**Иерархический алгоритм сегментации спутниковых изображений   
и его реализация с использованием технологии CUDA**

***Пестунов И.А., Рылов С.А.***

Институт вычислительных технологий СО РАН

Сегментация является одним из важнейших этапов анализа изображений. В основе наиболее распространенного подхода к сегментации спутниковых изображений лежат алгоритмы кластеризации данных [1], 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. Она может быть представлена в виде дендрограммы, которая содержит исчерпывающую информацию о структуре данных и служит удобным средством визуальной интерпретации результатов. Характерной особенностью иерархических алгоритмов является их высокая вычислительная сложность (порядка О(N^3)), поэтому "напрямую" они не применяются для обработки спутниковых изображений.  
В докладе представлен вычислительно эффективный иерархический плотностный алгоритм кластеризации CCAH для сегментации спутниковых изображений. В его основу положен сеточный подход, описанный в работе [2]. Группировке подвергаются не исходные данные, а элементы сеточной структуры (клетки), введенной в пространстве признаков. Клетки разбиваются на компоненты связности на основе заданного отношения эквивалентности. Компоненты связности соответствуют отдельным локальным модам плотности распределения данных и рассматриваются как базовые элементы иерархии. Для построения иерархической структуры между компонентами связности вводится специальное расстояние на основе оценки плотности распределения данных. Это позволяет избежать присущих иерархическим методам проблем, связанных с наличием пересекающихся кластеров [3]. При этом получаемая матрица расстояний обладает свойством ультраметрики, т.е. однозначно соответствует некоторой дендрограмме [4].  
Сеточный алгоритм способен эффективно обрабатывать большие массивы данных, однако результаты его работы существенно зависят от масштаба используемой сетки. Для увеличения устойчивости получаемых результатов и снижения влияния шума применяется ансамблевый подход [5]. В настоящее время работы, посвященные построению ансамблей иерархических разбиений, практически отсутствуют [4]. В данной работе предложен метод комбинирования результатов работы алгоритма CCAH, полученных при различных масштабах сетки. Разработанный ансамблевый иерархический алгоритм ECCAH позволяет повысить качество результатов кластеризации и их устойчивость, что подтверждают экспериментальные исследования на модельных данных и спутниковых изображениях.  
Доступность и непрерывно возрастающая производительность графических процессоров, а также современные технологии параллельных вычислений, делают их привлекательными для применения в задачах, связанных с обработкой изображений [6]. При реализации предложенных алгоритмов иерархической кластеризации использована технология NVIDIA CUDA, что позволило уменьшить время обработки спутниковых изображений в 10 и более раз.  
Работа выполнена при финансовой поддержке РФФИ (гранты №№ 14-07-00249а, 14-07-31320-мол\_а).  
  
Литература  
1. Пестунов И.А., Синявский Ю.Н. Алгоритмы кластеризации в задачах сегментации спутниковых изображений // Вестник КемГУ. 2012. № 4/2 (52). C. 110-125.  
2. Пестунов И.А., Бериков В.Б., Куликова Е.А., Рылов С.А. Ансамблевый алгоритм кластеризации больших массивов данных // Автометрия. 2011. Т. 47, № 3. С. 49-58.  
3. Lu Y., Wan Y. PHA: A fast potential-based hierarchical agglomerative clustering method // Pattern Recognition. 2013. Vol. 46, no. 5. P. 1227-1239.  
4. 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.  
5. Ghaemi R., Sulaiman M., Ibrahim H., Mustapha N. A Survey: Clustering Ensembles Techniques // World Acad. of Sci., Engineering and Technology. 2009. Vol. 3, no 2. P. 535-544.  
6. Боресков А.В. и др. Параллельные вычисления на GPU. Архитектура и программная модель CUDA. М.: Изд-во Московского университета, 2012. 336 с.