Geovisualization for knowledge construction and decision-support

Alan M. MacEachren
Mark Gahegan
William Pike
Isaac Brewer

GeoVISTA Center and Department of Geography
The Pennsylvania State University

Guoray Cai

GeoVISTA Center and School of Information Science & Technology
The Pennsylvania State University

Frank Hardisty

Department of Geography
University of South Carolina

Vast digital data resources are being produced that include geospatial referencing, from precise geographic coordinates (e.g., generated by GPS receivers in vehicles, PDAs, cell phones, and other devices), through street addresses, to codes for administrative and other regions (e.g., zip codes, census tracts FIPS codes, drainage basin indices). A few specific examples of geospatially-referenced data include: satellite remote sensing, meteorological measurements, telephone and credit card transactions (with both purchase and billing address), stream gauge readings, land use categories, transportation data (linked to intersections, highway segments, ticket offices), health statistics (collected with home and treatment addresses), tax and property records, and census enumerations (for population, agriculture, housing, manufacturing and other topics).

Geovisualization is both a *process* for leveraging these data resources to meet scientific and societal needs and, together with the broader discipline of Geographic Information Science (GIScience), a *field of research and practice* that develops visual methods and tools to support a wide array of geospatial data applications. While there have been substantial advances in geovisualization over the past decade, many challenges remain. To support real-world knowledge construction and decision making, some of the most important challenges involve *distributed geovisualization*; enabling geovisualization across software components, devices, people, and places.

Integrating and extending perspectives

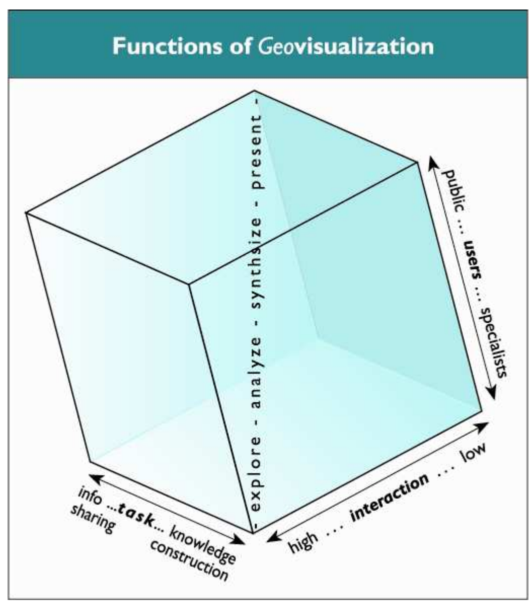
In her May/June 2003 *Viewpoints* paper, Theresa-Marie Rhyne highlighted some of the commonalities between cartographic and geographic information representation techniques and those in both scientific and information visualization [3]. Geovisualization draws upon these cartographic and geographic traditions, integrating their perspectives on representation and analysis of geospatial information with more recent developments in scientific and information visualization, exploratory data analysis (EDA), and image analysis. A general goal articulated for geovisualization is to integrate approaches from these domains "to provide theory, methods, and tools for visual exploration, analysis, synthesis, and presentation of geospatial data (any data having geospatial referencing)." [4] (see Figure 1).

The term *geographic visualization* (as well as the related, *cartographic visualization*) has an origin in the 1987 NSF report on visualization in scientific computing [5]. Research and practice in geovisualization, however, has roots at least a decade earlier, in Bertin's [6, French edition - 1977] cartographic and information design ideas for representing and exploring data [7]. Early work in geovisualization focused on the role of map-based, dynamic, visual displays as prompts for scientific insight [8] and on the methods through which dynamic visual displays might leverage perceptual-cognitive processes to facilitate scientific thinking [9].

In 1995, the International Cartographic Association established a *Commission on Visualization* that in 1999 expanded its focus to *Visualization & Virtual Environments*. This Commission has played an important role in stimulating geovisualization research and in articulating an international, interdisciplinary research agenda [4, 10]. That role has involved collaboration with the ACM SIGGRAPH Carto Project (http://www.siggraph.org/~rhyne/carto/).

Initial research prompted by the ICA Commission focused on development and implementation of highly interactive, exploratory methods targeted at knowledge construction by specialists [11], thus on support for visualization functions at the lower front corner of Figure 1 (this work balanced traditional cartographic research, focused on the top back corner, thus on presentation of existing information to the public).

