An Enriched Visualization Framework for Multidisciplinary Data Analysis

Kristina D. Mickus-Miceli* **and Gitta O. Domik**
Department of Computer Science
University of Colorado, Boulder
Boulder, CO 80309-0430
domik@cs.colorado.edu

Abstract

One of grand challenges in the Computational Aerosciences Project of the High Performance Computing and Communications Program involves performing multidisciplinary simulations in which two or more phenomena, such as aerodynamics and structures, are strongly coupled. This challenge has, in turn, given visualization researchers the challenge of developing methods to represent the interaction between disciplines effectively. Visual representations are useful only if they augment comprehension of the increasing amounts of data being generated by these simulations. Visualization researchers must ensure that these representations are useful and do not contain information that might lead to misinterpretation. This paper presents the development of a visualization tool to aid scientists in the analysis of multidisciplinary simulations. The authors define a visualization framework that supports the production of effective visualizations by adapting the visual representations to the goal of the scientist, the type of data to be displayed, software and hardware considerations, and rules of visual perception.

Introduction

As computer simulations become more complex, systems to aid scientists in analyzing the vast amount of information in a reduced amount of time are very desirable. Computer assistance with the development of visual representations of scientific data would no longer require the scientist to learn the many aspects of visualization. By providing intelligent assistance, the amount of time learning software and generating visualizations would be reduced, allowing the scientist to concentrate his/her time on the task at hand. Also, because the scientist is often unaware of the constraints imposed by computer software, hardware and the human visual system, storing this information in a knowledge based system would help in producing effective visual representations.

Data representation and knowledge based techniques are two of the AI methods that could benefit scientific visualization. Data representation involves organizing the data into appropriate forms so that the scientist can extract information from it easily. Knowledge bases would hold facts and rules from expertise in graphical presentation, visual perception and expertise in the scientific domains. This information would be coupled with the data representation in presenting the scientist with effective visual representations.

This could be applied to the development of a visualization system to assist with multidisciplinary simulations. Data representation would help to model the complexity of the data sets involved in multidisciplinary data. Knowledge based systems would store the facts and rules about the different scientific domains. These two sources of information would assist scientists in producing effective visualizations based on the goal of the scientist, the type of data to be displayed, software and hardware considerations, and rules of visual perception.

This paper details how a multidisciplinary visualization system might evolve around these AI-based concepts. A description of the computational environment as well as a short description of a proposed solution are presented in the next section. The following section reviews similar types of research systems that have evolved recently. A more thorough description of the visualization framework is then presented. Finally, the paper is concluded and future plans are discussed.

Multidisciplinary Visualization for Computational Aerosciences

Computational fluid dynamics (CFD), the study of fluid flow via numerical simulation, assists scientists and engineers in developing a better understanding of fluid flow and how it affects the flight characteristics of aircraft and aerospace vehicles. These simulations can involve large data sets composed of multiple grids and multiple physical variables. For exam-

*Currently under fellowship at NASA Ames Research Center, MS T045-1, Moffett Field, CA 94035, kmiceli@nas.nasa.gov

ple, the simulation of an F-18 aircraft involves a grid composed of 1.6 million points each with five physical variables per instant in time. To perform a thorough analysis of this data simply by studying the numbers is difficult, if not impossible. As a result, the technology to graphically display this data is essential. This technology is termed "scientific visualization" and many techniques have been developed to extract and present information from CFD data[Hultquist, 1992; Gerald-Yamasaki, 1992; Bryson and Levit, 1991; Globus *et al.*, 1991; Uselton, 1991]. Some of these techniques have been incorporated into visualization toolkits specifically designed for the study of CFD results such as Plot3D[Walatka and Buning, 1989], FAST (Flow Analysis Software Toolkit)[Bancroft *et al.*, 1990] and Superglue[Hultquist and Raible, 1992].

With increases in computing power, scientists have become interested in taking the technology one step further. Since factors other than fluid flow affect the flight characteristics of a vehicle, scientists are interested in the study of integrated multidisciplinary simulations. These simulations can involve interactions between fluid dynamics, chemistry, combustion, structures and controls. This additional complexity gives visualization researchers the challenge of developing methods to display the interaction between disciplines effectively.

Large, single discipline CFD data sets are complex and difficult to manage, which leads to difficulties in presenting data to the scientist. The combination of data sources from multiple disciplines makes it clear to visualization researchers that issues of data management and representation are necessary. Methods and models to organize data and represent relationships between data variables and data sets are equally important to visualization research as are visualization techniques.

