Science at the Speed of Thought

Judith E. Devaney, Steven G. Satterfield, John G. Hagedorn
National Institute of Standards and Technology
100 Bureau Dr., MS 8914
Gaithersburg MD 20899-8911
{judith.devaney,steven.satterfield,john.hagedorn}@nist.gov • http://math.nist.gov/mcsd/savg

ABSTRACT
In this paper we describe a flexible environment that combines scientific data mining with parallel computing in an immersive visualization environment. The goal is to minimize the time between the generation of a scientific hypothesis and the test of that idea, or science at the speed of thought.

Author Keywords
Discovery Science, Scientific Datamining, Immersive Visualization.

ACM Classification Keywords
H5 Information interfaces and presentation; I2 Artificial Intelligence; I3 Computer graphics

INTRODUCTION
Science advances through iterations of *theory*, *experiment*, and *analysis*. But the methods that scientists use to go about their work are changing[10]. Increasingly *experiment* means computational experiment as computers increase in speed and memory. Parallel computing assists in this by providing access to more processors and more memory. Consequently more complex models that run in feasible times become possible. Laboratory experiments as well are becoming parallel as combinatorial experiments become more common. Both types of experiment can produce large quantities of data where the time to analyze data can exceed the time to generate it. The generated data may also be non intuitive. For example nanoscience studies materials whose properties may change greatly as their size is reduced[1]. Thus analysis may benefit from new ways to examine and interact with data. But scientists also need access to their traditional analysis tools while taking advantage of new methods and technologies. Immersion is being actively studied as a tool for scientific visualization[11].

THE IMMERSIVE VISUALIZATION ENVIRONMENT
In our immersive visualization environment, we do not distinguish between information visualization[3] and scientific visualization, but rather seek to combine both to provide the maximum feedback to the scientist. We also provide an access path directly to the computing environment where the scientist can execute commands and have the result returned to the immersive environment. We use freely available data mining software as well as our own equation discovery software[4]. We have created a Glyph ToolBox to facilitate creating three dimensional glyphs that includes color and opacity parameters as well as location information. Additionally, we have created a suite of tools that allow us to display our data with differing views in space as well as time. Our environment has some aspects of a museum where there are different rooms that have multiple types of display such as wall hangings, rugs, sculptures, animations, and interaction. (In a *virtual* museum, you can even change the architecture (rooms) on the fly). It also has some aspects of a supermarket where variations on the same thing are presented in multiple ways and similar things are grouped together. We illustrate some some of these ideas with three data sets.

EXAMPLES
Figure 1 shows part of our visualization of the Iris data set [2]. (The full visualization contains multiple rooms with an alternate visualization of the same data set in each room, enabling a scientist to visit each of the rooms.) On the near side of the left wall is a parallel coordinate plot[6] of the cluster identified with the transparent envelope. On the far side of the left wall is a plot of the probability density distribution of each of the attributes in the data set. The right wall shows how the attributes rank with Information Gain[12]. In the foreground is a set of statistics that have been computed on the fly in response to a user command. The points of the data set are represented as glyphs where the attributes have been mapped to glyph attributes. The points are plotted in the central cube. A user can also interact with this visualization by turning the transparent envelops of the clusters on and off individually, and the

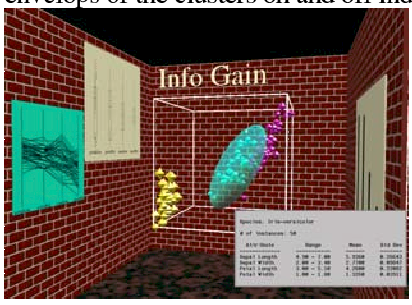

Figure 1. Part of an immersive visualization of the Iris data set. The full version contains multiple rooms.

parallel coordinate plots with them.

Figure 2 shows a combined surface volume visualization of Optical Coherence Tomography of a scaffolding used to grow cells on. The light gray is the scaffolding and the dark gray is the interface between the scaffolding and the pore space. The scaffolding can be viewed by itself or the segmentation (surface) can be viewed alone as well. The user clicks through the individual views to see each one. Separate views can also show a time sequence of the cells growing on the scaffolding; thus giving scientists insight into the growth process.

