NAHUM GERSHON, STEPHEN G. EICK AND STUART CARD

design

RVI/Rutt Video Interactive Attract Loop Sequence

Information Visualization

Introduction

Visualization provides an interface between two powerful information processing systems—the human mind and the modern computer. Visualization is the process of transforming data, information, and knowledge into visual form making use of humans' natural visual capabilities. With effective visual interfaces we can interact with large volumes of data rapidly and effectively to discover hidden characteristics, patterns, and trends. In our increasingly information-rich society, research and development in visualization has fundamentally changed the way we present and understand large complex data sets. The widespread and fundamental impact of visualization has led to new insights and more efficient decision making.

Much of the previous research in visualization was driven by the scientific community in its efforts to cope with the huge volumes of scientific data being collected by scientific instruments or generated by enormous supercomputer simulations [9]. Recently a new trend has emerged: The explosive growth of the Internet, the overall computerization of the business and defense sectors, and the deployment of data warehouses have created a widespread need and an emerging appreciation that visualization techniques are an essential tool for the broad business and technical communities.

Information visualization deals with new classes of data and their associated analytical tasks in business and information technology areas [10–12, 14]. Information visualization [6, 14] combines aspects of scientific visualization, human–computer interfaces, data mining, imaging, and graphics. In contrast to most scientific data [9], information visualization focuses on information, which is often abstract. In many cases information is not automatically mapped to the physical world (e.g., geographical space). This fundamental difference means that many interesting classes of information have no natural and obvious physical representation. A key research problem is to discover new visual metaphors for representing information and to understand what analytical tasks they support.

Information can come in huge quantities and in fast streams, creating an information avalanche. The largest information space is perhaps the World Wide Web (WWW), which contains millions of pages. Information visualization needs to enable users (e.g., in the commercial and the defense sectors) to get information fast, make sense out of it, and reach decisions in a relatively short time.

Figure 1 SeeSoft *S.G. Eick, J.L Steffen, and E.E Sumner, Lucent Technologies' Bell Lab*

Developments

Document and Software Visualization

Large bodies of text require the users to read many documents, understand them, and make sense of the information in them. Representing the information in a visual form could allow the user to browse through this ocean of information and find interesting pieces of text. Documents have contents and histories. In many cases both could be represented by visualization techniques. Looking at relationships within the contents of large sets of document histories is difficult to trace. A visualization tool, themescapes, was developed at Pacific Northwest National Lab [16]. A themescape is an abstract three-dimensional landscape of information extracted from the relationships among documents. Eick and others at Lucent Technologies/Bell Labs developed a method to represent properties of software programs comprising millions of lines [4]. This system, SeeSoft, uses an arrangement of colored columns and rows (see Figure 1) to represent the frequency of use or other attributes of the particular lines of source code. This system is currently being used to help detect lines of code associated with the Year 2000 problem [5].

Distortion Techniques

Many types of information involve relationships. One common way to visualize structured relationships is to use a graph, with nodes representing the entities and links the relationships between the entities. Graphs work well for small information sets (tens to hundreds of nodes and links) but are easily cluttered and become visually confusing for larger sets. One promising approach for increasing the density of information on graphs involves using distortion lenses, enabling the viewer to see the detail and the general context. A survey of the field as well as a new contribution that extends distortion techniques from two to three dimensions is given by Carpendale et al. [3]. Nevertheless, studies elucidating when distortion techniques are effective are needed.

Hierarchies

The traditional way of depicting hierarchical information is to structure it in a tree-like node and link diagrams. For large trees, however, these diagrams rapidly become cluttered and unusable. One of the earliest instances of information visualization—the cone tree, developed by Card et al. [14]—is a 3D representation of hierarchical information (see Figure 2). Tree

Nahum Gershon
The MITRE Corporation,
McLean, VA
gershon@mitre.org

Stephen G. Eick
Bell Labs/Lucent Technologies
Naperville, IL
eick@research.bell-labs.com