The 2001 interdisciplinary geovisualization research agenda (cited above) articulated a broader set of challenges that includes attention to visually-enabled information retrieval and decision-making tasks for a



1 Four functions for geovisualization are depicted along the central diagonal of this geovisualization use space. The space is defined by: kinds of task, kinds of user, and level of interaction enabled in the interface. *modified from figure PIII.1* [1].

wide range of users, and for groups as well as individuals. One component of a recent U.S. National Research Council (NRC) report builds on this agenda to identify challenges for IT research related to human interaction with geospatial information [12]. Particular geovisualization issues targeted in this NRC report include: advances in visualization to harness information volume and complexity (including attention to visual representation of knowledge); universal access and usability (including extensions of visualization to other modalities), mobile information acquisition, access, and use (including design of visualization methods suited to small, wireless devices); and collaborative work with geospatial information (including attention to the role of visual display as a mediator for both same- and different-place group discussion).

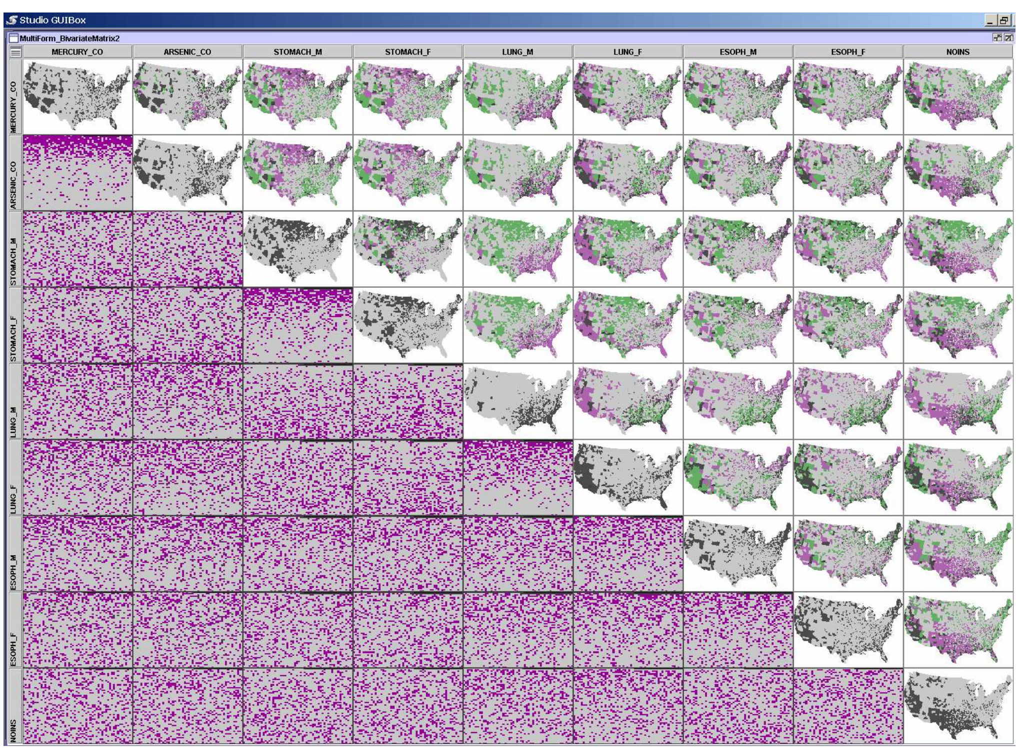
Application domains

The wide range of available geospatial data creates a potential for geovisualization to support activities in an equally wide range of application domains. Below, applications in three domains are highlighted, using examples from research underway in the GeoVISTA Center at Penn State.

Public Health

Geospatial data about health outcomes, interventions, and risk factors offer an opportunity to understand (and do something about) the varied geographic distribution of disease. These datasets, however, are highly multivariate and the complex multivariate relationships among variables are often unknown. Traditional statistical analysis methods are not well suited to uncovering spatial aspects of these relations. Integration of traditional cartographic methods with those from information visualization and EDA can provide researchers and analysts with a range of tools for visually (and statistically and computationally) exploring these relationships. To enable such integration, we have developed an open source application building environment, GeoVISTA *Studio,* that provides a visual programming interface for application developers to construct analytical tools (and other forms of visualization applications) by quickly integrating Java components (in the form of JavaBeans) – see sidebar.

