



Guest Editor's Introduction

Large-Scale Data Visualization

Kwan-Liu Ma

University of
California, Davis

Scientists nowadays have unprecedented computing and instrumental capability for studying natural phenomena at greater accuracy, resulting in an explosive growth of data. For example, the data generated by modeling next-generation accelerators can have from hundreds of millions to billions of particle paths. For Earth sciences turbulence calculations, each run can produce thousands of time steps of 1024^3 volume data. The increasing resolution of medical imaging instrumentation has resulted in data of unprecedented size, as demonstrated by the National Library of Medicine's Visible Human project. A new sensor technology known as MEMS (microelectromechanical systems) will constantly produce massive data streams that we don't know how to consume yet.

Since the publication of the National Science Foundation panel report¹ in 1987 recommending a new initiative in visualization in scientific computing, government agencies and universities have invested in tremendous research and development efforts, which have led to many research innovations in scientific visualization. However, after more than 10 years, current data handling and visualization capacities still seem orders of magnitude too small for scientists to interpret the voluminous and complex data they're capable of producing routinely.

The Department of Energy's Accelerated Strategic Computing Initiative (ASCI), which seeks to develop high-performance modeling and simulation tools for stockpile certification, has driven a second push for advancing visualization technology. Under the ASCI program, the most powerful parallel supercomputers ever built are now operated at several of the DOE labs. Each of the five ASCI alliance programs (dynamic response of materials, integrated turbulence simulation, astrophysical thermonuclear flashes, simulation of advanced rockets, and simulation of accidental fires and explosions) can produce terascale data beyond the reach of current visualization tools. These ASCI-class projects can lead to revolutionary advances in science and engineering only with appropriate data analysis and visualization support.

Recognizing the urgent need to solve the large data visualization problem, a few years ago the NSF and DOE sponsored a series of workshops on large-scale data management and visualization. Three were held in 1998. A

seminal report titled *Data and Visualization Corridors*² resulted from these workshops. This report provides a five-year roadmap for scientific visualization research and development. Another workshop held in May 1999, organized by Chris Johnson, John Reynders, and I, fostered further exchanges between visualization researchers and application scientists (see <http://graphics.cs.ucdavis.edu/~ma/LDV99>). A few months later, NSF announced an initiative in large scientific and software data-set visualization and subsequently funded 11 projects.

Solving the large-scale data visualization problem requires an integrated systems approach. On a low level, it's critical to develop an efficient data management mechanism coherently, supporting diverse data representations and access patterns of typical visualization calculations. The data management problem is exacerbated for geographically dispersed data. Therefore, we need novel designs in the areas of scalable databases, hierarchical storage systems, and parallel I/O. On a high level, four other areas are worth mentioning here: data reduction, scalable parallel visualization, high-resolution displays, and user interfaces for data visualization.

Data reduction

Generally, we don't need (nor is it feasible) to look at large data at full resolution at all times. In many instances, we'd rather trade resolution for interactivity, which helps us define key visualization parameters for high-fidelity visualizations. Thus, we need techniques to represent data at different resolution levels. A lower resolution version of the data may fit in or be quickly brought into the computer memory so we can efficiently explore the data's temporal, spatial, and parameter spaces. The other approach to data reduction is a pre-processing step that extracts from the data the physically based features that we specify. We can represent these features, if successfully extracted, in an economical way for interactive visualization.

Scalable parallel visualization

To visualize large-scale data at the highest possible resolution, we need the processing power and memory of a parallel computer. A parallel visualization system must be scalable to fully use the massively parallel supercomputers available at various government laborato-

ries and national supercomputing centers. Furthermore, every stage of the visualization pipeline must be parallelized to eliminate any bottleneck created by a serial process. PC clusters, which have become increasingly popular and affordable, make parallel visualization an even more attractive approach. In particular, we can now integrate the parallel visualization capability into the parallel numerical simulation to make runtime tracking of the simulation possible.

High-resolution displays

To visualize data produced from a state-of-the-art turbulence simulation, we need a display space with at least several millions of pixels, which a conventional desktop display can't provide. Over the past two years, developing tiled displays has become an area of active research.³ The increased display space not only permits the full display of high-resolution images but also offers new research opportunities for user-interaction techniques.

User interfaces

Investing computing and human resources in visualization and understanding large scientific data sets guarantees a high return, but it's always desirable to lower our investment. While improving the efficiency of visualization calculations—such as better resource use—can lower the cost, a mostly overlooked area is the design of user interfaces supporting reuse and sharing. We must go beyond the traditional graphical user interface design by coupling it with a mechanism that helps us keep track of our visualization experience, use it to generate new visualizations, and share it with others.⁴ Doing so can reduce the cost of visualization, particularly for routine analysis of large-scale data sets.

In this issue

This special issue features six articles that describe the recent research and experiences in managing and visualizing large-scale data. The first three articles introduce external memory methods for large-scale data visualization. The next two articles describe hardware-assisted approaches using a high-end multiprocessor graphics workstation and a PC cluster, respectively. The last article is a tutorial on parallel software volume rendering, which mainly considers rendering scalability for both large-scale distributed-memory and shared-memory architectures. Although I'm unable to address all the important aspects of the large data visualization problem in this issue, I hope I have drawn your attention to the research opportunities introduced by this challenging problem. ■

Acknowledgments

I would especially like to thank the 41 expert reviewers who helped us select five articles from a very large number of submissions. Thanks also go to Theresa-Marie Rhyne and the CG&A editorial staff. Without them, this special issue would not have been possible.

References

1. B.H. McCormick, T.A. DeFanti, and M.D. Brown, "Visualization in Scientific Computing," *Computer Graphics* (Proc. Siggraph 87), vol. 21, no. 6, ACM Press, New York, 1987.
2. P.H. Smith and J. van Rosendale, eds., "Data and Visualization Corridors," *Report on the 1998 Data and Visualization Corridors (DVC) Workshop Series*, tech. report CACR-164, California Institute of Technology, Pasadena, Sept. 1998, <http://www.cacr.caltech.edu/Publications/DVC>.
3. T. Funkhouser and K. Li, "Large-Format Displays," *IEEE Computer Graphics and Applications*, vol. 20, no. 4, July/Aug. 2000, pp. 20-21.
4. K.-L. Ma, "Visualizing Visualizations—User Interfaces for Managing and Exploring Scientific Visualization Data," *IEEE Computer Graphics and Applications*, vol. 20, no. 5, Sept./Oct. 2000, pp. 16-19.



Kwan-Liu Ma is an associate professor of computer science at the University of California, Davis, where he teaches and conducts research in computer graphics and scientific visualization. His career research goal is to improve the overall experience

and performance of data visualization through more effective user-interface designs, interaction techniques, and high-performance computing. He received his PhD in computer science from the University of Utah in 1993. He served as co-chair for the 1997 IEEE Symposium on Parallel Rendering, Case Studies of the IEEE Visualization Conference in 1998 and 1999, and the first National Science Foundation/US Department of Energy Workshop on Large-Scale Data Visualization. He recently received the Pecase award for his work in parallel visualization and large-scale data visualization.

Readers may contact Ma at the Dept. of Computer Science, Univ. of California, Davis, One Shields Ave., Davis, CA 95616-8562, email ma@cs.ucdavis.edu.

For further information on this or any other computing topic, please visit our Digital Library at <http://computer.org/publications/dlib>.