Stuart Card
Xerox Palo Alto Research Center
Palo Alto, CA
card@parc.xerox.com

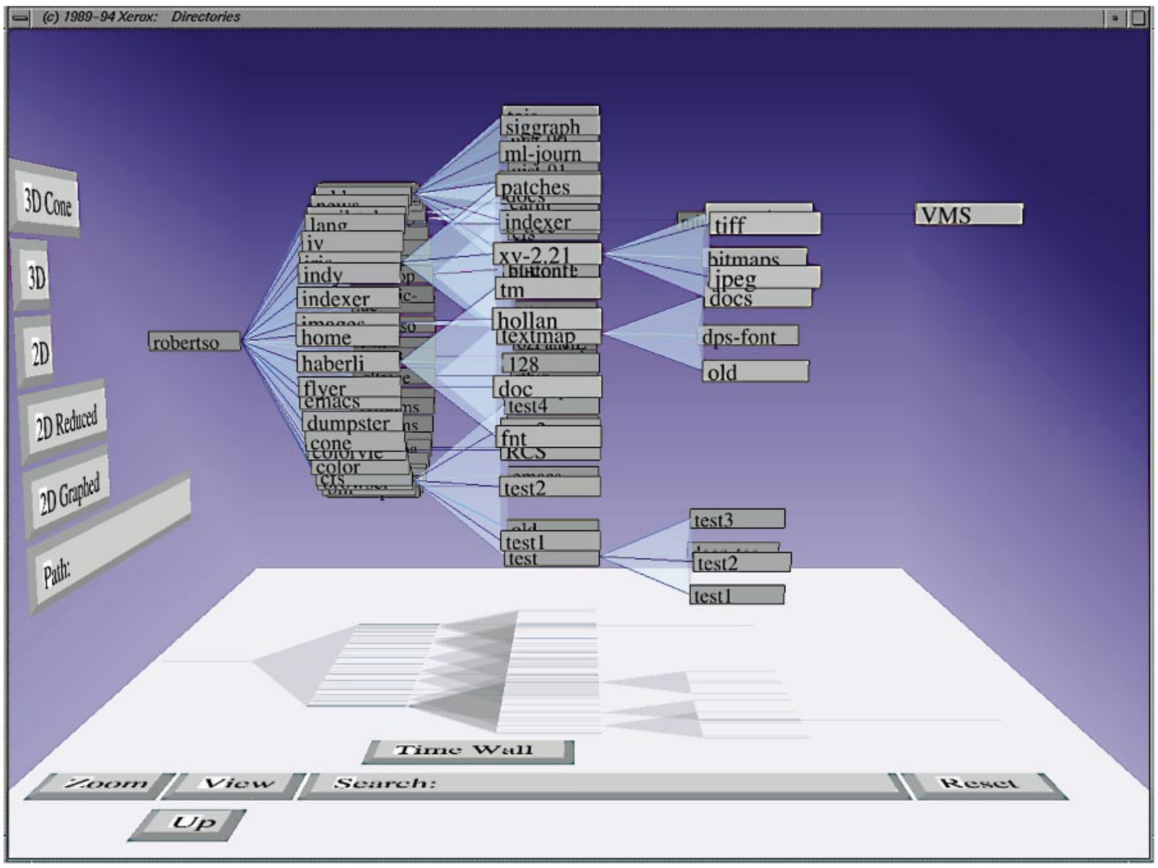

Figure 2 Cone Trees G.G. Robertson, S.K. Card and J.D. Mackinlay (Xerox PARC)