This project involves the creation of a visualization framework that attempts to describe the visualization environment in detail. This framework consists of a data model, a user model and a machine model. The integrating component of the framework is a knowledge base. The system uses information stored in the knowledge base, together with the information in the data, user and machine models to assist the scientist in generating visualizations by suggesting appropriate or "useful" visualization techniques.

The data model represents data from separate disciplines. This information, together with the relationships between data, is embodied in an object oriented model and serves as the base from which effective visualizations are built. This information expresses relationships between the data sets in order to define a natural organizational structure over the data. To complement the data model, the user model serves to describe the scientist and his/her characteristics. These characteristics include the interpretation aims of the user (what type of information the user is trying to extract), the background of the user, and any limitations the user may possess. The machine model incorporates information about the resources available in the scientist's computing environment.

The data model, user model and machine model form the base structure of the visualization framework. Visual representations are built by filtering information in the models through the knowledge base. The results of this query are the suggestion of a series of appropriate visual representations to the user.

AI Techniques in Visual Data Analysis

Visualization is essentially a mapping process, taking data attributes and mapping them onto visual representations. The mapping process is exploratory, creating representations that may or may not be useful to the scientist. Determining what constitutes a "useful" visualization is a difficult task. This research issue has been approached by researchers by studying the rules of perception and graphical presentation. For example, [Mackinlay, 1986] defined the words "effective" and "expressive" in attempting to determine a useful visual representation. "Expressive" describes a visualization that fully encodes the data attributes by presenting all relevant information to the user. The complementary word "effective" suggests that the technique exploits the capabilities of the output medium and human visual system, ensuring a correct and quick judgement of the data.

The criteria involved in the visualization mapping include the task or interpretation aims, data characteristics, user abilities and constraints, software and hardware environment, and human perceptual capabilities. The user is mainly concerned with her/his interpretations aims, being neither an expert on the system resources or human perception. Users lacking knowledge of the computing environment or human perception would benefit by having the system provide them with this information.

Visualization technology currently lacks the methods and models that would help to define these interfaces and create "useful" visualizations. Based on past experiences and the future promise of even more massive data sets, it is evident that the data must be central to the development of such a system. Not only does data drive the mapping process, but it must also be the underlying base for the user interface. This interface must give the scientist the capability to access the available data in a variety of ways as well as assistance in creating a visual representation in an appropriate form.

Enrichment of visualization systems using AI techniques is a new field of research. It consists of modeling the visualization process by carefully describing its components and making use of a knowledge base to map information into effective displays. Visualization models that are concerned with effective displays have emerged recently. A body of research on visualization

models and automatic generation of visual representations has recently developed. [Feiner *et al.*, 1992] call these new visualization systems "graphically articulate", forming the basis for the next generation of intelligent user interfaces.

Haber and McNabb [Haber and McNabb, 1990] developed an early model to aid complex visualizations of large scale numeric simulations. The model contains a description of the mainly quantitative data to be visualized. Decisions for the compositions of multiple data items are mapped into "abstract visualization objects" and these are rendered into displayable images. The resulting images are called "visualization idioms", referring to the abstract meaning of pictures as a visual representation of scientific data sets.

Mackinlay [Mackinlay, 1986] developed an early and complete model to automatically generate graphical presentations of relational information. The model is described by data, task and user directives. Data is described by a relational data model. Task related goals (e.g. relations to represent and/or relations to suppress) and user directives (such as prioritizing items or relations) define a first mapping into a set of graphical primitives which is expressed by a formal language consisting of syntax and semantics. To map from an internal representation to displayable images, evaluation criteria such as importance (ranking of tasks), expressiveness (encoding of data attributes) and effectiveness (psychophysical principles) are used. Although restricted to relational data and 2-d charts, Mackinlay's work is a foundation for complex visualization systems aiming towards automating the design of graphical presentations. The system that is based on these principles, APT (A Presentation Tool), contains a description of the information and task, defines evaluation criteria in formal terms, and is able to compose multiple items and relations into one effective display.