Figure 2 illustrates the use of a *MultiForm Bivariate Matrix* (part of an application built using *Studio*) to explore spatial and non-spatial relationships in a cancer mortality/risk factor data set. Aggregate county data are depicted for: two potential environmental risk factors (atmospheric emissions for arsenic and mercury), one health care access variable (proportion of individuals without health insurance), and a subset of age-adjusted cancer mortality rate data (for male and female stomach, lung, and esophageal cancer). The *Matrix* extends the well-known scatterplot matrix method into a generic visualization tool that accepts any bivariate representation forms. In this case, bivariate maps and space-filling visualizations are used, with the diagonal depicting univariate maps of each variable.



2. MultiForm Bivariate Matrix with spacefill and map components

Here, we applied a visual classification tool to *bin* the data (for each bivariate representation) into 4 classes: (1) counties with values in the lower ¾ of the data range for both variables (light gray on the maps), (2) counties with values in the highest ¼ of the data range for both variables (dark gray on the maps), (3) counties with values in the top ¼ on the column variable but not the row variable (purple on the map), and (4) counties with values in the top ¼ on the row variable but not the column variable (green on the map). The top row of maps matches data for atmospheric emission of mercury with data for all other variables. In that row, the male lung cancer mortality map contains a broad purple region in the southeast U.S. (indicating that this region is in the top ¼ for lung cancer mortality but not in the top ¼ for mercury emissions). The adjacent map (to the right) contains distinct regions (dark gray) in which the top ¼ female lung cancer mortality rates match with the top ¼ on mercury emissions (most noticeable in the far west, along the Gulf coast, and in Florida).

In the *spacefill* visualization, each county is depicted as a grid cell. In contrast to a scatterplot (a tool most potential users are familiar with) this depiction avoids over-plotting of identical or similar data values. Thus, some relationships which would be obscured in a scatterplot are evident in *spacefill* visualizations. The trade-off is that the tool is less familiar than a scatterplot (for most users), thus requires training to use. In the view shown, scan-line cell order (from lower left to upper right) depicts the column variable and color depicts the row variable (purple indicating values in the top ¼ on that variable). Other orderings (e.g., spiral) are user selectable.

The upper left *spacefill* view shows a strong positive relationship between mercury and arsenic emissions (the purple band at the top of the *spacefill* indicates that the two variables have substantial agreement in the top ¼ of counties). Male-female stomach and lung cancers both show similar (but weaker) relationships, while male-female esophageal cancer shows no relationship.

These and other components developed for integration with GeoVISTA *Studio* support many dynamic events that can be controlled by user action or by input from other components. For example, manual highlighting in any map or *spacefill* display causes selected entities to be highlighted in all displays, the order of Matrix columns and rows can be driven computationally, and manual or computational adjustment of the color scheme assigned to one map can propagate to all coordinated views.

A separate coordinator component (that takes advantage of Java's introspection capabilities) handles these cross-component connections – enabling distribution of visualization functions across software components that do not need to be developed with specific support for cross-component coordination in mind.

Environmental Science

Many of the same EDA methods and tools useful for applications in public health data analysis can also be applied to support research in environmental science. Figure 3 illustrates use of these geovisualization methods on large displays to facilitate collaborative land cover data exploration. The left panel of this large display depicts the design of an application in *Studio,* while the right panel depicts the resulting application. This application includes a dynamically linked scatterplot matrix, parallel coordinate plot, and a self organiz-

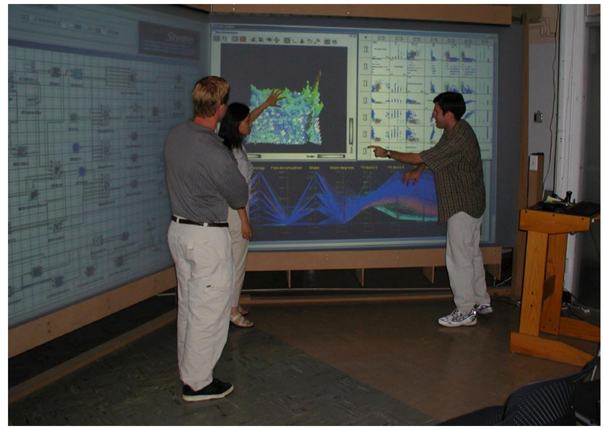


Figure 3. Collaboration in a land-cover classification task. *This dual stereo display (with 2 stereo screens) was created by staff of the Penn State Information Technology Services, Visualization Group, George Otto, Manager.*

ing map (SOM), the latter depicted in a 3D view. While not shown here, the screens can produce stereo views. The application's display depicts use of linked brushing among components (a region of dots selected on a scatterplot is highlighted in all other views, in blue). The analysis session is focused on land cover classification and the task of identifying anomalies in a remotely sensed data set that are resulting in failure of the SOM to distinguish among three similar vegetation types.