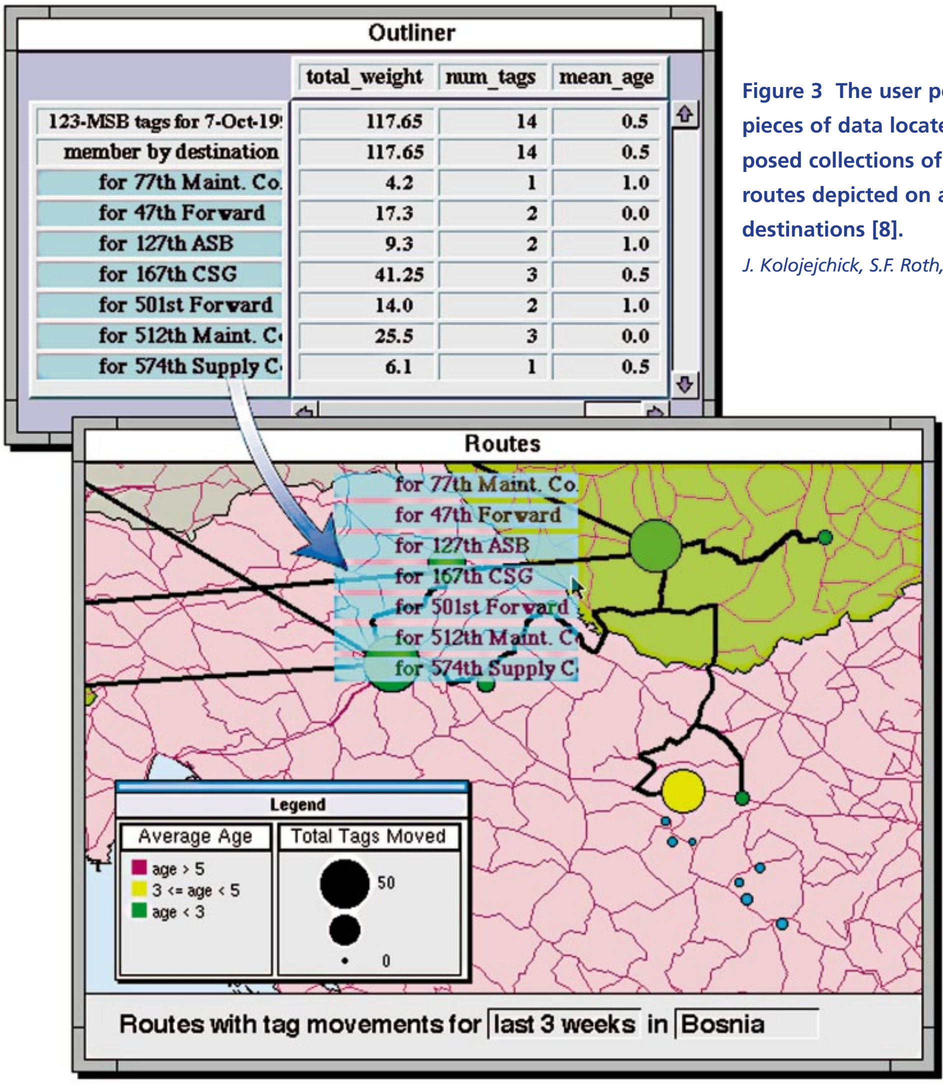

Figure 3 The user performs a recomposition of separate pieces of data located in separate collections. Newly recomposed collections of shipments are dragged to the original routes depicted on a map, where they are shown at their destinations [8].

J. Kolojejchick, S.F. Roth, and P. Lucas (CMU and MAYA Design Group, Inc)

maps, developed by Johnson and Shneiderman, visualize hierarchical information on a two-dimensional display by filling in available screen space [7].

Integrating Information Across Multiple Applications

Over the years, researchers have developed a diverse range of visualization techniques suitable for specific data and information types. Real-life situations may require visualization of information and data from diverse sources, whereas the application might require the use of different visualization tools. The lack of a common set of operations and ways to integrate information across multiple applications is a long-standing problem in research and application development. Kolojejchick et al. [8] have described new ways to create a coordinated suite of basic tools and specialized information appliances. One important contribution of Kolojejchick et al. is to combine many basic analysis and reporting tools into an integrated information workspace (see Figure 3).

World Wide Web

The World Wide Web has enabled the visualization community to provide visualizations to a large-scale audience. It is already changing the way visualizations are delivered to users. Information visualization can also facilitate the user interaction with information residing on the WWW. A good example is WebBook, by Card et al. [2]. WebBook allows rapid interaction with objects at a level higher than pages (e.g., by placing a series of pages in a booklike arrangement). A second example is Web Forager, an application that embeds WebBook and other objects in a hierarchical workspace (see Figure 4).