Robertson [Robertson, 1990] developed visualization guidelines by observing natural scenes and matching data and task characteristics to 2-d and 3-d scene parameters. His approach is different from previous techniques, in that he has a complete and coherent scene in mind already before the mapping stage begins. While restricting the user to predefined scenes, the natural scene paradigm guarantees perceptually valid mental models. De Ferrari [Robertson *et al.*, 1991] expanded the original paradigm to include a more elaborate data model, user directives and user interpretation aims (the latter two combined under the term "visualization specification") as input to the visualization system.

Wehrend and Lewis [Wehrend and Lewis, 1990] describe each visualization process by two dimensions: a) the characteristics of the information to be displayed and b) the specific perceptual task to be performed on the resulting images. The finite number of data characteristics and the finite number of perceptual tasks define a two dimensional matrix in which each element contains expressive and effective examples of visual representations. If more than one tuple (data, task) is to be represented in the same display, the user is responsible for setting priorities for the composition of representations.

The VISualization Tool Assistant [Senay and Ignatius, 1990] (VISTA) accomplishes complex visualization tasks automatically. The world to be represented is described by data characteristics. Knowledge about mapping procedures are stored in form of rules that observe expressiveness of data characteristics and effectiveness of displayable images. The data is first decomposed into partitions that can be mapped into individual visualization primitives, then combined into meaningful complex visual representations. A "visualization tree" describes mapping techniques and mapping structures. This tree can be edited by the user before final rendering is performed. Because of this modification stage, the system earns its name as "assistant", guiding rather than enforcing visual representations.

The Visualization Framework for Multidisciplinary Data Analysis

The visualization framework consists of a data model, a user model and a machine model. The integrating component of the visualization framework is that of a knowledge base. The information in the knowledge base is combined with the information in the data, user and machine models to suggest appropriate or "useful" visualization techniques to the scientist. The framework is implemented in an object oriented programming environment called Superglue which uses a dialect of Scheme [Hultquist and Raible, 1992]. This software is currently under development at NASA Ames Research Center.

The Data Model

Scientific data requires rich and flexible data models to support the often complicated relationships that occur. The ability to extend the data model is another essential component of the system. Object-oriented concepts form a good basis for a rich data model, allowing applications such as scientific data management to take advantage of its promising modeling capability. The object oriented model was chosen due to its characteristics and ability to handle diverse and complex data sets. The data model represents data from separate disciplines. This information, together with the relationships between data, serves as the base from which the visual representations are built.

The current data model contains information about the data: its structure, components, relationships and constraints. The basic structure of the data model is built around the application specific components that exist in a multidisciplinary fluids/structures simulation. The data model is decomposed hierarchically into two objects, a fluids object and a structures object. The fluids object contains a list of objects that

represent real entities such as a wing or a fuselage of an aircraft. These objects are further decomposed into timestep objects. Due to the iterative nature of the simulation, there are different data sets for different moments in time. Each timestep object consists of a grid object and a data object. The grid object has the grid information about the physical structure of the fluids object. The data object contains the data values that are calculated at each of the grid points in the grid object. In addition, information such as the name of the data variable and its attributes is incorporated into the data model. Data attributes include information such as whether the data is a scalar or vector field, how the data variable is mathematically related to other data variables and information specifying what visualization techniques (i.e. colormaps, contour lines, etc.) are most often used with the given variable.

A similar hierarchy exists for the structures data with modifications due the unstructured nature of the grids that specify structures objects. The structures objects correspond with the fluid objects, describing the same real entities (i.e. wing, fuselage). Each structures object contains a list of timestep objects, which each have a grid object and a data object. The grid object describes an unstructured grid as opposed to the structured grids used by fluid dynamicists. As a result, the data description is slightly different, involving a set of vertices and connectivity information. The vertices represent the physical locations of the grid points and the connectivity information describes how the vertices are connected to form the mesh around the physical object.

The data model is represented visually to the user. This will allow the user to investigate the data model and its components. The user will also be able to query the data model visually in the process of performing the data analysis. The specifics of the interface have not yet been determined.

The User Model

To complement the data model in this visualization framework, the creation of a user model serves to describe the user and his/her characteristics. The user model enhances the system by allowing it to tailor presentations to individual user preferences and requirements. The user model contains facts and rules about the user that may be relevant to the behavior of the system. This may include the interpretation aims of the user, the background of the user, and any limitations the user may possess.

Interpretation aims express the goal of the user and what information they are trying to extract from the data. Interpretation aims might include such tasks as 1) locating areas of interest in the data such as low pressure regions, 2) identifying characteristics of the data such as points on a wing where the velocity field is zero, or 3) correlating data variables to see how the pressure from a fluid affects the structural integrity of a wing. Background information about the user might include information such as the meaning of certain color tables, the meaning of shades, preference for certain visual representations and interaction techniques.

