The 2012 International Workshop on Modern Accelerator Technologies for GIScience (MAT4GIScience 2012) is held on September 18, 2012 in Columbus, Ohio, USA in conjunction with the 7th International Conference on Geographic Information Science (GIScience 2012). GIScience community has been an active user of high-performance computing (HPC) resources both for data-intensive and compute-intensive applications. The HPC technology, however, is shifting from homogeneous systems employing identical processing elements to hybrid computing architectures that employ multi-core processors in combination with special-purpose chips. New multi-core architectures combined with application accelerators hold the promise of increasing performance by exploiting levels of parallelism not supported by the conventional systems. Examples of such architectures include Cell-based RoadRunner, hybrid system of combined CPUs and GPUs like Keeneland, and FPGA-based Novo-G systems. However, while there is general trend towards hybrid systems, current GIScience applications require significant algorithmic re-design and software re-engineering efforts to take advantage of the diverse computing resources.

The 2012 International Workshop on Modern Accelerator Technologies for GIScience brings together developers of hybrid computing architectures and computing accelerators and experts involved with the development of applications on such systems to provide a collaborative forum for GIScientists, geographers, geoscientists, and computer scientists to discuss and share the state-of-the-art accelerator technologies and the pioneering applications, and promote collaboration on deploying the modern accelerator technologies in GIScience. The workshop is highlighted by Dr. Piyush Mehrotra from NASA Advanced Supercomputing (NAS) Division as keynote speaker who introduces the programmability and usability of accelerators as well as their performance on NASA relevant benchmarks and applications.

The MAT4GIScience 2012 workshop has accommodated about thirteen papers in two general categories. One is about how the accelerator technologies, such as GPUs, FPGAs, Multi-core & Parallel, and Multi-threading solutions, and heterogeneous computer architecture, are utilized in data and/or computation intensive GIScience applications, geospatial simulation and modeling, and remote sensing, social science research, as well as novel and parallel algorithms for geocomputation, spatial simulation and modeling, performance evaluation and optimization in high performance computing environment. The other is about the research challenges arising in GIScience, geography, geosciences, spatial simulation and modeling, and remote sensing that establishes the needs for high performance computing. By identifying the performance bottleneck in large scale geocomputation, the workshop promotes future collaboration on deploying modern accelerator technologies for high performance computing to advance fact-proven GIScience on the one hand, and advance the research in computational science and engineering, high performance computing, and geospatial algorithm development and optimization on the other.
Performance of NASA Applications on Modern Accelerator Technologies

Dr. Piyush Mehrotra
NASA Advanced Supercomputing (NAS) Division
NASA Ames Research Center, Moffett Field, CA
piyush.mehrotra@nasa.gov

The current trend in high performance computing systems is towards a cluster of multicore nodes enhanced with accelerators. These accelerators utilize lots of smaller multithreaded cores to provide greater computational capability at a much lower power draw. Examples of such accelerators include, NVIDIA’s General Purpose Graphics Processing Unit (GPGPU) and Intel’s Many Integrated Core (MIC) Architecture. The hybrid nature of such systems presents a major challenge to software developers. Both low-level and high-level programming models have been developed to address complex hierarchical structures at different hardware levels and to ease the programming effort. However, achieving the desired performance goal is still not a simple task.

At the NASA Advanced Supercomputing (NAS) Division, we have been experimenting with porting and optimizing several codes to such architectures. In this talk, we summarize our experiences focusing on the programmability and usability of accelerators as well as their performance on NASA relevant benchmarks and applications.
Parallel Primitives based Spatial Join of Geospatial Data on GPGPUs

Jianting Zhang
Department of Computer Science
The City College of the City University of New York
email: jzhang@cs.ccny.cuny.edu

Abstract. Modern GPU architectures closely resemble supercomputers. Commodity GPUs that have already been equipped with personal and cluster computers can be used to boost the performance of spatial databases and GIS. Parallel primitives refer to a collection of fundamental algorithms that can be run on parallel machines. In this study, we report our preliminary work on designing and implementing a spatial join algorithm on GPUs by using generic parallel primitives that have been well understood and efficiently implemented in many parallel libraries. In addition to help understand the inherent data parallelisms in spatial join operations, our experiments have shown that the reference implementation is able to achieve a 6.7X speedup when compared to an optimized CPU serial implementation. We believe that the implementations of parallel spatial algorithms on top of generic parallel primitives can be an important first step towards designing and developing high performance spatial-specific parallel primitives to make it easier to build parallel spatial databases and GIS.
Visualizing 3D/4D Environmental Big Data Using Manycore Compute Unified Device Architecture (CUDA) and Multi-core Central Processing Unit (CPUs)