Using the Web to deliver visualizations means that visualizations are no longer canned products predetermined by their provider. Data could be sent over the network and the visualization can be performed on the user (client) side. Not only can this save time in importing the visualization product over the network, it can also enable the user to tailor the displayed visualization to suit his or her problem, needs, or capabilities. These systems can be implemented using languages and standards such as Java and Virtual Reality Modeling Language (VRML). Another advantage of Web representation is the use of hyperlinks, visually interconnecting points in the visual representation to other objects. Some of these issues, together with some examples of new and potential applications, were addressed by Rohrer and Swing [15].

Applications to Real-World Problems

The importance of information visualization is its potential to solve real-world problems far beyond a purely academic interest. It is exciting to note that this relatively new discipline has already penetrated the commercial market. Some examples of commercial application include

- Solving problems in the financial market [17],
- Using applications of ideas developed at Xerox PARC (e.g., Perspective Wall, Cone Tree, Wide Widgets) to forge the next generation of user interface that will be more visual than current ones [13], and
- Using information visualization to represent knowledge extracted from large databases using data mining to provide decision support in commercial and other environments [1].

Challenges

UNDERSTANDING THE ADVANTAGES AND DISADVANTAGES OF THE NEW MEDIA

The media of visual computing and display are quite new, and we do not understand well their advantages and disadvantages. Many developers and users currently relate to the new medium of visual computing and display as if it were a replica of paper, to which we have grown accustomed over the past thousand years. These new technologies, however, allow us to do certain things beyond what is possible with paper (see e.g., [6]). It will take us time to develop sufficient understanding of these new media.

Emerging Discipline

As a discipline, information visualization is still emerging, tracking the revolution in networking and computing. It is being practiced

