Big Spatial Data

Rationale

Recent advances in computer hardware have made possible the efficient rendering of realistic 3D models in inexpensive PCs, something that was possible with high end visualization workstations only a few years ago. This class will cover the field of 3D Photography, the process of automatically creating 3D texture mapped models of objects, in detail. We will concentrate on the topics at the intersection of Computer Vision and Computer Graphics that are relevant to acquiring, creating, and representing 3D models of small objects or large urban areas. Many very interesting research questions need to be answered. For example: how do we acquire real shapes? how do we represent geometry? can we detect similarities between shapes? can we detect symmetries within shapes? how do we register 3D geometry with color images? etc. Applications that benefit by this technology include: historical preservation, urban planning, google-type maps, architecture, navigation, virtual reality, e-commerce, digital cinematography, computer games, just to name a few.

All of the above issues must be solved in a parallel processing environment. For example, the Nvidia GTX Titan GPUs with 2,688 cores that support 15*2048 concurrent threads, 6 GB memory and 1.3 and 4.5 Teraflops computing power (double and single precision, respectively) currently available from the market for around $1,000.

Course Description

The increasingly larger data volumes and more complex semantics of spatial information never cease to request more computing power to turn such data and information into knowledge to facilitate decision making in many applications, ranging from location based services to intelligent transportation systems. Current generation of spatial databases and moving object database technologies based on aged hardware architectures is incapable of processing data with reasonable effort and there are Spatial Big- Data (SBD) challenges. In particular, although locating and navigation devices (e.g. GPS, cellular/wifi network-based and their combinations) embedded in smartphones (nearly 500 million sold in 2011) have already generated large volumes of lo-

cation and trajectory data, the next generation of consumer electronics, such as Google Glasses, are likely to generate even larger volumes of location-dependent multimedia data where spatial and trajectory data management techniques will play critical roles in understanding the data. Graphics Processing Units (GPUs) are massively data parallel devices featuring a much larger number of processing cores and concurrent threads which make them significantly different from CPUs that currently support much fewer processing cores and concurrent threads. In addition, the current GPU memory bandwidth is more than an order of magnitude higher than that of CPUs and three orders higher than that of disks. Different from high-performance computing resources in the past that are typically only available to highly selective research groups, GPUs nowadays are quite affordable to virtually all research groups and individuals.

On the other hand, the Intel Xeon Phi accelerators based on its Many-Integrated-Core (MIC) architecture represent a hybridization of classic multi-core CPUs and GPUs and are suitable for speeding up a variety of applications.

1 Topic List

- Commodity Parallel Hardware

- Research Practices of Large-Scale Data Management

- Relational and Non-Relational Data

- OpenMP

- Nvidia CUDA

- Intel TBB based parallel programming techniques

- Parallel Indexing and Query Processing on Multidimensional Spatial and Trajectory Data

  - Grid- and tree-based indexing

  - Selectivity estimation

  - Various types of spatial joins and their optimization

2

- Identifying inherent parallelisms in processing large-scale multi-dimensional data

- High-level parallel primitives

- Multi-core CPUs, GPUs, and Intel MICs

2 Learning Goals

- Be able to understand the variety of kinds of parallel processing

- Be able to understand the different kinds of programming for parallel processing

- Understand how to identify parallelisms in a large multi-dimensional data set

- Be able to solve spatial data problems in a parallel processing environment

3 Assessment

Grading will be based on the attendance, student presentation, homework completion and the final research report and project. Students can work in groups if they desire so for the final project, upon the consent of the instructor. A list of possible topics that would be appropriate for the final project and report can be provided. Students can pick a topic from this list or can also work on any 3D Photography related topic approved by the instructor.

- 50% for group or individual projects

- 30% for presentation(s)

- 20% for class participation