The user model is represented using a stereotype hierarchy, as defined in [Kass and Finin, 1991]. This method uses stereotypes to describe a class of users by defining facts and rules about the users. The hierarchy is composed of two main stereotypes, namely the fluid dynamicists and the structural dynamicists. Each class of users has a specific academic background which has, in turn, led to the use of certain types of physical models, software, computing environments, etc. These facts and rules can be of two types, either "definite" or "default". Definite facts and rules always apply to the given class of users. For example, fluid dynamicists always use certain mathematical relationships to describe the relationships between data variables. Structural dynamicists have other mathematical relationships that describe the phenomena that they are interested in. Default facts and rules can be overridden as the stereotype hierarchy defines smaller classes of users. For example, a default rule might be that all fluid dynamicists use structured grids for the computations. However, this can be overridden if a certain group of fluid dynamicists uses the unstructured grid technique to solve their problem.

The stereotype hierarchy will be further decomposed into smaller user classes and eventually to single users, which are represented as leaf nodes of this hierarchy. The acquisition of the information for the user model can take place in two modes, either explicitly or implicitly. Explicit acquisition occurs before the user begins work, by filling in a template or by answering a series of questions. The implicit acquisition of information happens during a working session, by making conclusions based on the work habits of the user. The user model will interact with the data modelOB and machine model through the knowledge base. The information from these models is useful in developing visual representations that satisfy the needs and the desires of the user.

The Machine Model

The machine model is perhaps the most straightforward of the components comprising the visualization framework. This model simply details the capabilities of the computing environment available for scientific use, both in terms of software and hardware resources. The scientist should not be expected to know what type of environment he/she has available to them. This information is important, however, impacting what type of visualization will be appropriate for a given hardware configuration. This information can be encoded easily into the machine model, specifying the type of system, monitor resolution, graphics

capabilities, color facilities, CPU, memory, etc.

For this project, the machine model will be restricted to only the current machine used by the developer of this project, namely a Silicon Graphics workstation. The development of the machine model is not a main emphasis of the project, but is still a necessary part. This model can be expanded to other machines as the project progresses.

The Knowledge Base

The knowledge base will be intertwined with the three models, connecting them with rules obtained from graphical presentation, visual perception and experience in the given scientific domains. The information stored in the knowledge base will assist the scientist in developing useful visualizations. The visualizations will be judged based on the expressiveness and effectiveness criteria initially developed by Mackinlay [Mackinlay, 1986]. Expressiveness criteria determine how well a graphical representation expresses the information stored in the data and effectiveness criteria determine if the graphical representation takes advantage of the output medium and the human visual system.

Data characterization is based on the capability to describe effective and expressive qualities of the data. There are many dimensions involved in data characterization such as the properties of sets of data, including how they are ordered (quantitative, ordinal, nominal), measures of quantity or coordinates and their domain of membership (whether they involve domains of time, space, temperature, mass, etc). The system will display the data based on the data characterization and visualization techniques specified in the knowledge base.

Future Work

At the time of this paper, this project was in the beginning stages. A data model had been developed to represent the data involved in a multidisciplinary fluids/structures interaction. Visual tools were available to allow the user to view what information was in the data set. The object oriented nature of the model gave promising flexibility to adding, deleting and updating information and adjusting the model as was deemed necessary during the evolution of the data model. Building relationships between the data variables was a result of implementing the laws of fluid and structural dynamics and by incorporating information obtained from the scientists. Future plans for the data model include further enhancements to its structural and behavioral properties, more visual tools to help the user peruse the information visually, methods for querying the data visually and overall enhancements to the general representational model.

The user model was in beginning stages of development, mostly conceptual models on paper. The model involved considerations such as how to obtain information from the user and how to access the model and integrate information with the data model. The user model requires a great deal of thought as to how to create the model and the level of complexity desired.

The knowledge base, the integrating component of the visualization model, will be incorporated into the framework by obtaining information about different forms of graphical presentation and perception. This process will continue over the duration of the project.

The data analysis tool will provide the user with an interface to select the data and representation in question, display the different views of the data, and allow the user to manipulate the resulting image. The results from the separate visualizations will be displayed separately or merged together if the representations do not detract from each other. If the results are displayed separately, the visualizations will be coordinated so that their orientations will match.