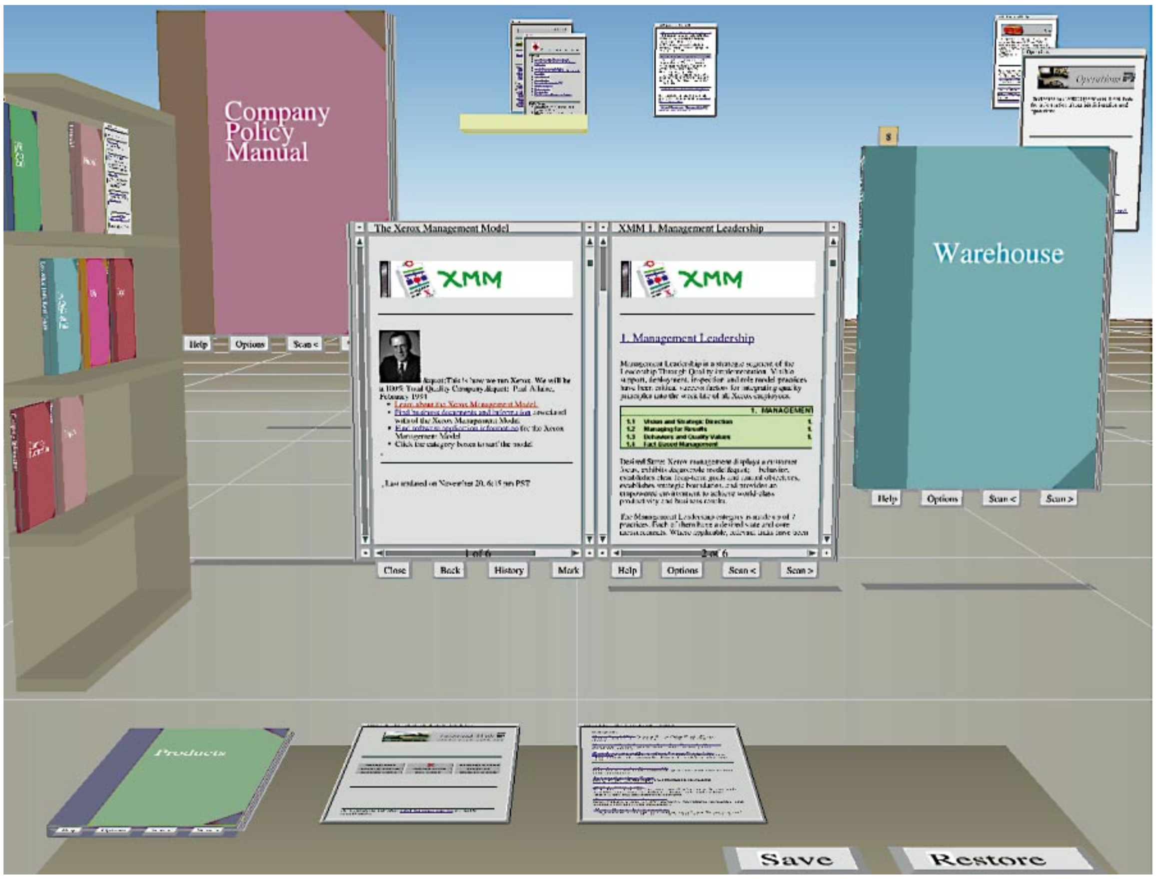

Figure 3 Web Forager *S.K. Card, G.G. Robertson, and W. York (Xerox PARC)*

by skilled practitioners who are hand crafting the current systems. An emerging discipline progresses through four stages. First, it starts as a craft, practiced by artisans using heuristics. Second, researchers formulate scientific principles and theories to gain insights about the processes. Third, engineers eventually refine these principles and insights into production rules. Fourth, the technology becomes widely available. For information visualization many of these stages are happening in parallel.

Ease of Use

In contrast to scientific visualization, which focuses on highly trained scientists, interfaces created for manipulating information may be broadly deployed among a diverse and potentially nontechnical community. The demand for good and effective visualization of information embraces all walks of life and interests. This user community is diverse, based on varying levels of education, backgrounds, capabilities, and needs. We need to enable this diverse group to use visual representations that will be tailored to their specific needs and the specific problem at hand.

Application of Visualization to Real-World Problems

How to use information visualization to solve real-world problems is one of the major challenges of information visualization technology. The motive of a large part of the research and development in this area is academic interest rather than applicability, thus creating a "bag of tricks" (Ken Boff, private communication). An alternative and more demanding approach is to study the problem first and then look for appropriate solutions, for example, using information visualization, if appropriate. We need to realize that in many real-world applications, visualization is just one component of a complex system rather than a stand-alone entity, and we need to understand the system and user needs in order to create effective visualizations.

Three-dimensional vs. Two-dimensional

With the widespread deployment of three-dimensional graphics chips, desktop PCs will soon be capable of much more sophisticated 3D graphics and animations. The challenge, however, is how best to exploit this forthcoming capability. Currently, we do not always understand when 3D is more effective and is not more effective than 2D. As better software is created to make producing visualizations easier, it will be important to use these new capabilities when they are appropriate and information-conveying.

Human- and Usage-Centered Visualization

Tailoring visualization systems that will be seriously based on human capabilities of perception and information processing is another challenge. We need to better understand how human beings interact with information, how they perceive it visually and nonvisually, how the mind works when searching for both known and unknown information, and how it solves problems. A good human-computer interaction is a must, but it is not enough. We also need to better use the information we have about how humans understand and interact with information and the perceptual system in designing visualization systems. Another human-based challenge is the need to learn how to create flexible user interfaces, navigation tools, and search methods that will be appropriate for each of the existing types of users, applications, and tasks.

Text vs. Images

Recent developments in visual display computer hardware on the one hand and computer graphics and visualization methods and software on the other have generated new interest in images and visual representations. "A picture is worth a thousand words," goes the popular saying. However, images may have some disadvantages, and words are sometimes more effective (or powerful) than pictures. The disadvantages of using images include difficulty in representing information clearly; dependency of visual and information perception on memories, experiences, beliefs, and culture; and difficulty in making effective use of color. To use images effectively in science, education, art, and life in general, we need to understand the power and frailty of images. We need to understand when they are equivalent to words, when they are more appropriate to represent information than words, and when they are not. This issue has become extremely important with the spread of the Web, whose many document authors use graphics inappropriately.

