CUDA-ACCELERATED HD-ODETLAP: A HIGH DIMENSIONAL GEOSPATIAL DATA COMPRESSION FRAMEWORK

By

You Li

A Thesis Submitted to the Graduate

Faculty of Rensselaer Polytechnic Institute

in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major Subject: COMPUTER SCIENCE

Approved by the
Examining Committee:

W. Randolph Franklin, Thesis Adviser

Mukkai S. Krishnamoorthy, Member

Peter Fox, Member

Michael J. Wozny, Member

Rensselaer Polytechnic Institute
Troy, New York

July 2011
(For Graduation August 2011)

ABSTRACT

This thesis describes HD-ODETLAP, a geospatial data compression technique to lossily compress high dimensional geospatial datasets. A five dimensional (5D) geospatial dataset consists of several multivariable 4D datasets, which are sequences of time-varying volumetric 3D geographical datasets. These datasets are typically very large in size and demand a great amount of resources for storage and transmission. HD-ODETLAP consists of work from two steps.

Firstly, we build the foundation of HD-ODETLAP method from 3D-ODETLAP method, which targets at compressing 3D geospatial datasets. With proper point selection, our 3D-ODETLAP method approximates uncompressed 3D data using an over-determined system of linear equations. Then this approximation is refined via an error metric. These two steps work alternatively until a predefined satisfying approximation is found. This chosen representative sample set of original 3D dataset is then encoded using simple Run Length Encoding (RLE) and prefix coding technique.

Secondly, based on 3D-ODETLAP, we present 5D-ODETLAP, a lossy compression algorithm and implementation for high dimensional geospatial data. 5D-ODETLAP exploits the spatial dependency and autocorrelation in every dimension in these large datasets. This is an advance on traditional methods that compress only lower dimensional slices. 5D-ODETLAP greedily selects a characteristic subset of the original 5D dataset, chosen to minimize information loss. The selected set of points is further compressed using a coder built from classic encoding methods. That coded set of points is the compressed representation of our dataset. To uncompress the data, 5D-ODETLAP recomputes the values at each point in 5D by solving a sparse over-determined linear system of equations.

After preliminary test of 5D-ODETLAP, we optimize it by using a much more advanced encoding method than the simple RLE and prefix coding. The second

advance in 5D-ODETLAP is our incorporation of a CUDA-based conjugate gradient linear solver into this framework. That exploits the massive, and inexpensive, parallelism available in modern GPUs. We have interfaced CUDA with Matlab to maximize programming efficiency and to minimize data transfer overhead. We have tested 5D-ODETLAP with various datasets and error metrics. With the same mean percentage error, compressed file size by 5D-ODETLAP is 7.67 and 2.14 times as small as that by JPEG2000 and 3D-SPIHT respectively in our eight test datasets on average. 5D-ODETLAP's advantage is even larger under the same maximum percentage error. 5D-ODETLAP has no restrictions in the data types, and it has the flexibility to properly adjust the parameter setting for other datasets with spatial and temporal redundancy.