LBNL/PUB-886

Visual Data Exploration and Analysis – Report on the Visualization Breakout Session of the SCaLeS Workshop

Wes Bethel, LBNL (Co-Chair)
Randy Frank, LLNL
Sam Fulcomer, Brown University
Chuck Hansen, University of Utah (Co-Chair)
Ken Joy, UC Davis
Jim Kohl, ORNL
Don Middleton, NCAR

June 24-25, 2003

1 What is Visualization?

Scientific visualization is the transformation of abstract information into images, and it plays an integral role in the scientific process by facilitating insight into observed or simulated phenomena. Visualization as a discipline spans many research areas from computer science, cognitive psychology and even art. Yet the most successful visualization applications are created when close synergistic interactions with domain scientists are part of the algorithmic design and implementation process, leading to visual representations with clear scientific meaning. Visualization is used to explore, to debug, to gain understanding, and as an analysis tool. Visualization is literally everywhere – images are present in this report, on television, on the web, in books and magazines – the common theme is the ability to present information visually that is rapidly assimilated by human observers, and transformed into understanding or insight.

2 Visualization Impact

As an indispensable part a modern science laboratory, visualization is akin to the biologist's microscope or the electrical engineer's oscilloscope. Whereas the microscope is limited to small specimens or use of optics to focus light, the power of scientific visualization is virtually limitless: visualization provides the means to examine data that can be at galactic or atomic scales, or at any size in between. Unlike the traditional scientific tools for visual inspection, visualization offers the means to "see the unseeable." Trends in demographics or changes in levels of atmospheric CO_2 as a function of greenhouse gas emissions are familiar examples of such unseeable phenomena.

Over time, visualization techniques evolve in response to scientific need. Each scientific discipline has its "own language," verbal and visual, used for communication. The visual language for depicting electrical circuits is much different than the visual language for depicting theoretical molecules or trends in the stock market. There is no "one visualization tool" that can serve as a panacea for all science disciplines. Instead, visualization researchers work hand in hand with domain scientists as part of the scientific research process to define, create, adapt and refine software that "speaks the visual language" of each scientific domain.

3 Visualization Research Areas

The sections that follow present a number of visualization research topics. They are a blend of computer science technologies for realizing needed growth in visualization capacity and capability, as well as new visualization technologies that address specific science needs. The challenges posed by modern computational science performed on large-scale computer systems are acute: not only is the amount of data becoming larger, the complexity of the data itself is growing. Due to their fundamental design, visualization tools from earlier periods simply do not exhibit the capacity to process large scientific data sets. Similarly, the capabilities of earlier tools are not adequate to effectively present the meaningful information inherent in large, multidimensional data.

3.1 High Capacity Visualization

One of the most significant challenges facing visualization is the need to process and display very large scientific datasets. Significant early advances by the visualization community in this area have identified areas requiring research to meet the needs of emerging computational science programs. One such area is data models and algorithms for processing and visualizing time-varying data. Technologies already used to accelerate static data processing, i.e. multiresolution representations, will benefit access and processing of dynamic, time-varying data, thereby increasing the power of tools available to scientific researchers. New, related visualization technology that focuses upon effective visual display of time-varying data will enable better scientific understanding of complex dynamic phenomena. Achieving these objectives requires careful attention to the architecture of pipelined and parallel visualization processing tools, along with effective use of high-resolution displays.

3.2 Remote Visualization

Remote visualization is an integral part of all our lives. When we watch the weather forecast on television, we are viewing a presentation of data assembled from a number of remotely located sources: satellite imagery, regional ground-based stations, weather balloon observations and computer simulations that predict tomorrow's weather. This same metaphor applies to modern computational science, where large datasets are generated on supercomputers and are analyzed or viewed by remotely located researchers . The trend towards consolidated centers that provide extreme computing capabilities as centralized resources, combined with the increased size of generated data, produce an acute need for remote visualization capabilities. As research teams are increasingly composed of geographically distributed scientists, interactive collaborative remote visualization technologies can help to accelerate scientific discovery while reducing the costs associated with travel.