The Future

Research and development in the emerging field of information visualization has increased dramatically in the last few years. With hardware and software prices decreasing and speed increasing, more diverse groups of users are going to use this technology. In addition, as the use of the Web becomes more sophisticated and more commonly used, it will have a major impact on the way we deliver information visualization. With the increasing use of visual media, the population's visual literacy will increase, and as a result users will become more comfortable dealing with visuals and thus will get more information from them. Developers, in turn, will include human considerations, creating more and more usable visualization systems. The development of scientific and engineering principles for the generation of visualizations (to users with diverse needs and capabilities) and a methodology for solving problems with information visualization are badly needed.

References

1. Becker, B.G. "Using MineSet for knowledge discovery." *Computer Graphics and Applications* (July-Aug. 1997), pp. 75–78.

2. Card, S.K., Robertson, G.G., and York, W. "The WebBook and the Web Forager: An information workspace for the World Wide Web." In *Proceedings of CHI '96, ACM Conference on Human Factors in Software.* ACM, New York, 1996, pp. 111-116.

3. Carpendale, M.S.T., Cowperthwaite, D.J., and Fracchia, F.D. "Extending distortion viewing from 2D to 3D" *Computer Graphics and Applications* (July–Aug. 1997), pp. 42–51.

4. Eick, S.G., Steffen, J. L., and Sumner, E. E. "SeeSoft—A tool for visualizing software." *IEEE Transactions on Software Engineering 18* (1992), pp. 957–968.

5. Fyock, D.E. "Using visualization to maintain large computer systems." *Computer Graphics and Applications* (July–Aug. 1997), pp. 73–74.

6. Gershon, N. and S.G. Eick. "Visualization's new tack: Making sense of information." *IEEE Spectrum* (Nov. 1995), pp. 38–56.

7. Johnson, B. and Shneiderman, B. "Tree-maps: A space-filling approach to the visualization of hierarchical information structures." In *Proceedings of IEEE Visualization '91,* Computer Society Press, Los Alamitos, CA, 1995, pp. 284–291.

8. Kolojejchick, J., Roth, S.F., and Lucas, P. "Information appliances and tools: Simplicity and power trade-offs in the Visage exploration environment." *Computer Graphics and Applications* (July–Aug. 1997), pp. 32–41.

9. McCormick, B.H, Defanti, T.A., and Brown, M.D. Visualization in scientific computing." *Computer Graphics 21,* 6 (Nov. 1987).

10. *Proceedings of Information Visualization Symposium '95.* Gershon, N. and Eick, S.G. (eds.), Computer Society Press, Los Alamitos, CA, 1995.

11. *Proceedings of Information Visualization Symposium '96.* Card, S., Eick, S.G., and Gershon, N. (eds.), Computer Society Press, Los Alamitos, CA, 1996.

12. *Proceedings of Information Visualization Symposium '97.* Dill, J. and Gershon, N. (eds.), Computer Society Press, Los Alamitos, CA, 1997.

13. Rao, R. "From research to real world with Z-GUI." *Computer Graphics and Applications* (July–Aug. 1997), pp. 71–72.

14. Robertson, G. G., Card, S.K., and Mackinlay, J.D. "Information visualization using 3D interactive animation." *Communications of the ACM 36,* 4 (1993), pp. 57–71.

15. Rohrer, R.M., and Swing, E. "Web-based information visualization." *Computer Graphics and Applications* (July–Aug. 1997), pp. 52–59.

16. Wise, J. A. , Thomas, J.J., Pennock, K., Lantrip, D., Pottier, M., and Schur, A. "Visualizing the non-visual: Spatial analysis and interaction with information from text documents." *Proceedings of Information Visualization Symposium '95,* Gershon, N. and Eick, S.G. (eds.), Computer Society Press, Los Alamitos, CA, 1995, pp. 51–58.

17. Wright, W. "Business visualization applications." *Computer Graphics and Applications* (July–Aug. 1997), pp. 66–70.