

**МОСКОВСКИЙ АВИАЦИОННЫЙ ИНСТИТУТ  
(НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)**

**Институт №8 «Информационные технологии и прикладная математика»  
Кафедра 806 «Вычислительная математика и программирование»**

**Лабораторная работа №3  
по курсу «Программирование графических процессоров»**

**Классификация и кластеризация изображений на GPU.**

Выполнил: М.А.Трофимов

Группа: 8О-408Б

Преподаватели: К.Г. Крашенинников,  
А.Ю. Морозов

Москва, 2021

## Условие

**Цель работы.** Научиться использовать GPU для классификации и кластеризации изображений. Использование *константной памяти*. **Вариант:** 1, Метод Максимального Правдоподобия.

## Программное и аппаратное обеспечение

### Характеристики GPU "NVIDIA GeForce GTX 950"

CUDA Driver Version / Runtime Version 11.4 / 11.4  
CUDA Capability Major/Minor version number: 5.2  
Total amount of global memory: 1997 MBytes (2094137344 bytes)  
(006) Multiprocessors, (128) CUDA Cores/MP: 768 CUDA Cores  
GPU Max Clock rate: 1278 MHz (1.28 GHz)  
Memory Clock rate: 3305 Mhz  
Memory Bus Width: 128-bit  
L2 Cache Size: 1048576 bytes  
Maximum Texture Dimension Size (x,y,z) 1D=(65536), 2D=(65536, 65536), 3D=(4096, 4096, 4096)  
Maximum Layered 1D Texture Size, (num) layers 1D=(16384), 2048 layers  
Maximum Layered 2D Texture Size, (num) layers 2D=(16384, 16384), 2048 layers  
Total amount of constant memory: 65536 bytes  
Total amount of shared memory per block: 49152 bytes  
Total shared memory per multiprocessor: 98304 bytes  
Total number of registers available per block: 65536  
Warp size: 32  
Maximum number of threads per multiprocessor: 2048  
Maximum number of threads per block: 1024  
Max dimension size of a thread block (x,y,z): (1024, 1024, 64)  
Max dimension size of a grid size (x,y,z): (2147483647, 65535, 65535)  
Maximum memory pitch: 2147483647 bytes  
Texture alignment: 512 bytes

### Характеристики CPU Intel i5-4460

# of Cores 4  
# of Threads 4  
Processor Base Frequency 3.20 GHz  
Max Turbo Frequency 3.40 GHz  
Cache 6 MB Intel® Smart Cache  
Bus Speed 5 GT/s  
Intel® Turbo Boost Technology 2.0 Frequency 3.40 GHz  
TDP 84 W

### Характеристики RAM

Total 15 Gi  
Swap 2 Gi

Операционная система: Ubuntu 20.04 LTE

IDE Sublime Text 3

Compiler nvcc for cuda 11.4

## Метод решения

Сущность решения в том, что мы считаем оценки мат. ожидания и ковариации, а потом, применяя ММП классифицируем пиксель по максимальному значению ММП.

Параллелизация реализована с помощью константной памяти в виде массива из 32ух элементов типа ClassData. Каждый поток перебирает возможные классы и выбирает номер класса тот, что больше всего подходит.