Jing Li, Yunfeng Jiang, Chaowei Yang*, Qunying Huang
Center for Intelligent Spatial Computing, and
Department of Geography and GeoInformation Science,
MS 6A2, College of Science,
George Mason University, Fairfax, VA, 22030-4444
*Corresponding Author
{jlih, yjiang7, cyang3, qhuang1}@gmu.edu

Abstract. Visualizing 3D/4D environmental Big Data is critical to understand and predict environmental phenomena for relevant decision making. This research explores how to best utilize Graphics Process Units (GPUs) and Central Processing Units (CPUs) collaboratively to speed up the visualization process. Taking the visualization of dust storm as an example, we developed a systematic visualization framework. To compare the potential speedup of using GPUs versus that of using CPUs, we implemented visualization components based on both multi-core CPUs and many-core GPUs. We found that 1) multi-core CPUs and many-core GPUs can improve the efficiency of mathematical calculations and graphics rendering using multithreading techniques; 2) when increasing the size of blocks of GPUs for reprojecting, interpolating and rendering the same data, the executing time drops consistently before reaching a peak; 3) GPU based implementations is faster than CPU-based implementations. However, the best performance of rendering with GPUs is very close to that with CPUs. Therefore, visualization of 3D/4D environmental
A Parallel Algorithm to Solve Near-Shortest Path Problems on Raster Graphs

F. Antonio Medrano and Richard L. Church
Dept. of Geography, Univ. of California at Santa Barbara
{medrano, church}@geog.ucsb.edu

Abstract. The Near-Shortest Path (NSP) algorithm (Carlyle and Wood, 2005) has been identified as being effective at generating sets of good route alternatives for designing new infrastructure. While the algorithm itself is faster than other shortest path set approaches including the Kth shortest path problem, the solution set size and computation time grow exponentially as the problem size increases, and requires the use of high-performance parallel computing to solve for real-world problems. We present a new breadth-first search parallelization of the NSP algorithm. Computational results and future work for parallel efficiency improvements are discussed.
CUDA-Accelerated HD-ODETLAP:
Lossy High Dimensional Gridded Data Compression

W. Randolph Franklin, You Li, Tsz-Yam Lau, and Peter Fox
Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY, USA 12180
mail@wrfranklin.org, liy13@cs.rpi.edu, laut@cs.rpi.edu, pfox@cs.rpi.edu

Abstract. We present High-dimensional Overdetermined Laplacian Partial Differential Equations (HD-ODETLAP), a high dimensional lossy compression algorithm and CUDA implementation that exploits data correlations across multiple dimensions of gridded GIS data. Exploiting the GPU gives a considerable speedup. In addition, HD-ODETLAP compresses much better than JPEG2000 and 3D-SPIHT, when fixing either the average or the maximum error.
Pursuing Spatiotemporally Integrated Social Science over Cyberinfrastructure

Xinyue Ye¹ and Xuan Shi²

¹ Center for Regional Development (Department of Commerce EDA University Center)
School of Earth, Environment and Society, Bowling Green State University
Bowling Green, OH 43403
Email: xye@bgsu.edu

² Department of Geosciences, University of Arkansas
Fayetteville, AR 72701
Email: xuanshi@uark.edu

Abstract. The rich details of space-time complexity in social science remain largely unexplored because of the challenge of data intensive computing. Hence, the current space-time simulation and statistics for social science research can only deal with a limited amount of data. We introduce a pilot study about how to deploy the modern accelerator technology and hybrid computer architecture and systems to extend the National Institute of Justice-funded Near-repeat calculation. This pilot study demonstrates the promising feature of high performance computing on the solution of large-scale space-time interaction, which is a challenging statistical issue in forming spatiotemporally integrated social science.
Simulation and analysis for cluster-based caching replacement based on
temporal and spatial locality of tile access

Rui Li*, Xinxing Wang, Jingjing Wang, Huayi Wu
State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing,
Wuhan University,
Wuhan, Hubei, P.R.China, 430079
RuiLi@whu.edu.cn