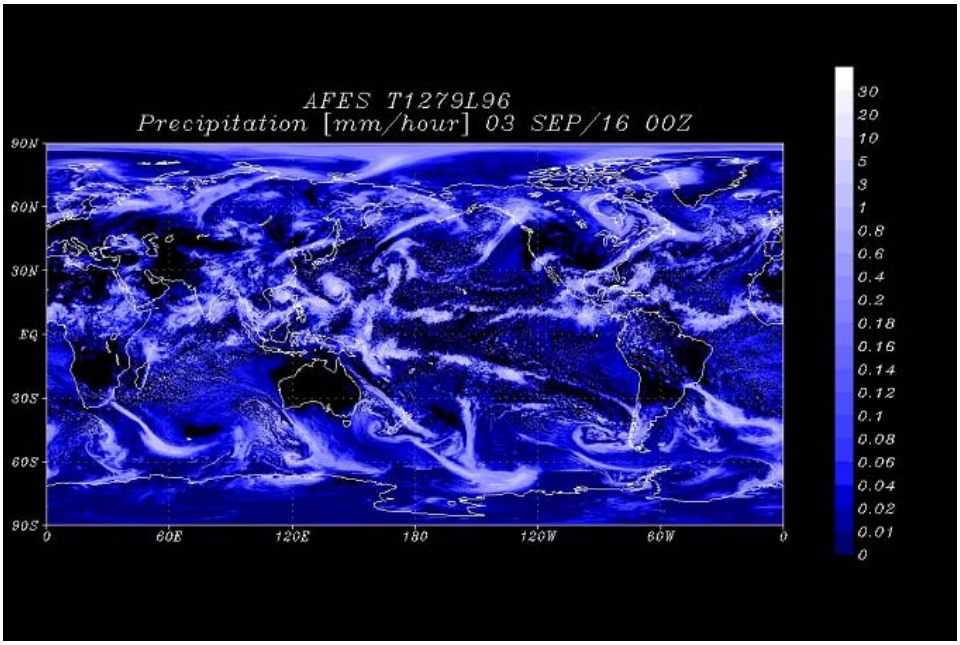

Figure 1. High-resolution datasets are computed at centralized facilities, but viewed by remotely located researchers. Remote visualization techniques help scientists make effective use of centralized facilities. *Don Middleton, NCAR.*

3.3 Multiresolution Methods

One avenue for addressing the problems posed by remote visualization is to enable the researcher to examine data at different resolutions. A quick examination of a low-resolution model or a statistical summary might reveal that no further inspection is necessary, thereby resulting in a significant time and resource savings. Alternately, a low-resolution model can provide a visual roadmap for high-resolution exploration, allowing a researcher to select small, high-resolution subsets of a dataset for more thorough analysis. Creating such multiresolution representations for specific scientific domains is a research area unto itself. However, creating effective methods for visually presenting such multiresolution representations and enabling the interactive transition between visual depictions are both active areas of visualization research.

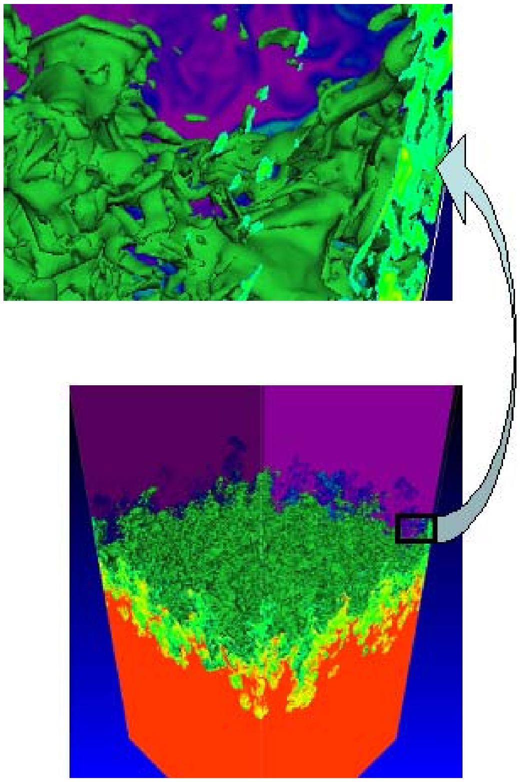

Figure 2. Multiresolution visualization requires specialized data models. *Randy Frank, LLNL.*

3.4 Multidimensional and Multivariate Visualization

Scientific computing has evolved to simulate phenomena at ever-increasing levels of fidelity and accuracy. Accurate modeling of phenomenon often requires solving for more and more unknown variables. To facilitate scientific advances and provide insight into these complex systems, visualization technology is needed that can effectively display many variables simultaneously. The visualization challenge is compounded by the scientific need for simultaneous visual comparisons of experimental and simulation data, as well as data obtained or computed over a period of time.

3.5 Coupling Analysis, Visualization and Data Management

At the core of the visualization processing pipeline is technology for accessing, manipulating and processing data. As data models and data management systems evolve to accommodate ever-increasing dataset sizes and locations, there is a corresponding need for visualization tools to take advantage of these emerging technologies that store, retrieve, characterize and analyze data. Statistical analysis forms an integral part of data understanding, yet there exist few techniques for visualization of uncertainty and other statistical features. Identification and characterization of "interesting features" is highly domain specific. Automatic detection and display of such features is a blend of statistical analysis, data management and domain-specific visualization techniques. Through advances in visualization technology that include closer ties to data management technology, including processing and display of statistical information, computational science programs benefit from increased visual data analysis capacity and capability.

