Social-Aware Collaborative Visualization for Large Scientific Projects

Kwan-Liu Ma and Chaoli Wang
Department of Computer Science
University of California at Davis
Davis, CA 95616
{ma, wangcha}@cs.ucdavis.edu

ABSTRACT

This paper discusses how to better support collaborative work for large scientific projects using visualization. Particular considerations are given to knowledge sharing and the social aspect of collaboration. The goal is to provide a roadmap for creating next-generation collaborative technologies with visual and social augmentation.

KEYWORDS: Collaborative work, information visualization, social data analysis, scientific visualization, user interface design, collective analysis

1. INTRODUCTION

Greater advances can only be made with collective efforts. Most large-scale scientific investigations are highly interdisciplinary and collaborative, with project investigators often geographically distributed. How can these investigators effectively share their data, knowledge, and research findings without time/place constraint? A good solution is web-based collaboratories.

Many major science projects use web-based collaboratories. Most of these collaboratories, however, are simply data repositories with web interfaces for looking up information. The next-generation collaborative space should support beyond browsing the stored information. There is valuable knowledge to be extracted and utilized. For example, the associations between data and users are as valuable as the data itself. These associations can be large and complex, and thus hard to comprehend. In addition, understanding the interaction and communication among users, which reflect the social aspect of the collaboration, can help improve the workspace. Confronting such rich information, visualization can play a key role in both knowledge discovery and collaborative space management. Views tailored to the user's background and interests are particularly desirable. Furthermore, additional cares are needed for handling results produced in a collaboratory which include a variety of forms such as text, table, graph, image, animation, and visualization. As such, new metrics or techniques are necessary for measuring and normalizing heterogeneous results so that they can be arranged appropriately in a common layout.

In this paper, we discuss areas of collaborative work where visualization can facilitate and greatly enhance performance with a focus on large-scale scientific investigations. We consider, in particular, knowledge sharing and the social aspect of collaboration.

2. SPACE FOR COLLABORATION

The term "collaboratory" was defined by Wulf [1] as a "center without walls", in which the nation's researchers can perform their research regardless of their physical locations, interacting with colleagues, making use of instrumentation, sharing data and computational resources, and accessing information in digital libraries. With the flourishing of the Internet in the 1990s, web-based collaboratories have been established and put into operation for many major science areas. For example, the Upper Atmospheric Research Collaboratory (UARC) [2] is a web-based, multidisciplinary research collaboration that gives space scientists around the world a virtual, shared workspace for conducting real-time experiments as well as various asynchronous collaborations. Kilman et al. [3] created TeleMed, a prototype international health care collaboratory that allows physicians located at different places to consult on a patient record. Both realtime interaction and offline communication using textual or audio annotations are supported in the consultations. More recent efforts in the U.S. Department of Energy's program included Particle Physics Data Grid Collaborative Pilot, Earth System Grid II, National Fusion Collaboratory, and Collaboratory for Multi-Scale Chemical Science.

Most collaboratories to date have been applied in major scientific research projects with various degrees of success. The collaboratory has proven itself as a valuable solution to

relax the constraints of place and time for scientists and researchers through building a virtual organization with communication and computing technologies. So far, the research and development of collaboratories has followed a tool and/or data centric approach. However, a collaboratory is beyond a data warehouse. It should support the very nature of collaboration in the scientific context. Collaboration is driven by the need to share both data and knowledge about the data. In particular, the substantial knowledge about the use and evolution of a collaboratory should be extracted and utilized to continuously enhance and enrich the collaborative space.

3. SOCIAL CONTEXT AND ACTIVITIES IN A COLLABORATIVE SPACE

Shared data is useful only if sufficient context about the data is given so that collaborators may understand and apply it appropriately. Therefore, it is important to know how a piece of data relates to the overall data space, user space, and application space. The interactions among collaborators are as valuable as the data itself. By focusing on the dynamics of information exchange, Henline [4] argues that the key challenges in creating a collaboratory may be social rather than technical. Cogburn [5] also pointed out that a collaboratory is a new networked, organizational form that also includes social processes. According to a report [6], from 1992 to 2000 the total use ranged from 17 to 215 users per collaboratory. The communication among collaborators can thus be large and complex, and becomes difficult to comprehend. Extending the collaboratory concept to include both social and behavioral research could provide opportunities of learning more about the social infrastructure that supports a distributed knowledge network. Understanding the interaction and communication among users, which reflect the social aspect of the collaboration, can potentially help improve the work space.