Abstract. Cluster-based caching system can accelerate uses’ access in large-scale network service. One of the difficult points for realizing a cluster-based caching system is how to configure all kinds of parameters, to make cluster-based caching servers cooperate with each other, to share cached data, and to obtain the optimal performance of system. With the trace-driven experiment based on the log files from the digital earth, this paper analyzes characteristic of access to tiles in network geography information systems, and simulates cluster-based caching and the collaborative approach between cluster-based caching servers, quantitatively analyzes each parameter of the cluster-based caching system, and obtain the global optimal parameter combination. It provides a reference for realize cluster-based caching system in network geography information application, to improve the quality of service for network geography information systems.
Large-Scale Pulse Compression for Costas Signal with GPGPU

ZHOU Bin¹, YEH Chun-mao², LI Wen-wen³, ZHANG Wei-jie⁴

¹Institute of Oceanographic Instrumentation, Shandong Academy of Sciences, Qingdao, China
synosy@gmail.com

²Beijing Institute of Radio Measurement, Beijing, China
chunmaoyeh@gmail.com

³GeoDa Center for Geospatial Analysis and Computation, Arizona State University, Tempe AZ
wenwen@asu.edu

⁴Science and Technology on Electronic Information Control Laboratory, Chengdu, China
arrowbit@gmail.com

Abstract. Costas signal has been widely acknowledged as having superb ambiguity properties, and is able to provide good time and frequency resolution simultaneously. Consequently, this kind of signal is widely adopted in modern radar systems, especially for low interception performance. However, the pulse compression for Costas signal requires multi-velocity-channel processing. Hence the computational load is increased accordingly. In this paper, a pulse compression scheme is proposed for Costas signals by taking the advantage of “General Purpose Graphics Processing Unit (GPGPU)”. The performance, as well as the effectiveness, is validated by experiments. This processing scheme will provide a good reference for the design of radar signal processors with high computing performance.
Parallelizing ISODATA Algorithm for Unsupervised Image Classification on GPU

Fei Ye\textsuperscript{1} and Xuan Shi\textsuperscript{2}

\textsuperscript{1} School of Computational Science and Engineering
Georgia Institute of Technology
266 Ferst Dr. NW, Atlanta, GA 30332
Email: yefei@gatech.edu

\textsuperscript{2} Department of Geosciences
University of Arkansas
Fayetteville, AR 72701, USA
Email: xuanshi@uark.edu

Abstract. Iterative Self-Organizing Data Analysis Technique Algorithm (ISODATA) is commonly used for unsupervised image classification in remote sensing applications. Although parallelized approaches were explored, previous works mostly utilized the power of CPU clusters. We deploy the many-cores in the Graphics Processing Unit (GPU) to accelerate the unsupervised image classification over GPU. The proposed solution is scalable and satisfactory to speed up the computational time, while the quality of classification is almost the same as that from ERDAS, a well know remote sensing software.
A High-Concurrency Web Map Tile Service Built with Open-Source Software

Huayi Wu, Xuefeng Guan*, Chao Li, Tianming Liu, Jin Li
State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing,
Wuhan University, 129 Luoyu Road, Wuhan 430079, P. R. China
wuhyai@whu.edu.cn, guanxuefeng@whu.edu.cn

Abstract. As an Open Geospatial Consortium (OGC) standard service, Web Map Tile Service (WMTS) has been widely used in many fields to fast and conveniently share geospatial information with the public. However, in practice when client users increase dramatically, the torrent requests from users bring overwhelming pressure to the web server where WMTS is deployed, and then cause significant response delay and serious performance degradation of WMTS. It is of great necessity to study the architecture of high-concurrency WMTS which can automatically scale with the requests of client users. In order to tackle this problem, this paper proposes a prototype of one high-concurrency WMTS, which is totally built with excellent open-source software, such as Nginx, GeoWebcache, and MongoDB. Several experiments were carried out to test the efficiency and scalability of the proposed high-concurrency WMTS with Web-bench. The results illustrate that our WMTS can still function well when enduring more than 30,000 concurrent connections. The request throughput of the proposed high concurrency WMTS is twice as large as that of traditional WMTS deployed in a single web server.
Opportunities and Challenges for Urban Land-use and Land-cover Change Modeling in High-performance Computing

Qingfeng Guan\textsuperscript{1}, Xuan Shi\textsuperscript{2}

\textsuperscript{1} Center for Advanced Land Management Information Technologies  
School of Natural Resources  
University of Nebraska – Lincoln  
3310 Holdrege Street, Lincoln, NE 68583, USA  
Email: qguan2@unl.edu

\textsuperscript{2} Department of Geosciences  
University of Arkansas  
Fayetteville, AR 72701, USA  
Email: xshi@cast.uark.edu

