Parallel Quadtree Coding of Large-Scale Raster Geospatial Data on GPGPUs
Jianting Zhang¹, Simin You² and Le Gruenwald³
1 City College of New York 2 CUNY Graduate Center 3 University of Oklahoma

Abstract
Global remote sensing and large-scale environmental modeling have generated huge amounts of raster geospatial data. While the inherent data parallelism of large-scale raster geospatial data allows straightforward coarse-grained parallelization at the chunk level on CPUs, it is largely unclear how to effectively exploit such data parallelism on massively parallel General Purpose Graphics Processing Units (GPGPUs) that require fine-grained parallelization. In this study, we have developed an efficient spatial data structure called BQ-Tree to code raster geospatial data by exploiting the uniform distributions of quadrants of bitmaps at the bitplanes of a raster. A fine-grained parallelization scheme has been implemented using Nvidia CUDA. Experiments show that the GPGPU implementation is capable of decoding a BQ-Tree encoded 16-bits NASA MODIS geospatial raster with 22,658*15,586 cells in 190 milliseconds, i.e., 1.86 billion cells per second, on an NVIDIA C2050 GPU card. The performance achieves a 5.9X speedup when compared with the best dual quadcore CPU implementation and a 36.9X speedup compared with a highly optimized single core CPU implementation.

Proposed Solution
Coding geospatial rasters as sets of Bitplane Quadtrees (BQ-Tree)
• BQ-Tree is an efficient and cache-friendly data structure that is suitable for both CPUs and GPUs
• BQ-Trees can be used for data compression and query processing simultaneously
• Decoding can be accelerated using massively parallel GPGPU technologies

Structure of BQ-Tree

Experiments

<table>
<thead>
<tr>
<th>Compressed Size</th>
<th>Chunk Size</th>
<th>1024*1024</th>
<th>4096*4096</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Volume</td>
<td>171,751,596</td>
<td>401,308,369</td>
<td></td>
</tr>
<tr>
<td>Last Level Quadrant Size</td>
<td>272</td>
<td>272</td>
<td>472</td>
</tr>
<tr>
<td>Node Array Size</td>
<td>133,277,609</td>
<td>35,390,341</td>
<td>31,270,652</td>
</tr>
<tr>
<td>LLQS Array Size</td>
<td>136,290,830</td>
<td>191,154,696</td>
<td>136,290,199</td>
</tr>
<tr>
<td>Total Encoded Size</td>
<td>297,568,419</td>
<td>230,855,837</td>
<td>260,560,531</td>
</tr>
<tr>
<td>Compression Ratio</td>
<td>34.6%</td>
<td>29.3%</td>
<td>33.2%</td>
</tr>
</tbody>
</table>

Data: NASA MODIS band1 of the North America 2003097 imagery, 22,658*15,586 cells and 786,295,176 bytes
Hardware: SGI Octane III equipped with two Intel Xeon E5520 CPUs, 48 GB memory and two Nvidia C2050 GPUs (only one node is used)
Software: gcc 4.6.1 (with O3 for speed) on CPU and CUDA 4.0 toolkit on GPU
Results: (1) 3X compression ratio (2) 36.9X speedup of GPGPU decoding when compared with single CPU thread (7005 ms/190 ms)

Future Work
• Integrate the BQ-Tree encoding engine with query processing frontends in databases (e.g., SciDB and FastBit)
• Develop an end-to-end visual analytical prototype system accelerated by GPGPUs for large-scale geospatial rasters