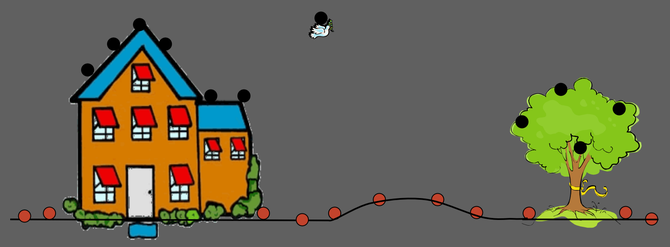
MADALGO
CENTER FOR MASSIVE DATA ALGORITHMICS


Anders Hessellund Jensen
University of Aarhus

TerraSTREAM – Terrain Point Classification

Problem

- Modern sampling techniques such as LIDAR, makes it possible to efficiently acquire detailed terrain data (point samples) for large areas
- A large fraction of the samples do not originate from the bare terrain surface (but from objects above the terrain such as vegetation, buildings, cars, etc)
- Such samples must be removed to obtain a terrain model



All black samples must be removed to obtain a terrain model

Challenges

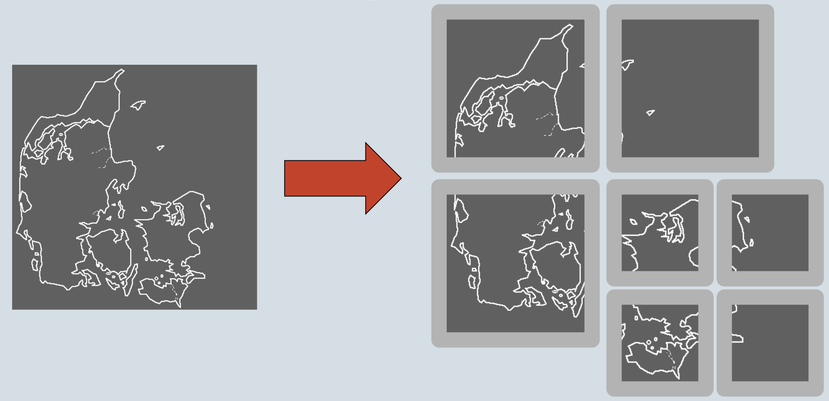
- Point samples are huge
  - Denmark is terabyte sized when sampled at meter resolution
  - I/O-efficient algorithm is needed
- Hard to classify which point are terrain, since e.g.
  - Real terrain may contain discontinuities and rapid changes in slope
  - Areas with dense vegetation may contain very few terrain samples

Normal Classification Algorithm

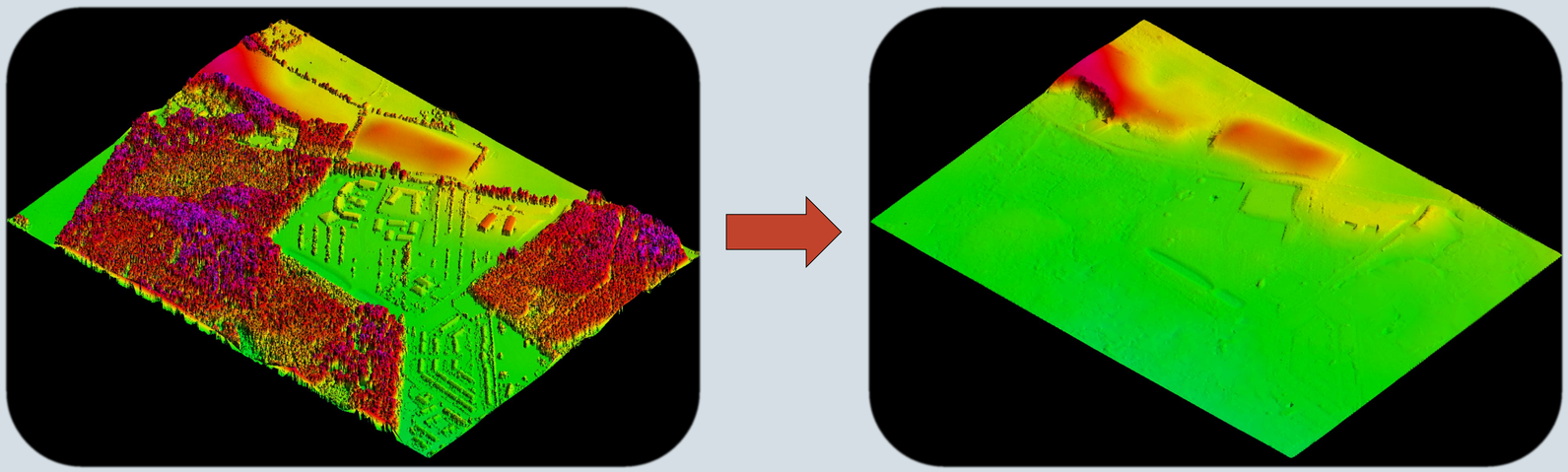
- Initially, a small set of samples originating from the terrain (classified as terrain) is selected and triangulated. The triangulation constitutes a rough model of the terrain surface
- In an iterative process, more samples originating from the terrain are added to the triangulation. Each iteration results in a more refined model of the terrain
- Several different criteria are used to decide what samples should be added
  - Initially, a simple but computationally inexpensive criterion is used
  - In later stages, more complex criterions are used (e.g. to handle difficult areas with discontinuities)

I/O-Efficient Classification Algorithm

- The normal classification algorithm is made I/O-efficient by I/O-efficiently dividing the sampled area into smaller regions. Each region is classified using the normal algorithm
- The size of each region is chosen in a manner adaptive to the density of the input. Dense areas are divided into smaller regions



- Naively dividing into disjoint regions would result in inaccuracies near the boundary of each region. Therefore, an overlap from neighboring regions is included with each region.



Samples *Terrain model*