Abstract. Simulating urban land-use and land-cover change (LULCC) has been a challenging theme because of the spatial and temporal complexity of interrelationships and interactions between the urban land system and related natural and socio-economic systems. The difficulty of modeling urban LULCC can be aggravated by the massive computational intensity caused by complicated algorithms and large data volume that are often required in the simulation. In order to reduce both the modeling and computational complexities in spatio-temporal simulations, researchers often had to make subjective and/or simplifying assumptions. However, such simplifying approaches raised some serious scientific questions in regard to the validity and soundness of the findings resulted from these models, because whether these assumptions could generate reliable calibration and simulation results and lead to unbiased and accurate scientific conclusions has not been sufficiently studied yet. To investigate the potential problems and advance our understanding and theories of urban land dynamics, we must device approaches to reduce or even eliminate these assumptions. Recent advancements in high-performance computing (HPC) infrastructure provide promising solutions to the above problems. Emerging advanced computing technologies, including Graphics Processing Units (GPUs), and heterogeneous computer systems that combine GPU accelerators and multi-core CPUs, have been significantly improving the performance of scientific computation in a variety of domain knowledge. Therefore it is time for geographers and spatial scientists to examine, validate and advance urban LULCC theories as the technological solutions and computing infrastructure are increasingly mature and efficient for such kind of investigations.
Parallelizing Max-P-Regions Spatial Clustering Method in PySAL

Xing Kang\textsuperscript{1} and Sergio J. Rey\textsuperscript{1}
\textsuperscript{1} GeoDa Center for Geospatial Analysis and Computation, School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, AZ 85287-5302.
E-mail: \{xkang4, srey\}@asu.edu

Abstract. In this paper, we report on a series of experiments applying parallelization techniques to optimize PySAL, a Python library with a rich set of spatial analysis methods. Specifically, we focus on the Max-P-Regions spatial clustering method which forms mutually exclusive, exhaustive, and spatially contiguous groups of regions given a set of areas so that the regions satisfy a minimum size criterion but also minimize intraregional homogeneity. We examine the performance of parallelization through comparing various parallelized implementations with the original sequential version on a multi-core desktop environment. Our experiment consists of a range of sample sizes for synthetically created data and we consider the tradeoffs between solution time and the quality of the partition solution in terms of overall regional homogeneity.
Accelerating Mean Shift Segmentation Algorithm on Hybrid CPU/GPU Platforms

Liang Men, Miaqing Huang, John Gauch
Department of Computer Science and Computer Engineering
University of Arkansas
{mliang, mqhuang, jgauch}@uark.edu

Abstract. Image segmentation is a very important step in many GIS applications. Mean shift is an advanced and versatile technique for clustering-based segmentation, and is favored in many cases because it is non-parametric. However, mean shift is very computationally intensive compared with other simple methods such as $k$-means. In this work, we present a hybrid design of mean shift algorithm on a computer platform consisting of both CPUs and GPUs. By taking advantages of the massive parallelism and the advanced memory hierarchy on Nvidia's Fermi GPU, the hybrid design achieves a $20\times$ speedup compared with the pure CPU implementation when dealing with images bigger than $1024\times1024$ pixels.
Use Geosimulation Data to Assess the Inferential Power of Statistics

Ninghua (Nathan) Wang
Joint Ph.D. Program in Geography
San Diego State University and UC Santa Barbara
Email: wagn@rohan.sdsu.edu

Li An
Department of Geography
San Diego State University
Email: lan@mail.sdsu.edu

Abstract. Inferential statistical models have been long used to uncover underlying mechanisms of land-cover change processes. For example, Allen and Barnes (1985) used a linear regression to identify causes of deforestation in developing countries, and Wu and Yeh (1997) employed a logistic regression to evaluate driving factors in urbanization. Innovative methods (e.g. survival analysis) were also introduced in response to the proliferation of spatial-temporal data (An and Brown 2008). While valuable insights can those statistical models bring to us, their usage is subject to uncertainty. Most land-cover change processes involve complex effects, such as feedbacks, chaos, and emergence (Lambin and Geist 2006), and it remains unclear how statistical models function in presence of those effects; second, most statistical models are borrowed from other fields so their assumptions may not be fulfilled in geographical applications; furthermore, data availability in terms of spatial and temporal resolutions may severely constrains the performance of certain models. In short, a thorough assessment of the inferential power of statistical models seems necessary. However, traditional assessment based on a few empirical case studies is inadequate, because of unknown underlying mechanism, limited conditions, and data incompleteness. A novel approach is thus called for.