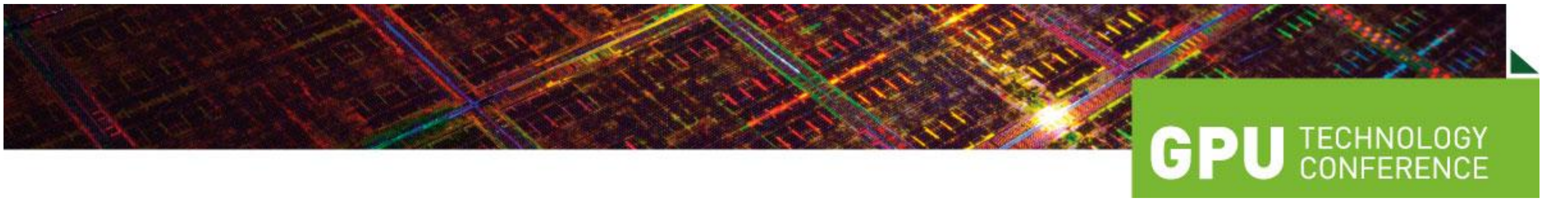

GPU TECHNOLOGY CONFERENCE

GPU Technology Conference, May 14-17, 2012
McEnery Convention Center, San Jose, California
www.gputechconf.com

Sessions on Machine Learning & AI (subject to change)

IMPORTANT: Visit http://www.gputechconf.com/page/sessions.html for the most up-to-date schedule.

S0223 - Rapid Training of Acoustic Models Using GPUs
Jike Chong (Carnegie Mellon University), Ian Lane (Carnegie Mellon University Co)
Day: Tuesday, 05/15 | **Time:** 3:00 pm - 3:50 pm
Topic Areas: Audio, Image and Video Processing; Machine Learning & AI
Session Level: Intermediate

Learn how to realize robust and accurate speech recognition systems by training acoustic models on GPUs. For common languages, state-of-the-art systems are now trained on thousands of hours of speech data, which can take weeks even with a large cluster of machines. To overcome this development bottleneck, we propose a new framework for rapid training of acoustic models using highly parallel GPUs. With a single NVIDIA GTX580 GPU, our proposed approach is shown to be 51x faster than a sequential CPU implementation, enabling a moderately sized acoustic model to be trained on 1000-hour speech data in just over 9 hours.

S0314 - Efficient k-Nearest Neighbor Search Algorithms on GPUs
Nikos Pitsianis (Aristotle University, Greece), Xiaobai Sun (Duke University)
Day: Tuesday, 05/15 | **Time:** 4:30 pm - 4:55 pm
Topic Areas: Machine Learning & AI; Databases, Data Mining, Business Intelligence; Algorithms & Numerical Techniques
Session Level: Beginner

Come see how to select the k smallest elements from an unsorted list. We present a selection and combination of different algorithms that perform exact k-nearest neighbors search (k-NNS) on GPUs and outperform the competition. In this session we present four different selection algorithms designed to exploit differently the parallelization of the GPU according to the relative size of the corpus data set, the size of the query set and the number of neighbors sought. We show the application of Logo Retrieval with SIFT vector matching on two different GPUs, the Tesla C1060 and the Fermi GTX480.

S0052 - Fast High Quality Image and Video Background Removal with CUDA
Timo Stich (NVIDIA)
Day: Wednesday, 05/16 | **Time:** 4:30 pm - 4:55 pm
Topic Areas: Audio, Image and Video Processing; Machine Learning & AI
Session Level: Intermediate

S0133 - Improving Mars Rover Image Compression Via GPUs and Genetic Algorithms

Brendan Babb (University of Alaska Anchorage)
Day: Thursday, 05/17 | **Time:** 9:00 am - 9:25 am
Topic Areas: Machine Learning & AI; Audio, Image and Video Processing; Development Tools & Libraries
Session Level: Beginner

Learn how to use Jacket to accelerate genetic algorithm (GA) image compression. Our research uses a GA to optimize lossy compression transforms that outperform state-of-the-art wavelet-based approaches for a variety of image classes, including fingerprints, satellite, medical, and images transmitted from the Mars Exploration Rovers. A typical training run evolves a population of transforms over many generations; since each transform must be applied to each image from the training set, each run entails thousands of independent, parallelizable fitness evaluations. By using MATLAB, and Jacket to perform 2D convolution on the GPU, we have greatly reduced the total computation time needed.

S0038 - Designing Killer CUDA Applications for X86, multiGPU, and CPU+GPU

Robert Farber
Day: Thursday, 05/17 | **Time:** 4:00 pm - 4:25 pm
Topic Areas: Machine Learning & AI; Supercomputing; Databases, Data Mining, Business Intelligence; Computer Vision
Session Level: Intermediate

CUDA redefined software development with 10 to 1000-times faster GPU applications. Now a single CUDA source tree can support the x86 mass market (no GPU required) and 1/3 billion CUDA-enabled GPUs. MultiGPU and CPU+GPU apps utilize all system resources. GPUdirect, UVA, caches, prefetching, ILP (Instruction level Parallelism), automated analysis tools and more offer ease, capability, and performance. The overall impact on software investment, scalability, balance metrics, programming API, and lifecycle will be considered. Working real-time video and other examples from my book, “CUDA Application Design and Development” provide practical insight to enable augmented reality and your killer apps.