Our recent work in environmental applications has combined data visualization methods and tools derived from EDA and cartography with graph-based concept visualization methods and tools derived from Information Visualization. We are developing a distributed concept mapping tool, ConceptVISTA that runs in both a standalone mode on a desktop or handheld device and through a Web portal used for scientific collaboration. Figure 4 depicts a portion of one researcher's concept map representing the vulnerability of people and places to environmental change. Such concept maps provide a vehicle for researcher teams to create and share depictions of complex knowledge. We are developing concept similarity measures for use within ConceptVISTA that help reveal levels of agree-

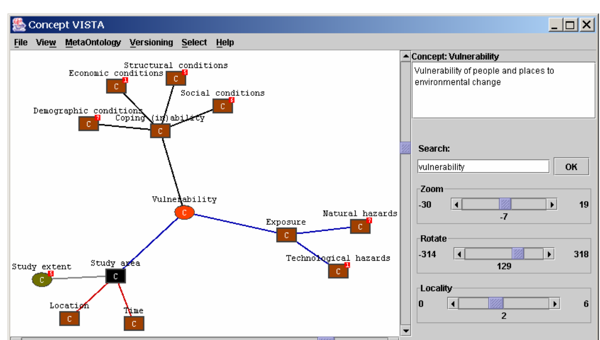


Figure 4. ConceptVISTA -- Concept graph depicting components of water system vulnerability. *This component builds on an open source graph drawing tool called TouchGraph,* www.touchgraph.com

ment between concept maps created by different people or for different problems. ConceptVISTA also has the ability to encode semantic relationships between the researchers, places, data, software tools, and analysis tasks depicted in a map; this information can represent a problem solving approach much as a GeoVISTA *Studio* design does, but at a different level of abstraction. As a result, we anticipate that users will be able to build visual representations of problems using a ConceptVISTA-style interface, which *Studio* can use to select and connect appropriate data and components.

Crisis Management

Geovisualization is not limited to support for science. Rapid advances in geographic information systems and related technologies have created a potential for dynamic geovisualization methods to be integrated with GIS in support of a range of decision making tasks. Crisis management is a prototypical example where a visual, map-based display can be used to integrate, assess, and apply multi-source geospatial information.

In time-critical crisis situations, it is imperative that access to geospatial information is not impeded by constraints inherent in the software or interface. Moreover, Emergency Operations Centers (EOCs) have been outfitted with large screen displays that provide collaborators with up-to-date information about hazards and their impact. In response to both of these factors, new interfaces are needed that enable users who lack GIS training to quickly access complex geospatial information displayed on these large screens. Such interfaces should support untethered access to data exploration tools, such as those shown in figure 3.

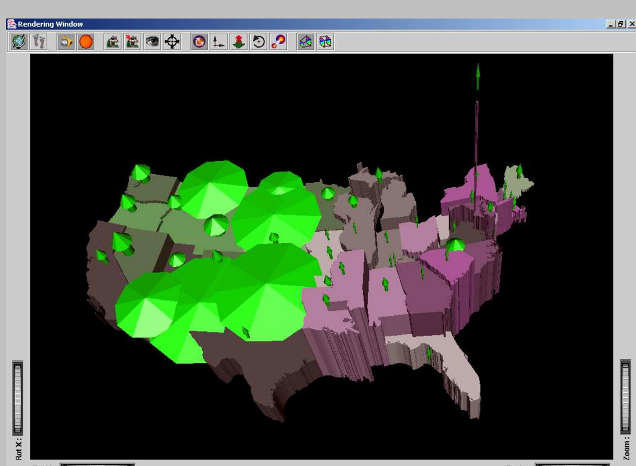
New collaborative geographical visualization environments that support decision-making activities must address two related challenges: (1) the interruptions in cognitive problem solving and collaborative discourse caused by mouse/keyboard input, and (2) the potential for cognitive overload from multiple visualization tools and their controls. First, traditional visualization interfaces (using keyboard and mouse) demand user attention, thus they distract users from thinking about and discussing the problem at hand. Second, geovisualization used in crisis management must often depict complex, multivariate information. Such depictions coupled with a complex interface will force a choice between devoting cognitive resources to understanding the display and understanding the display controls. Particularly for time-critical decision making, it is important to minimize the cognitive resources that must be directed to geovisualization controls.

