

Иерархический плотностной алгоритм кластеризации мультиспектральных изображений и его реализация на графических процессорах NVIDIA

РЫЛОВ СЕРГЕЙ АЛЕКСАНДРОВИЧ

Институт вычислительных технологий СО РАН (Новосибирск), Россия
e-mail: RylovS@mail.ru

Одним из основных подходов к автоматической сегментации изображений является использование алгоритмов кластеризации, которые можно разделить на две большие группы – иерархические и неиерархические. Неиерархические алгоритмы позволяют получить фиксированное разбиение данных, а иерархические – систему вложенных разбиений, соответствующих различным уровням иерархии. Характерной проблемой традиционных иерархических алгоритмов является их высокая вычислительная сложность (порядка $O(N^3)$), поэтому "напрямую" для обработки изображений они не применяются.

В докладе представлен вычислительно эффективный иерархический плотностной алгоритм кластеризации для сегментации многоспектральных изображений. В основу его положен сеточный подход, описанный в работе [1]. Элементами иерархии выступают «компоненты связности», которые соответствуют локальным модам плотности распределения данных. Для построения иерархической структуры вводится специальное расстояние между компонентами связности на основе оценки плотности распределения данных. Такая схема позволяет избежать присущих иерархическим методам проблем, связанных с наличием пересекающихся кластеров [2], и уменьшить их вычислительную сложность.

В последние годы для повышения устойчивости и качества результатов кластеризации активно применяются ансамблевые методы. Однако в настоящее время практически отсутствуют работы, посвященные построению ансамблей иерархических разбиений [3]. В данной работе предложен метод комбинирования результатов работы иерархического алгоритма, полученных при различных масштабах сетки. Разработанный на его основе ансамблевый иерархический алгоритм позволяет значительно повысить качество и устойчивость результатов кластеризации, что подтверждают экспериментальные исследования на модельных данных и спутниковых изображениях.

Доступность и непрерывно возрастающая производительность графических процессоров делают их привлекательными для применения в задачах, связанных с обработкой изображений. В данной работе предлагается технология реализации разработанного алгоритма кластеризации на графических процессорах NVIDIA с помощью технологии CUDA [4]. Приводится анализ производительности алгоритма и сравнение его с реализацией на центральном процессоре, а также с параллельной реализацией на CPU с помощью стандарта OpenMP.

Работа выполнена при финансовой поддержке РФФИ (грант № 14-07-31320-мол_а).

Список литературы

1. Куликова Е.А., Пестунов И.А., Синявский Ю.Н. Непараметрический алгоритм кластеризации для обработки больших массивов данных // Тр. 14 конф.

-
- «Математические методы распознавание образов». Изд-во MAKS Press, 2009. С. 149-152.
2. Lu Y., Wan Y. PHA: A fast potential-based hierarchical agglomerative clustering method // Pattern Recognition. 2013. Vol. 46, no. 5. P. 1227-1239.
 3. Mirzaei A., Rahmati M. A novel hierarchical-clustering-combination scheme based on fuzzy-similarity relations // IEEE Trans. Fuzzy Syst. 2010. Vol. 18, no. 1. P. 27-39.
 4. Боресков А.В. и др. Параллельные вычисления на GPU. Архитектура и программная модель CUDA. М.: Изд-во Московского университета, 2012. 336 с.