November 2012

MADALGO seminar by Constantinos Tsirogiannis, Aarhus University

Fast Generation of Multiple Resolution Raster Data Sets

Abstract:

In many GIS applications it is important to study the characteristics of a raster data set at multiple resolutions. Often this is done by generating several coarser resolution rasters from a fine resolution raster. In this talk we describe efficient algorithms for different variants of this problem.

Given a raster G of \sqrt{N} \times \sqrt{N} cells we first consider the problem of computing for every 2 \leq m \leq \sqrt{N} a raster G_m of \sqrt{N}/m \times \sqrt{N}/m cells such that each cell of G_m stores the average of the values of m \times m cells of G. We describe an algorithm that solves this problem in Theta(N) time when the handled data fit in the main memory of the computer. We also describe three algorithms that solve this problem in external memory, that is when the input raster is larger than the main memory. The first external algorithm is very easy to implement and requires O(sort(N)) data block transfers from/to the external memory. The second algorithm requires only O(scan(N)) transfers, however this algorithm is cache-aware and assumes that the main memory of the computer can store a considerable number of data blocks. The third algorithm is a fresh, yet-to-be-published upgrade from the previous results; this algorithm requires O(scan(N)) transfers, is cache-oblivious, and makes no assumption on the size of the main memory.

We also study a variant of the problem where instead of the full input raster we handle only a connected subregion of arbitrary shape. For this variant we describe an algorithm that runs in Theta(U \log N) time in internal memory, where U is the size of the output.