HIGH-DIMENSIONAL DATA ACQUISITION, COMPUTING, AND VISUALIZATION

Data-centered collaboration on emerging cyberinfrastructures is expected to revolutionize science and engineering. Many experts predict grids of geographically distributed mass storage, high-end computers, and visualization platforms connected via high-speed networks will let scientists and engineers collaborate remotely. They envision a new era of integrated massive data from distributed, diverse resources that will help researchers achieve data-driven discovery and problem solving across multiple spatiotemporal scales and traditional discipline boundaries (see www.cise.nsf.gov/evnt/reports/toc.htm). Data sharing and collaboration on cyberinfrastructures also will democratize science and engineering by diversifying participation from broader institutions and communities.

Central to this revolution are the techniques and tools needed to acquire, share, analyze, visualize, and discover knowledge from distributed,

AIICHIRO NAKANO
University of Southern California
JIM X. CHEN
George Mason University

high-dimensional, information-rich data sets, which are the focus of this special issue. However, these techniques and tools pose enormous technical challenges. Only cross-disciplinary collaboration between application and computer scientists, the latter in the areas of databases, data mining, visualization, high-performance computing, and systems software, will push us to the next level. We hope this issue provides a forum for such scientists to exchange their ideas and project future directions.

The first half of this theme topic appeared in the March/April 2003 issue of *CiSE*, and it dealt with visualization, data mining, and Internet server technologies for large-scale and time-varying data sets in broad application areas such as materials and Earth sciences. The second part expands the scope by featuring three articles covering data mining, visualization, and database technologies applied to astronomy, neuroscience, and genetics.

In "Migrating a Multiterabyte Archive from Object to Relational Databases," Ani Thakar, Alex Szalay, Peter Kunszt, and Jim Gray share lessons learned from porting a commercial object-oriented database engine to relational database technology, with custom tools for data mining the multiterabyte Sloan Digital Sky Survey archive. Such information will be valuable for those involved in similar scientific and engineering data-sharing projects.

"Dimension Reduction and Spatiotemporal Regression: Applications to Neuroimaging," by Kerby Shedden and Ker-Chau Li, discusses methods to characterize spatiotemporal variation in brain activity measurements. The authors use statistical dimension reduction to locate temporal components in data that best preserve the spatiotemporal regression structure.

"Gene Expression Clustering and 3D Visualization," by Yonggao Yang, Jim X. Chen, and Woosung Kim, presents methods for processing DNA microarray data. These methods include a self-organizing map to cluster data and discover gene patterns, principal components analysis to reduce data dimension, and 3D plotting to interactively visualize the reduced data sets. Such a system will provide a useful tool for deciphering the meaning behind the mysterious gene expressions that create life.

We hope these excellent articles will fire your own imagination to enrich your research and application efforts in the emerging data-driven science and engineering focus in cyberinfrastructure.

Aiichiro Nakano is an associate professor of Computer Science with joint appointments in Physics and Astronomy, Materials Science and Engineering, and the Collaboratory for Advanced Computing and Simulations at the University of Southern California. His interests include scalable scientific algorithms, Grid computing on geographically distributed parallel computers, and scientific visualization. He is a member of the IEEE, ACM, APS, and MRS. Contact him at anakano@usc.edu.

Jim X. Chen is an associate professor in the Department of Computer Science at George Mason University. He is currently the director of the Graphics Lab at GMU, the Visualization Corner department editor for *CiSE* magazine, and the program cochair of VR2003. His research interests include graphics, visualization, virtual reality, networking, and simulation. He received his PhD in computer science from the University of Central Florida, and is a member of the IEEE Computer Society. Contact him at jchen@cs.gmu.edu; www.cs.gmu.edu/~jchen.

How to Reach *CiSE*