To make GIS and geovisualization tools more accessible to crisis managers working with large screen map displays, we have integrated solutions from natural language and speech processing, vision-based gesture recognition, and conversational dialogue technologies to enable multimodal dialogues with interactive maps served from geographical information systems [13]. Figure 5 illustrates our Dialogue-Assisted Visual Envi-

GeoVISTA *Studio*

The main objective of the GeoVISTA Studio project is to improve geoscientific analysis by providing an environment that operationally integrates a wide range of problem-solving components and activities, including those both computationally and visually based [2]. Through support for geographic visualization and knowledge discovery, *Studio* enables researchers to explore data, construct hypotheses, discover, refine and test knowledge, construct analyses tasks and evaluate results. It offers a number of specific features and advantages, including: (i) ease of program construction, by visual programming—users drag components from a palette into the "design box" and link them together to create systems which they can run and test in real time, (ii) an open (non-proprietary) architecture based on the JavaBeans environment (iii) a shared code-base—the *Studio* source tree and applications are distributed through SourceForge (http://www.sourceforge.net/projects/geovistastudio), (iv) simple component-based integration using Java introspection methods to expose Bean functionality and a sophisticated event coordination harness that maps user interactions in one component to equivalent actions in others, (v) on-the-fly design modification, and (vi) advanced deployment methods using serialization, automatic application and applet creation and Java WebStart, to facilitate the rapid construction, sharing and deployment of tools developed.

This versatility has the potential to change the nature of systems development, use and deployment for the geosciences, providing better mechanisms to coordinate complex functionality. As a consequence, analyses and decision making processes may be improved by closer integration of software tools and better engagement of the human expert.



Bivariate color represents % age 18-29, % female in each state and the arrows depict % black (height), % divorced (length), and % native American (thickness) ronment for Geoinformation (DAVE_G). In DAVE_G, natural hand gestures and spoken requests are recognized, allowing completely device free interaction [14]. Dialogues between the user and the system are mixed-initiative and collaborative, allowing cognitive loads to be shared between human and the system.

The human-map dialogue in DAVE_G is facilitated by a dialogue manager, which is a computational agent playing the role of an intelligent information assistant, similar to the role of human GIS specialists in current EOCs. The dialogue manager recognizes the user's goals and acts on their user's behalf in spatial data retrieval as well as generation of visual displays. The system is competent in various human-like dialogue strategies for resolving ill-defined requests, ambiguities, and vagueness of spatial concepts [15]. The overall goal of this research is to free the user from the cognitive burden of complete and accurate data query and GIS command specification, allowing smoother, more natural interaction with the geospatial information. We are currently extending the system to support multiple users working collaboratively.

Some challenges

In a recent Viewpoints column, Shalf and Bethel argued that:

> *A new grid-aware framework is needed for distributed visualization that's easy to use, modular, extensible, and permits reuse of existing investments in visualization technology* [16].

Similar challenges must be faced to achieve the goal of distributed geovisualization that crosses the boundaries of software applications, devices, distance, and individual use.

Current geovisualization tools start with an assumption that a user's task will involve geovisualization exclusively (or at least primarily). That is an unrealistic assumption, particularly as geovisualization matures and the potential to play a role in a wide array of activities increases. A component-based approach to geovisualization tools, that distributes functionality among a set of independent modules, has the potential to support more flexible integration of geovisualization with other information access and analysis tools as well as geovisualization that works across devices. The distributed Grid-based architecture that Shalf and Bethel envision is also critical to the challenge of support for same- and different-place collaborative visualization.