3.6 "Behavioral" Visualization

As computer simulations increase in complexity, there is a growing need for visual representations of complex processes. One example is the behavior of optimization calculations in combinatorial algorithms. Visualization of algorithmic behavior, decision trees, and related "behavioral processes" provide insight into the operation and improvement of complex scientific software. A good example is how the search space in protein conformation is pruned to identify minimal energy conformations in complex molecules. Another example is the visual display of chemical pathways in combustion simulations, or metabolic pathways in cells. The evolution of simulation programs requires new visualization techniques to facilitate scientific insight.

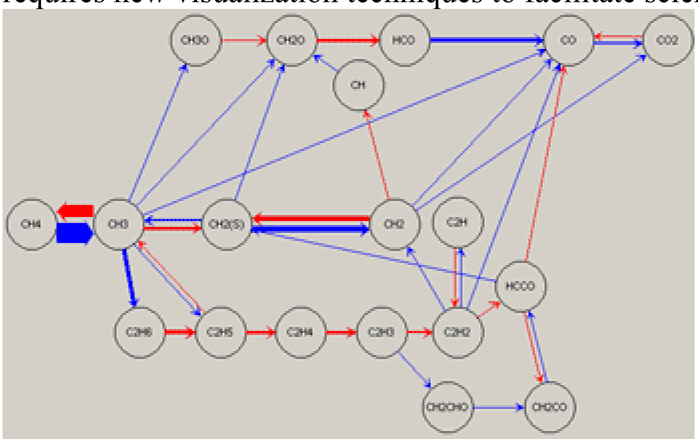

Figure 3. Chemical Pathway Visualization. The nodes represent species, and the edges represent flow of a conserved quantity, such as transfer of a particular element. *Mark Day, LBNL.*

4 Delivering Visualization Technology to Application Scientists

Application scientists have indicated that the best software tools are those specifically tailored for their domain. Such tools provide results in a familiar "language" that are readily comprehensible and applicable to scientific research. To develop such tools, visualization researchers must be part of the multidisciplinary science team performing the research. Even though each discipline needs tailored software tools, careful general-purpose software design and implementation will result in a "toolbox" of compatible components that can be combined in various ways to provide domain-specific solutions. Such components, with supporting data models, provide the "standards" to which disparate teams of visualization and science researchers can create compatible software tools. The evolution of a community-defined and supported software technology base will accelerate the growth of visualization research and its application to scientific domains through reduced duplication of effort and software engineering practices that promote reuse.

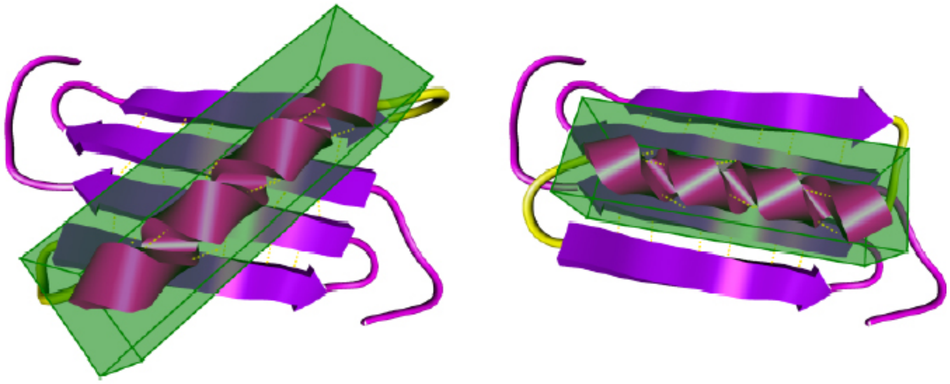

Figure 4. Visualization and manipulation of protein molecules is performed using "units" familiar to computational biologists – alpha helices and beta sheets. *Oliver Kreylos, UC Davis/LBNL; Silvia Crivelli, LBNL; Nelson Max, UC Davis, LLNL and LBNL. W. Bethel, LBNL, B. Hamann, UC Davis/LBNL.*

5 Resources Required (and Barriers)

The current model for funding visualization research and development tends to emphasize technology demonstrations. In contrast, science researchers need stable, production quality software. The cost of ongoing software maintenance, documentation, training and evolution far exceeds the cost of initial research and development. However, there is no funding mechanism to sustain these crucial activities. The traditional economic model of technology transfer from research into commercial products does not apply to scientific software, particularly visualization. The size of the consumer market for such specialized products and services simply does not exist.

