Streaming processing of spatial data

David Millman

joint work with

Martin Isenburg,
Shawn Brown and
Jack Snoeyink
Scanning Artifacts

scanning of “Michelangelo’s David” courtesy of Marc Levoy
Modeling & Games

screen shot of “The village of Gnisis”, The Elder Scrolls III
Geometry processing applications

- Computer-aided design (CAD): modeling
- Graphics/Games: level of detail rendering
- Finite element (FEM): analysis
- Robotics: path planning
- Geographic Information Systems (GIS): Terrain maps \([M01,vK97]\)
  - TIN triangulation \([PFLM76]\)
  - Contour lines
  - Raster DEM \([T90]\)
LIDAR to TIN to contour or raster

0.5 Billion raw points (11 GB) via Delaunay Triangulation (40 GB) rasterized onto Digital Elevation Grid
Pipes

data in $\rightarrow$ filter $\rightarrow$ results out

ps aux | grep oeyi > myprocs.txt

- Easy to combine simple tools
- Avoids writes to temporary files
- OS handles buffers & time, some apps have their own buffers
- Data must stream in a format that the application expects
Advantages of pipes

- **Composable** - string modules together to solve complex tasks with high throughput.
- **Low Latency** - Allow output from the last module in the string while the first is still reading its input.
- **Memory Coherence** - As data is finalized its memory may be released.
- **Parallelism** - The operating system does load balancing for additional processors or multi-core chips.
- **Modularity** - Prototype modules can be tested and later replaced with "better" techniques.
Spatial locality

• Waldo Tobler’s 1st Law of Geography:
  – “Everything is related to everything else, but near things are more related than distant things.”
  – This law affects how data is used and how it is collected.

• Finalization: document in the data format when an item is last used.
  – For meshes
  – For point sets
TIN format
List coords for verts & indices for $\triangle$s

Batch: Must read all coords before processing any $\triangle$
Streaming Mesh [105]

- Interleave vertices & triangles
- **introduce** and **finalize** vertices

<table>
<thead>
<tr>
<th>v</th>
<th>1.32</th>
<th>0.12</th>
<th>0.23</th>
</tr>
</thead>
<tbody>
<tr>
<td>v</td>
<td>1.43</td>
<td>0.23</td>
<td>0.92</td>
</tr>
<tr>
<td>v</td>
<td>0.91</td>
<td>0.15</td>
<td>0.62</td>
</tr>
<tr>
<td>f</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>done</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>v</th>
<th>0.72</th>
<th>0.34</th>
<th>0.35</th>
</tr>
</thead>
<tbody>
<tr>
<td>f</td>
<td>4</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>done</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Vertices introduced = vertices not used by preceding triangles
- Vertices finalized = vertices not used by subsequent triangles
St. Matthew Statue

- standard indexed format

186 million vertices
372 million triangles
St. Matthew Statue

- standard indexed format
- streaming format

6 GB

186 million vertices
372 million triangles
Advantages of finalization

- Can flush data no longer needed
- Pipeline geometry processing
- Simple API

- Disadvantage: can’t do everything... fixed traversal; not progressive
Stream Processing
with a small memory footprint

processed region
in-cache buffer
output boundary
input boundary
unprocessed region
Points to contours in 7 modules
Points to raster DEM (TO BE DELETED)

- Need the sort phase, but can at least produce raster in strips...
- Effectively produces rasters that are larger than memory size.
Thank you!