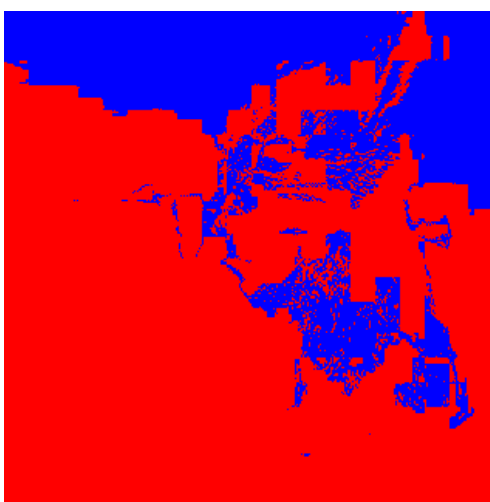
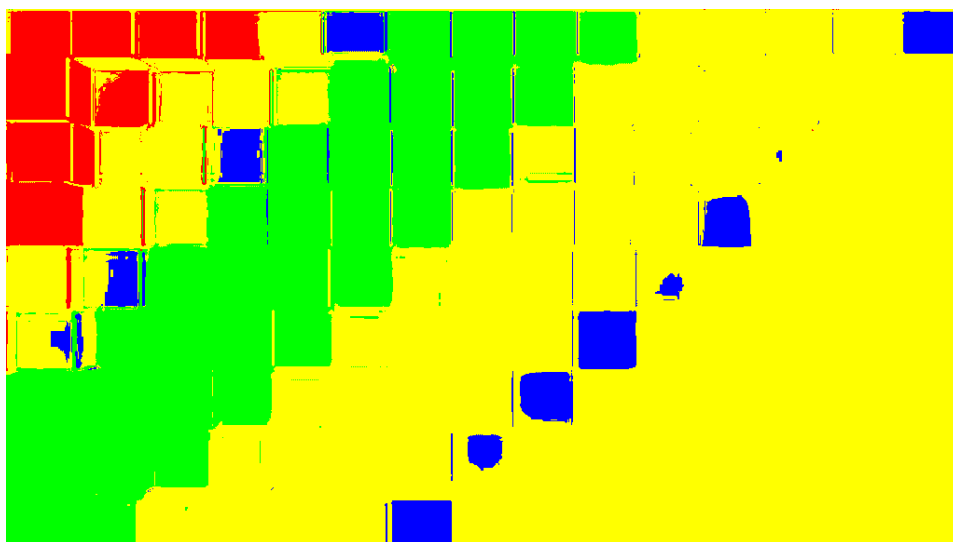
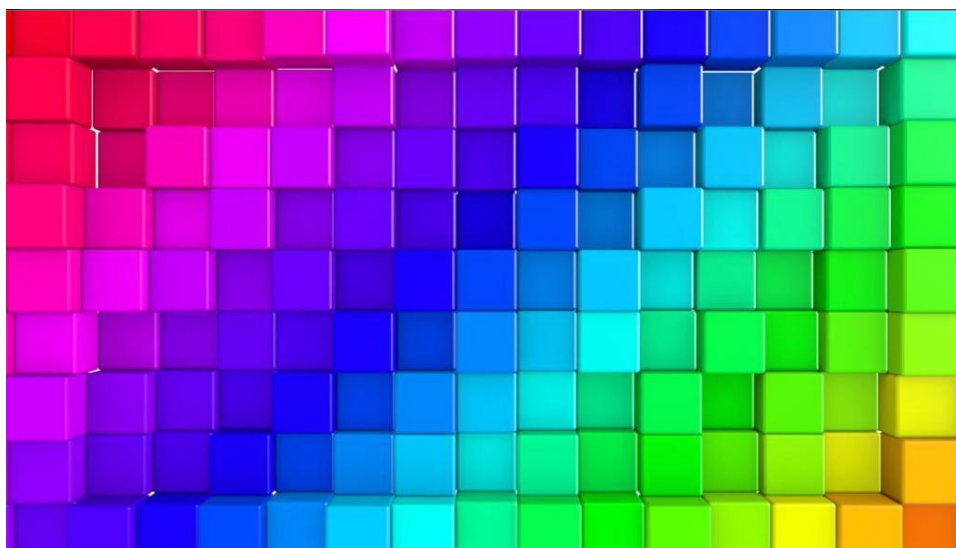
## Описание программы

Всё писалось одним файлом main.cu, в котором функция main, read\_file - функция для считывания из входного файла данных, write\_file - записи получившихся данных в файл, calc\_metric для считывания из stdin выборок и подсчёта для каждой выборки среднее, обратную матрицу ковариации и её определитель, сохраняя в класс ClassData, который содержит необходимые метрики для определения принадлежности к текущему классу. Compute\_mmp - функция device (видеокарты) для подсчёта значения ММП для проверки принадлежности данного пикселя к данному классу.

## Результаты

Конфигурация	тест картинки 320x320	тест картинки 500x500	тест картинки 889x906
На CPU	143.62 ms	388.15 ms	1268.70 ms
1,32	84.63 ms	384.57 ms	1026.32 ms
32, 32	3.82 ms	20.96 ms	59.77 ms
64, 64	2.75 ms	15.00 ms	43.04 ms
512, 512	2.62 ms	14.39 ms	40.29 ms
1024, 1024	2.63 ms	15.32 ms	40.34 ms
2048, 1024	2.63 ms	13.37 ms	40.36 ms

## Демонстрация работы



## Выводы

Как видно, параллелизация классификации весьма сильно улучшает время работы до 300 раз, что очень сильно, но большое число блоков и тредов не всегда

нужно. Как видно, минимальное время достигается на конфигурации 64, 64, и больше не уменьшается. Скорее всего, это из-за того, что достаточно мало работы приходится на каждый тред, поэтому прирост эффективности не происходит.