The next-generation science collaboratory should be augmented by cultivating a social infrastructure to facilitate knowledge discovery, consolidation, and utilization. New methods and interface designs need to be developed for visualizing social context and activities in the collaborative space, to derive collective knowledge from separate findings, and to eventually enhance existing cyber collaboratories with social augmentation.
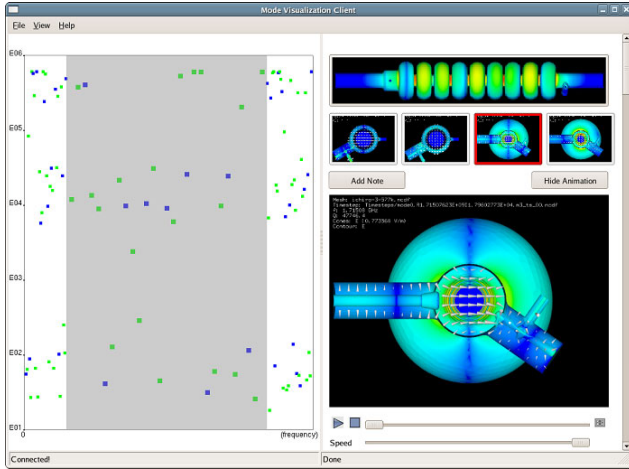
From a macroscopic level, most collaboratories follow a similar structure. They consist of multiple users working with a set of tools and input data, both of which might be heterogeneous, provided on potentially diverse computing platforms. Multiple tools are sometimes used in sequence or alongside other intermediate or external experimentation, which could be called a workflow. Both individual tools and entire workflows have certain configurations, which are advantageous for specific tasks. Tool usage and other experimentation generate output or findings, and the users discuss or annotate these findings or synthesize them into more significant results. Making sense of such a complex web of connected components can be quite daunting, especially when hoping to uncover patterns for which one is not explicitly searching. Statistical techniques are well suited for matching incoming data against known patterns, but are weak for pure data exploration. Because it leverages the pattern-recognition power of the human visual system, a multi-level, interactive visualization approach can help understand the social relationships of online collaboratories.

At the topmost level, the collaboratory consists of an intertwined mesh of heterogenous components (authors, tools, annotations, etc.) connected with several relations. Tools for social network analysis, especially those equipped to navigate and display heterogeneous graphs such as Onto-Vis [7], are usable for providing an overview of the entire collaboratory. By allowing the users to freely define and select a cross-section of both components and relations, these systems help pull back the shroud of clutter and complexity and reveal underlying patterns.

Such a visual overview would be useful for answering certain high-level questions, but would also provide a mechanism for a clean navigation to the different, lower-level visualizations. A unified visual overview also permits the annotation of heterogeneous components in a simple, consistent way. Clement [8] points out how facilitating discussion of tools by a relatively non-technical user base led to the emergence of informal learning and spontaneous local expertise. Data sets, previous publications, and intermediate findings could likewise benefit from such collaborative social understanding, and a consistent interface lessens the burden on the user. The annotations themselves are sources of social information currently undervalued in the collaboratory. Jung et al. [9] introduce "Annotizer" - an online annotation system that improves collaboration by allowing users to create, share, and search annotations on existing HTML contents.

The International Linear Collider (ILC) project [10] involves researchers from SLAC (Stanford Linear Accelerator Center), KEK (High Energy Accelerator Research Organization) in Japan, DESY (Deutsches Elektronen Synchrotron) in Germany, and various U.S. national laboratories. Scientists on this project may run the same simulation code with different parameter settings. They looked at not only the output data they generated but also those by others. When they examined the data, they could make visual-
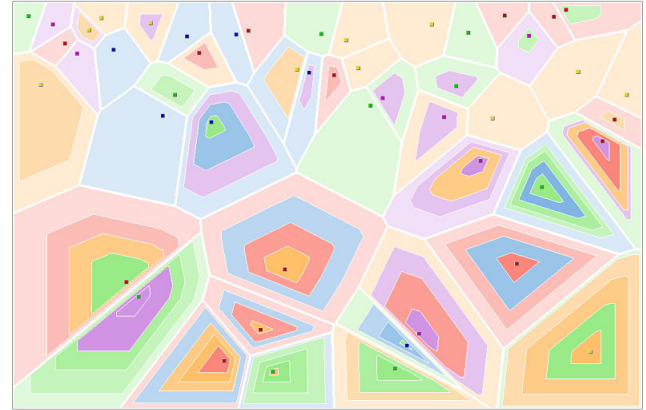
Figure 1. An interface for visualizing simulation results shared by a group of scientists [11].

