

Эффективная реализация алгоритма поиска лиц людей на изображении для архитектуры NVIDIA CUDA

И.И. Зиновьев

Владимирский государственный университет имени А.Г. и Н.Г. Столетовых

Комплекс алгоритмов для обнаружения объектов, предложенный в 2001 году Полом Виолой (Paul Viola) и Майклом Джонсом (Michael Jones) [1], логически состоит из двух компонентов: обучение классификатора и непосредственно обнаружение объектов на изображении. На практике основные требования к скорости работы предъявляются ко второму алгоритму. В данной работе рассматривается реализация этого алгоритма для архитектуры NVIDIA CUDA, которая позволяет в несколько раз ускорить его работу по сравнению с реализацией на центральном процессоре, представленной в библиотеке OpenCV.

Алгоритм Viola-Jones реализован как одно CUDA-ядро. Это позволило минимизировать пересылки данных через глобальную память графического ускорителя. Правильный выбор размера блока потоков и количества используемых регистров позволяет достичь значения коэффициента использования мультипроцессора порядка $2/3$, что является достаточным для нивелирования влияния латентности при работе с памятью.

С целью эффективного использования особенностей работы различных видов памяти графического ускорителя исходные для алгоритма данные, а именно каскад классификаторов и анализируемое изображение, размещаются в текстурной и константной памяти. Это позволяет реализовать эффективный алгоритм масштабирования путем выборки из текстуры с аппаратной билинейной фильтрацией. Из-за большого числа обращений к памяти при работе рассматриваемого алгоритма разделяемая память используется как управляемый L1 кэш. Применение технологии «zero copy», позволяет сократить накладные расходы при пересылке данных между хостом и ускорителем практически к нулю за счет их асинхронного выполнения.

Вычисление интегрального изображения является ключевым шагом в работе алгоритма Viola-Jones. Для облегчения процесса расчета интегрального изображения в данной работе предлагается использовать частичное интегральное изображение, которое вычисляется независимо для участков исходного изображения, уже размещенных в разделяемой памяти мультипроцессоров графического ускорителя. Благодаря этому работа отдельных мультипроцессоров становится независимой, что исключает временные затраты на синхронизацию их работы.

В ходе экспериментов было достигнуто ускорение работы алгоритма на графическом ускорителе от 2 до 5 раз в сравнении с однопоточным вариантом, работающем на центральном процессоре. Такой результат в несколько раз превышает тот, что представлен автором работы [2]. Полученная реализация алгоритма показывает хорошую масштабируемость, что позволяет рассчитывать на еще большее ускорение на более современных графических ускорителях.

В данной работе предложен оригинальный вариант эффективной реализации алгоритма Viola-Jones для архитектуры графических ускорителей производства NVIDIA. Следует отметить, что сам алгоритм не слишком приспособлен для реализации на графической карте. Большое число обращений к памяти, простой потоков при работе каскада классификаторов негативно сказываются на производительности. Но даже при этом полученная реализация алгоритма на современных графических ускорителях работает в несколько раз быстрее, чем аналоги на центральном процессоре.

Литература

1. Paul Viola, Michael J. Jones. Robust real-time face detection // International journal of computer vision 57(2), 137–154, 2004
2. J. P. Harvey. GPU Acceleration of Object Classification Algorithms Using NVIDIA CUDA
URL: https://ritdml.rit.edu/bitstream/handle/1850/10894/35445_pdf_00B0B24A-DFD8-11DE-9A30-D21AD352ABB1.pdf?sequence=1. (дата обращения: 11.10.2010)