Visualizing Ultra-Scale Data

Kwan-Liu Ma
University of California at Davis

Parallel supercomputers enable scientists to model complex physical phenomena and chemical processes in efforts to answer some of the most important and difficult questions in science. The output from these simulations is so voluminous and complex that advanced visualization technologies are necessary to interpret the calculated results. Even though visualization technology has progressed significantly in recent years, we are barely capable of visualizing and analyzing terascale data to its full extent, and petascale datasets are on the horizon. The Institute for Ultrascale Visualization (http://ultravis.org), funded by the US Department of Energy, is a 5-year project involving several universities and national laboratories to address the upcoming petascale visualization challenges facing computational science and engineering. The Institute is expected to revolutionize the very process of scientific discovery by equipping scientists with tools that shed light on the knowledge hidden in previously incomprehensible datasets.

This talk will present four key technologies: parallel visualization, knowledge-based visualization, in situ visualization, and visualization interfaces. Important advancements made by the Institute and best practices will be introduced. The objective is to foster exchange between the visualization and SIGGRAPH communities.

Parallel visualization is absolutely needed if the goal is to see the full extent of the data at the highest possible fidelity, but not without its own challenges especially across our diverse scientific user community. The Institute brings together leading experts from visualization, high-performance computing, and science application areas to make parallel visualization technology a commodity for SciDAC scientists and the broader community.

Most of the data sets can be reduced to some degree without removing the most important information in the data. While we can always employ data compression methods such as quantization and transform coding, a more effective way to reduce data is physically based. That is, with scientists’ knowledge about some specific feature of interest, it is also possible to use that knowledge to effectively reduce the data. The reduction may significantly facilitate subsequent data transfer and visualization. We can consider this as knowledge-assisted visualization.

So far, data visualization has been almost exclusively done as a post-processing step. For runtime visualization, the cost of moving the simulation output to a visualization machine could be too high to make interactive visualization feasible. A better approach is not to move the data, or to keep the data that must be moved to a minimum. That is, both simulation and visualization calculations run on the same parallel supercomputer so the data can be shared. Such in situ processing can render images directly or extract features, which are much smaller than the full raw data, to store for on-the-fly or later examination. As a result, reducing both the data transfer and storage costs early in the data analysis pipeline can optimize the overall scientific discovery process. The petascale challenge can be best addressed with in situ visualization.

Over the past twenty years, many novel visualization techniques have been invented but few have been deployed in production systems and tools. Why? The UI aspect of these techniques were often neglected, and the results were tools that are hard to use. The Institute has initiated several projects on the design and deployment of appropriate user interfaces for advance visualization techniques. In addition, it becomes clear that next generation visualization technology should be built upon further exploitation of human perception to simplify visualization, advanced hardware features to accelerate visualization calculations, and also make use of machine learning to reduce the complexity, size, and high-dimensionality of data.

We are not too far from petascale and exascale computing. The investment made by the DOE SciDAC program in ultra-scale visualization is timely and ensures that challenges will be addressed. The Ultravis Institute has made some success in creating and deploying new visualization technologies. This presentation calls for further research and experimental studies involving a greater community.