Figure 2. A visual summary of available annotations using Voronoi partitioning of the space [12].

ization and add notes summarizing their findings. Such an analysis activity resulted in a potentially very large collection of output data, images, and notes, which were not organized for convenient browsing and comparative analysis. In cooperation with the scientists, we have created a web-based interface displaying the results of each simulation run in terms of visualization and animations, along with notes made by those who have examined the simulation results [11, 12].

Figure 1 shows one such interface. Each point in the left half of the window represents a simulation run. A selected point corresponds to multiples values along with the collection of images and animations displayed in the right half of the window. The user can switch to a separate visualization to see how many notes have been written for each simulation run, as shown in Figure 2. The points correspond to the data points displayed in Figure 1. Color encodes authorship, the number of cell layers corresponds to the number of annotations, and the thickness of each layer indicates the size of each annotation. The visualization can also show authorship and the evolution of the annotation creation over a given time period. For example, users can discover patterns of annotation authorship by selecting different subsets of authors. Perhaps two people frequently work on the same data points because they have compatible ideas, work habits, or simply because they inspire each other. The visualization highlights such patterns easily, while they are essentially invisible without a visual means.

What we need to do next is to study how such a facility benefits the project and if it changes how scientists work. The scientists can obtain a quick overview of what have been done by whom, what have not been extensively stud-

ied, which simulation output has received the greatest attention, etc. Improved understanding of the overall work and findings will help them better direct their further effort.

There is also the need to compare different simulation codes for the same modeling problem. In that context, visualization support for collaborative annotations appears to be equally useful.

4. COLLECTIVE ANALYSIS

In a large collaborative research team, quite often, the same research topic may lead to different approaches and techniques among the participating groups. As a result, different data are produced and different parameters are chosen to derive the outcomes, which may be similar, different, or even conflicting. Scientists need to compare, analyze, and evaluate all the information in order to answer important questions leading to collective findings that may actually become knowledge or suggesting further investigation due to inconsistent results.

However, most current collaboratories simply collect research highlights from the participating groups and publish them in the same report. This practice is convenient but the critical aspect of information presentation and knowledge consolidation is missing. The rich information that scientists are dealing with is multidimensional and heterogeneous. It includes experimental data, assumptions, procedures, parameters, and results, which themselves may be in a variety of forms such as text, table, graph, image, animation, and visualization. Designing methods for comprehending disparate information in the collaboratory is essential for scientists to go beyond loosely connected reports and derive collective knowledge from separate findings.

The first step towards this research goal is to represent the multidimensional parameter space for the subject of study and provide techniques for visual exploration. A wealth of research has been conducted to represent and interact with high-dimensional parameter spaces. For example, Feiner and Beshers [13] introduced an interaction metaphor called the "worlds within worlds": nested heterogeneous coordinate systems allowing users to view and manipulate high-dimensional functions. On the other hand, the parameter space in the collaboratory varies from group to group due to the different approaches taken. Scientists need to understand the similarities and differences among disparate parameter spaces from multiple groups. They can first choose a subset of parameters from their own parameter space for the analysis. Since the parameter space could be huge, an interface should be provided for an easier selection by categorizing and/or prioritizing parameters according to their properties and/or importance. More importantly, such an interface should enable scientists to perform cross-comparison of their parameters with parameters published by other groups. The interface can also be intelligent, which means that it can perform semantic parameter matching and comparison automatically whenever possible and prompt users to resolve when ambiguity arises. Furthermore, the interface should highlight the similarities and differences among multiple parameter sets with proper visual attributes for drawing users' attention.