Scientific visualization also places extreme demands upon computing infrastructure. All aspects of the computing pipeline are subject to significant demands for multi-terabyte datasets: storage systems that serve as repositories ; CPU and memory systems that process the data; networks that transport the data, and graphics systems that display it. The same maladies that plague the general scientific computing hardware market are present in the high performance graphics and visualization world: the needs of the scientific visualization community are largely ignored by graphics hardware manufactures. Those vendors instead respond to the needs of the computer gaming industry, which uses only benchmarks that measure the number of frames per second generated when playing one of several different computer games. These ratings do not correlate to scientific visualization needs.

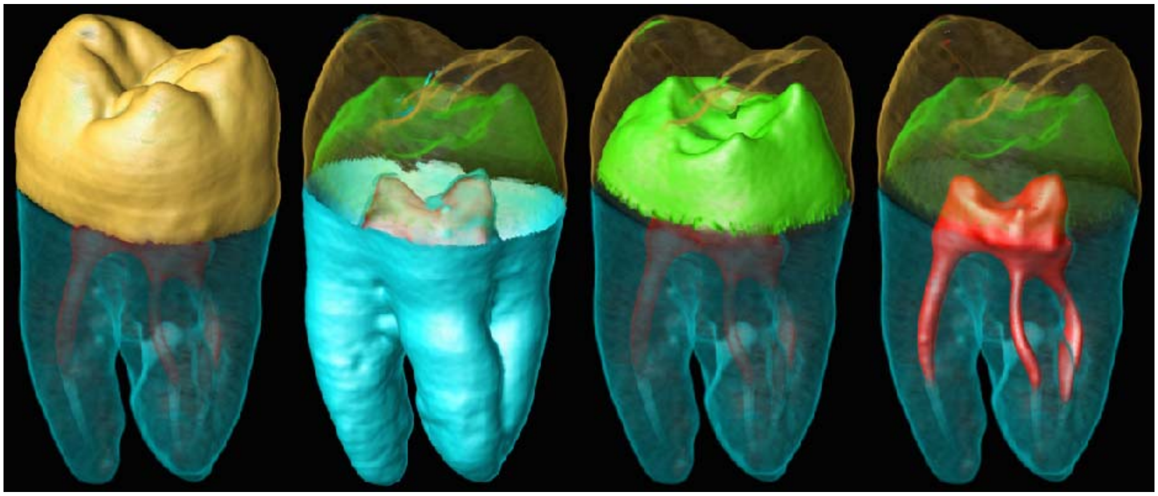

Figure 5. Advanced rendering features like three dimensional transfer functions are not provided by graphics hardware vendors because they aren't used by computer games. *Chuck Hansen, University of Utah.*

Given the central role of the remote visualization metaphor in modern scientific computing, there is an alarming lack of networking capacity to connect remote users with centralized facilities. Large-scale computer systems provide massive computational capacity, but are often linked to the outside world using networks of inadequate capacity. Commodity Gigabit Ethernet hardware for desktop platforms is very inexpensive, yet the networks connecting major sites typically can support only two and a half such users operating at full capacity. Beyond the trunk lines themselves is the acute need for hardware that connects sites to the network. Effective use of centralized facilities requires high-speed network connectivity to deliver results to remotely located researchers.

When designing large-scale platforms, the needs of computational science research programs are taken into account by considering grid resolution, number of unknowns, number of time steps, and related variables to estimate the approximate amount of computing power required for a given class of algorithms. On the other hand, visualization processing is typically delegated to relatively small computing platforms that have nowhere near enough computing power. A disparity of several orders of magnitude in computing power is typical – simulations are run on platforms that can reach tens of teraflops, yet visualization is delegated to machines that are capable of only a few gigaflops. A substantial increase in funding for visualization computing platforms is critical to "impedance match" the capacity of simulation and analysis platforms. Similarly, an increase in visualization research staffing is needed to support projected growth trends to meet the needs of science research programs.

6 Metrics of Success

Visualization success can be characterized using several metrics. First and foremost is the degree to which visualization helps advance science. The most obvious metric is the number of scientific discoveries facilitated by visualization. However, achieving these discoveries requires close coupling between visualization and scientific researchers so that visual data analysis tools are effectively designed and applied. Another metric is longevity, or the temporal lifetime of visualization technology. The current visualization funding model encourages exploration of ideas, but does not provide for the critical ongoing maintenance and lifecycle support activities needed to ensure that today's research prototypes form the basis for tomorrow's staple software tools. Another metric is the degree to which visualization, analysis, and data management are interoperable. Future research programs in visualization must include interoperability as a central theme to promote both widespread usage by a large population as well as longevity.

7 Acknowledgement

This work was supported by the Office of Science, Computational and Technology Research, U. S. Department of Energy under Contract No. DE-AC03-76SF000098.