Like scientific and information visualization, geovisualization is maturing as a field of research as well as a domain of practice. The potential is there to apply geovisualization as a tool for addressing critical issues in the fields of public health, environmental science, crisis management, and others. Achieving this potential will require multidisciplinary collaboration that integrates perspectives from cartography and GIScience with those from computer graphics, information and scientific visualization, computer-supported cooperative work, diagrammatic reasoning, cognitive science, HCI, cognitive systems engineering, and other domains.

Figure 5. Natural dialogue with a GIS, mediated by a map-based display. *photo by Greg Grieco, Penn State*

Acknowledgement

Research presented in this paper is supported in part by the U.S. National Science Foundation (grants #EIA-9983451, BCS-0113030, SBE-9978052, EIA-0306845) and by the National Cancer Institute (grant number CA95949)

References

[1] A. M. MacEachren, "Visualization in modern cartography: Setting the Agenda," in *Visualization in Modern Cartography*, A. M. MacEachren and D. R. F. Taylor, Eds. Oxford, UK: Pergamon, 1994, pp. 1-12.

[2] M. Gahegan, M. Takatsuka, M. Wheeler, and F. Hardisty, "Introducing GeoVISTA Studio: an integrated suite of visualization and computational methods for exploration and knowledge construction in geography," *Computers, Environment and Urban Systems*, vol. 26, pp. 267-292, 2002.

[3] T.-M. Rhyne, "Does the difference between information and scientific visualization really matter?," *Computer Graphics & Applications*, vol. 23, pp. 6-8, 2003.

[4] A. M. MacEachren and M.-J. Kraak, "Research challenges in geovisualization," *Cartography and Geographic Information Science*, vol. 28, pp. 3-12, 2001.

[5] B. H. McCormick, T. A. DeFanit, and M. D. Brown, "Visualization in scientific computing — A synopsis," *IEEE Computer Graphics & Applications*, vol. July, pp. 61-70, 1987.

[6] J. Bertin, *Graphics and Graphic Information Processing*. Berlin: Walter de Gruyter, 1981.

[7] A. M. MacEachren, "An evolving cognitive-semiotic approach to geographic visualization and knowledge construction," *Information Design Journal*, vol. 10, pp. 26-36, 2001.

[8] D. DiBiase, "Visualization in the earth sciences," *Earth and Mineral Sciences, Bulletin of the College of Earth and Mineral Sciences, Penn State University*, vol. 59, pp. 13-18 (also: www.geovista.psu.edu/publications/others/dibiase90/swoopy.html), 1990.

[9] A. M. MacEachren and J. H. Ganter, "A pattern identification approach to cartographic visualization," *Cartographica*, vol. 27, pp. 64-81, 1990.

[10] A. M. MacEachren and M. J. Kraak, "Exploratory cartographic visualization: Advancing the agenda," *Computers & Geosciences*, vol. 23, pp. 335-343, 1997.

[11] M.-J. Kraak and A. M. MacEachren, "Visualization for exploration of spatial data (editorial introduction to special issue)," *International Journal of Geographical Information Science*, vol. 13, pp. 285-287, 1999.

[12] R. R. Muntz, T. Barclay, J. Dozier, C. Faloutsos, A. M. Maceachren, J. L. Martin, C. M. Pancake, and M. Satyanarayanan, *IT Roadmap to a Geospatial Future, report of the Committee on Intersections Between Geospatial Information and Information Technology*. Washington, DC: National Academy of Sciences, 2003.

[13] R. Sharma, M. Yeasin, N. Krahnstoever, I. Rauschert, A. MacEachren, G. Cai, K. Sengupta, and I. Brewer, "(invited) Role of Speech/Gesture Driven Multimodal Interfaces for Crisis Management," *Proceedings of the IEEE*, in press.

[14] I. Rauschert, P. Agrawal, S. Fuhrmann, I. Brewer, H. Wang, R. Sharma, G. Cai, and A. MacEachren, "Designing a human-centered, multimodal gis interface to support emergency management," presented at ACM GIS'02, 10th ACM Symposium on Advances in Geographic Information Systems, Washington, DC, USA, 2002.

[15] G. Cai, H. Wang, and A. MacEachren, "Communicating Vague Spatial Concepts in Human-GIS Interactions: A Collaborative Dialogue Approach," presented at COSIT 2003: Conference on Spatial Information Theory, Ittingen, Switzerland, in press.

[16] J. Shalf and E. W. Bethel, "The grid and future visualization system architectures," *Computer Graphics & Applications*, vol. 23, pp. 6-9, 2003.