What we have described above is a parameter-centric forward approach: scientists start from the parameter space and explore how different parameters lead to different results. Alternatively, we can also take a result-centric backward approach: scientists start from experimental data or results of interest and compare across different groups. Through this study they can put focus on data and results and then backtrack what parameters make the differences. Several visualization user interfaces can be used to present findings made by scientists about the same subject or data of interest. For instance, Design Gallery [14] and Image Graphs [15] are two such interfaces. Design Gallery can automatically generate image rendering results by varying the input parameter vector and arrange the results in a 2D layout, from which users simply select desired ones. Using this interface, scientists can arrange their data and results, and quickly identify interesting point clouds which may lead to collective findings and scattered points which may require further investigation. Image Graphs can be integrated into the layout of data or results to show detailed information about how they are generated. This visual tool allows scientists to efficiently backtrack and unveil the relationship between the parameters or settings and the data or results produced at different stages. Instead of imagining the whole process in their minds, scientists can leverage
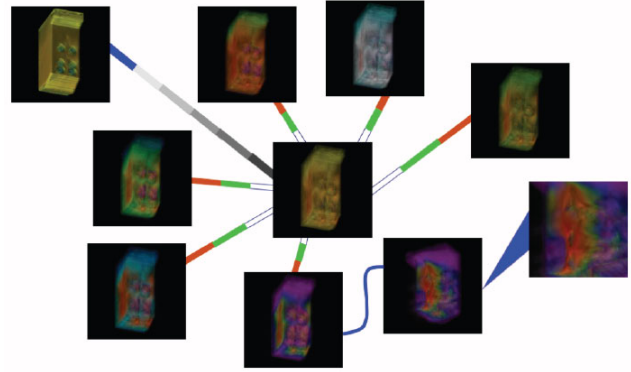
Figure 3. Image graph of a visualization process [15].

this interface to assist them in figuring out potential problems and actions to take. Figure 3 presents an image graph. Each image shows a visualization of a furnace data set, and each edge represents the operation connecting the images.

Arguably, the primary goal of an online collaboratory is to focus the collective efforts of the group in order to produce some significant, useful results. Yet the path to understanding might be just as valuable as the end results, especially if it can assist the discovery process for subsequent tasks or other endeavors entirely. Capturing and visually analyzing the discovery process has been studied for the task of visualization [16]. The idea of making visualization by analogy [17, 18, 19] can be utilized to match processes created by different groups and to possibly configure existing or assemble new processes effectively.

Similar techniques could be created to capture the interwoven tapestry of discussion, data, tool usage, and cross-referencing that leads to general scientific discovery based on the social data collected in a collaboratory. Such techniques can help understand how researchers form collective conclusions, delineate the differing viewpoints, and identify the significant contributors. Finally, social awareness, which provides a context for your own activity [20], is critical to successful collaborations. Awareness can be achieved through a notification system. A flexible notification system should adopt strategies that propagates one user's actions to others synchronously (such as instant messenger) or asynchronously (such as email or bulletin board), depending on message priority as well as user preference.

5. CONCLUSION

Scientific collaboration is intrinsically social interaction among collaborators. The next-generation web-based collaboratories should be enhanced with social augmentation towards a more efficient and effective collaboration among

collaborators. Leveraging visual means and visualization techniques, we can help collaboratory users discover complex relationships and interactions hidden in the collaborative space, facilitate communication and interaction for a better utilization of aggregated software, hardware, and human resources, and ultimately propel knowledge discovery.

Our work is motivated by our participation in the U.S. Department of Energy's SciDAC (Scientific Discovery through Advanced Computing) program [21, 22], which sponsors a dozen high-profile science projects. It is imperative for the partners in each project, who are distributed in the country, to share their creations and findings. There is thus a pressing need to provide support for collaborative work. In this paper, we attempt to address the need to support social-aware collaborative data analysis with visual means, and to direct the visualization research community towards the development of needed technologies. The former effort started with the development of VICA [12] - a web-based collaborative data analysis facility, and the latter has led to the creation of an online collaboratory called VisFiles [23] for sharing data sets and visualization experience. VisFiles essentially invites others to join us in working on some of the most challenging scientific visualization applications. In the future, we will continue our efforts to build the social-aware VisFiles collaborative visualization space, and incorporate user evaluation into our research and development process for an enduring success.

ACKNOWLEDGMENTS

REFERENCES

[1] Wulf, W. A., "The Collaboratory Opportunity," *Science*, Vol. 261, No. 5123, 1993, pp. 854-855.

[2] [2] Subramanian, S., Malan, G. R., Shim, H. S., Lee, J. H., Knoop, P., Weymouth, T. E., Jahanian, F., and Prakash, A., "Software Architecture for the UARC Web-Based Collaboratory," *IEEE Internet Computing*, Vol. 3, No. 2, 1999, pp. 46-54.

