MADALGO
CENTER FOR MASSIVE DATA ALGORITHMICS

TerraSTREAM: Terrain Processing Pipeline

What
TerraSTREAM is a collection of software modules for computation on very large digital terrain models.

Problem
Modern sampling techniques yield datasets in the order of hundreds of gigabytes, which cannot be processed by standard software.

Solution
Use provable efficient algorithms specifically tailored to terrains much larger than the size of the memory. These algorithms tries to minimize the number of I/Os performed.

Properties
- Where it makes sense, all modules work on both grid and triangulated terrains.
- Works on GNU/Linux, Mac OS X and Windows.
- Many of the modules export several different parameters that can be tweaked by the user for maximum flexibility.

Usage
The modules presented on this poster are designed independently of any frontend. Frontends designed for:

ArcGIS   GRASS   Cmd Line

Standalone   MapInfo

Point Classification
Given the raw input samples the classification module identifies points on buildings and vegetation. These can be removed leaving only points representing the actual terrain.

Model Construction
This module constructs a terrain model (a raster grid or a triangulation) from the points representing the terrain, which has usually been through the classification module.

Conditioning
Module removes insignificant sinks from a terrain model. Significance is user defined in terms of the height, area and volume of the sinks. Sinks that are judged insignificant are "flooded".

Area   Height   Volume

Flow Routing
Module determines how water flows locally at each point of the terrain model. Supports several different flow models and can easily be extended with new ones. The input model is usually prepared for flow computation using the conditioning module.

Flow Accumulation
Module computes area upstream of each point in a terrain model. Water is routed along the flow paths computed by the routing module. The module can optionally use an auxiliary input model defining the initial flow and compute arbitrary functions at each point.

Pfafstetter Labeling
Module decomposes a river network into a hierarchy of watersheds. The pfafstetter labels define a certain hierarchy, which is easy for humans to visualize.

Contour Maps
Module constructs contour maps of the entire terrain. The contours generated from high resolution terrain models tend to be very jagged and visually unpleasing. Using conditioning with careful definitions of which sinks are insignificant, the generated contours are improved.

Quality Metric
Module computes the distance from each cell in grid terrain to nearest original input points. Gives a numerical value that can be interpreted as the quality of individual grid cells. When overlayed with orthophotos the effects of the classification algorithms are clearly seen.

Fitting Everything Together
The modules can be combined to form a pipeline. This figure shows the typical order in which the modules are invoked:

Point Classification → Grid/TIN Construction → Hydrological Conditioning → Flow Routing → Flow Accumulation → Pfafstetter Labeling

Grid/TIN Construction → Grid Quality Metric

Hydrological Conditioning → Contour Map Generation