Conclusions

In order to analyze the vast amounts of data being generated today by complex computer simulations, scientists require the help of some type of assistance or intelligence. Without some type of assistance from an intelligent interface, the scientist will be required to learn the complexities of generating visual representations and understanding the toolkits that build them. The use of techniques derived from artificial intelligence can enhance the data analysis process by presenting the user with appropriate visual representations thereby allowing the scientist to concentrate on research instead of learning new software tools and techniques.

Acknowledgments

KM would like to thank Sam Uselton, Jeff Hultquist, and Eddy Pramono for their guidance throughout the development of this project.

Kristina Mickus-Miceli is employed by Computer Sciences Corporation and this work is performed for NASA under government contract NAS 2-12961. Gitta Domik is supported under NASA grant NAGW-1902.

References

Bancroft, Gordon V.; Merritt, Fergus J.; Plessel, Todd C.; Kelaita, Paul G.; McCabe, R. Kevin; and Globus, Al 1990. FAST: A multi-processed environment for visualization of computational fluid dynamics. In Kaufman, Arie, editor 1990, *Proceedings of Visualization '90*. IEEE Computer Society. 14–24.

Bryson, Steve and Levit, Creon 1991. The virtual windtunnel: An environment for the exploration of three-dimensional unsteady flows. In Nielson, Gregory M. and Rosenblum, Larry, editors 1991, *Proceedings of Visualization '91*. IEEE Computer Society. 17–24.

Feiner, Steven; Mackinlay, Jock; and Marks, Joe 1992. Automating the design of effective graphics for intelligent user interfaces. In *Tutorial at the 1992 Conference on Computer Human Interaction.*

Gerald-Yamasaki, Michael 1992. Cooperative visualization of computational fluid dynamics. Technical Report RNR-92-007, NASA Ames Research Center, MS T045-1, Moffett Field, CA 94035.

Globus, Al; Levit, Creon; and Lasinski, Tom 1991. A tool for visualizing the topology of three-dimensional vector fields. In Nielson, Gregory M. and Rosenblum, Larry, editors 1991, *Proceedings of Visualization '91.* IEEE Computer Society. 33–40.

Haber, Robert B. and McNabb, David A. 1990. Visualization idioms: A conceptual model for scientific visualization systems. In Nielson, G.M. and Shriver, B., editors 1990, *Visualization in Scientific Computing.* IEEE Computer Society Press.

Hultquist, J.P.M. and Raible, E.L. 1992. Superglue: A programming environment for scientific visualization. In *Proceedings of Visualization '92.*

Hultquist, J.P.M. 1992. Constructing stream surfaces in steady 3d vector fields. In *Proceedings of Visualization '92.*

Kass, Robert and Finin, Tim 1991. General user modeling: A facility to support intelligent interaction. In Sullivan, Joseph W. and Tyler, Sherman W., editors 1991, *Intelligent User Interfaces.* ACM Press.

Mackinlay, Jock 1986. Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics* 5(2):110–141.

Robertson, Philip K.; Ferrari, Lisa De; Fletcher, Peter; and Vezina, Guy 1991. Visualisation on a massively parallel supercomputer. Technical Report TR-HJ-91-04, CSIRO Division of Information Technology, Center for Spatial Information Systems, GPO Box 664, Canberra, ACT 2601, Australia.

Robertson, Philip K. 1990. A methodology for scientific data visualization: Choosing representations based on a natural scene paradigm. In Kaufman, Arie, editor 1990, *Proceedings of Visualization '90.* IEEE Computer Society. 114–123.

Senay, Hikmet and Ignatius, Eve 1990. Vista: Visualization tool assistant for viewing scientific data. In *SIGGRAPH 1990 Course Notes: State of the Art in Data Visualization.*

Uselton, Samuel P. 1991. Volume rendering for computational fluid dynamics: Initial results. Technical Report RNR-91-026, NASA Ames Research Center, MS T045-1, Moffett Field, CA 94035.

Walatka, P.P. and Buning, P.G. 1989. *Plot3D User's Manual Version 3.6.* NASA Ames Research Center.

Wehrend, Stephen and Lewis, Clayton 1990. A problem-oriented classification of visualization techniques. In Kaufman, Arie, editor 1990, *Proceedings of Visualization '90.* IEEE Computer Society. 139–143.