[3] Kilman, D. G., and Forslund, D. W., "An International Collaboratory based on Virtual Patient Records," *Communications of the ACM*, Vol. 40, No. 8, 1997, pp. 110-117.

[4] Henline, P., "Eight Collaboratory Summaries," *Interactions*, Vol. 5, No. 3, 1998, pp. 66-72.

[5] Cogburn, D. L., "HCI in the So-Called Developing World: What's in it for Everyone," *Interactions*, Vol. 10, No. 2, 2003, pp. 80-87.

[6] Sonnenwad, D. H., "Expectations for a Scientific Collaboratory: A Case Study," In Proceedings of ACM SIGGROUP Conference on Supporting Group Work, 2003, pp. 68-74.

[7] Shen, Z., Ma, K.-L., and Eliassi-Rad, T., "Visual Analysis of Large Heterogeneous Social Networks by Semantic and Structural Abstraction," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 12, No. 6, 2006, 1427-1439.

[8] Clement, A., "Cooperative Support for Computer Work: A Social Perspective on the Empowering of End Users," In Proceedings of ACM Conference on Computer Supported Cooperative Work, 1990, pp. 223-236.

[9] Jung, B., Yoon, I., Lim, H., Ramirez-Weber, F. A., and Petkovic, D., "Annotizer: User-Friendly WWW Annotation System for Collaboration in Research and Education Environments," In Proceedings of IASTED Web Technologies, Applications, and Services, 2006.

[10] Ko, K., "The International Linear Collider," *SciDAC Review*, No. 1, 2006, pp. 17-20.

[11] Ma, K.-L., Schussman, G., Wilson, B., Ko, K., Qiang, J., and Ryne, R., "Advanced Visualization Technology for Terascale Particle Accelerator Simulations," In Proceedings of ACM/IEEE Supercomputing Conference, 2002.

[12] Wang, Y., Shearer, J., and Ma, K.-L., "VICA: A Voronoi Interface for Visualizing Collaborative Annotations," In Proceedings of International Conference on Cooperative Design, Visualization, and Engineering, 2007, pp. 21-32.

[13] Feiner, S., and Beshers, C., "Worlds within Worlds: Metaphors for Exploring N-Dimensional Virtual Worlds," In Proceedings of ACM Symposium on User Interface Software and Technology, 1990, pp. 76-93.

[14] Marks, J., Andalman, B., Beardsley, P. A., Freeman, W., Gibson, S., Hodgins, J., Kang, T., Mirtich, B., Pfister, H., Ruml, W., Ryall, K., Seims, J., and Shieber, S., "Design Galleries: A General Approach to Setting Parameters for Computer Graphics and Animation," In Proceedings of ACM SIGGRAPH Conference, 1997, pp. 389-400.

[15] Ma, K.-L., "Image Graphs - A Novel Approach to Visual Data Exploration," In Proceedings of IEEE Visualization Conference, 1999, pp. 81-88.

[16] Ma, K.-L., "Visualizing Visualizations: User Interfaces for Managing and Exploring Scientific Visualization Data," *IEEE Computer Graphics and Applications*, Vol. 20, No. 5, 2000, 16-19.

[17] Tzeng, F.-Y., and Ma, K.-L., "Intelligent Feature Extraction and Tracking for Visualizing Large-Scale 4D Flow Simulations," In Proceedings of ACM/IEEE Supercomputing Conference, 2005.

[18] Scheidegger, C. E., Vo, H. T., Koop, D., Freire, J., and Silva, C. T., "Querying and Creating Visualizations by Analogy," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 13, No. 6, 2007, pp. 1560-1567.

[19] Lu, A., and Ebert, D. S., "Example-Based Volume Illustrations," In Proceedings of IEEE Visualization Conference, 2005, pp. 83-90.

[20] Dourish, P., and Bellotti, V., "Awareness and Coordination in Shared Workspaces," In Proceedings of ACM Conference on Computer Supported Cooperative Work, 1992, pp. 107-114.

[21] SciDAC, `http://www.scidac.org`.

[22] Ultravis Institute, `http://www.ultravis.org`.

[23] VisFiles, `http://vis.cs.ucdavis.edu/VisFiles`.