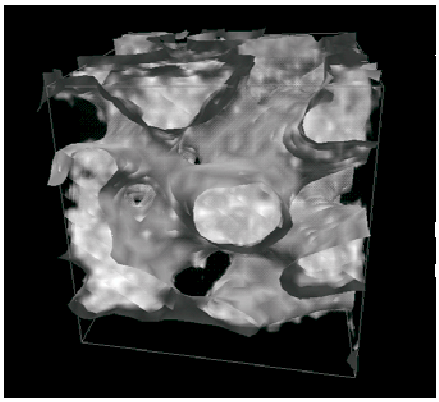

Figure 2. A combined surface volume visualization of Optical Coherence Tomography of a scaffolding used to grow cells on.

According to Louis Pasteur "In the field of observation, chance favors only the prepared mind" [8]. Scientists are prepared, but can we increase chance? We believe that with the tools of immersive visualization, data mining, interaction, and multiple views, it is possible to accelerate concept development. Figure 3 shows a visualization of a shake gel[9]. Because of this visualization, scientists were able to determine that it is the water's *oxygen* atoms, instead of the *hydrogen* atoms as previously thought, that attach to the clay. When a scientist can get access to their data at the scale of the data, where access means multiple types of visualizations in space and time as well as the ability to compute and recompute values on the fly, the possibility of discovery increases as in the case of the shake gel.

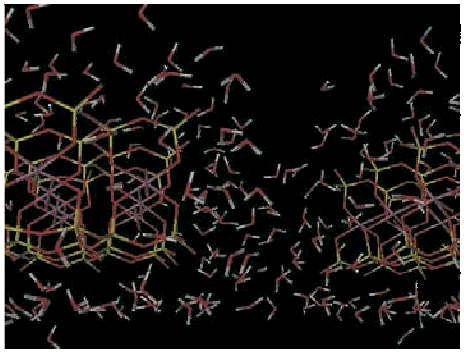

Figure 3. Visualization of a Shake Gel.

REFERENCES

1. Alivisatos, A.P., Semiconductor Clusters, Nanocrystals, and Quantum Dots, *Science*, 271, 933-937 (February 16, 1996)
2. C.L. Blake and C.J. Merz. UCI repository of machine learning-databases,1998. http://www.ics.uci.edu/~mlearn/MLRepository.html.
3. Card, S.K., Mackinley, J.D., Shneiderman, B., Readings in Information Visualization, Morgan Kaufmann, 1999.
4. Devaney, J.E., The Role of Choice in Discovery, Lecture Notes in Computer Science, vol. 167, eds. S. Arikawa and S. Morishita (Springer, 2000, pp.247-251).
5. DIVERSE: Device independent virtual environments-reconfigurable,scalable,extensible http://diverse.sourceforge.net/.
6. Inselberg, A., The plane with parallel coordinates. *The Visual Computer*, 1:69–91, 1985.
7. Kelso, J., Satterfield, S.G., Arsenault, L.E., Ketcham, P., Kriz, R., *DIVERSE:* A Framework for Building Extensible and Reconfigurable Device Independent Virtual Environments and Distributed Asynchronous Simulations , Presence: Teleoperators and Virtual Environments, 12 (1) , February 2003.
8. Louis Pasteur, Lecture, 1854, http://www.quotationspage.com/quotes/Louis_Pasteur/
9. Recipe for a Shake Gel, NIST Techbeat, 8/27/2003 www.nist.gov/public_affairs/techbeat/tb2003_0827.htm
10. Sims, J.S., George, W.L, Satterfield, S.G., Hung, H.K. Hagedorn,J.G., Ketcham, P.M., Griffin, T., Hagstrom, S.A., Franiatte, J.C., Bryant, G.W., Jaskolski, W., Martys, N.S., Bouldin, C.E., Simmons, V., Nicolas, O.P., Warren, J.A., am Ende, B.A., Koontz, J.E., Filla, B.J., Pourprix, V.G., Copley, S.R., Bohn, R.B., Peskin,, A.P., Parker, Y.M., Devaney, J.E., Accelerating Scientific Discovery Through Computation and Visualization II,, NIST Journal of Research, 107 (3) , May-June, 2002, pp. 223-245.
11. van Dam, A., Laidlaw, D.H., Simpson, R.M, Experiments in Immersive Virtual Reality for Scientific Visualization, Computers and Graphics, 26 (2002), 535-555, Elsevier Publications, 2002.
12. Witten, I.H., Frank, E. Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kauffann, San Mateo, 1999.