);
std::cout << " DONE" << std::endl;
std::cout << "GPU TIME: " << gpuDuration.count() << " seconds" << std::endl;

// Free memory
cudaFree(ptrImageDataGpu);
stbi_image_free(imageData);

return 0;
}

```