VISUALIZING MULTIDIMENSIONAL (MULTIVARIATE) DATA AND RELATIONS -- PERCEPTION vs GEOMETRY

Panelists:

Georges Grinstein, University of Mass.-Lowell & MITRE Co.

Andreas Buja, AT&T Bell Laboratories Murray Hill, NJ

Daniel Asimov, NASA Ames Research Center

Organizer:

Alfred Inselberg, Multidimensional Graphs Ltd, IBM Research & USC Dept of Comp. Sci.

Introduction

The fascination with **dimensionality** predates Aristotle and Ptolemy who argued that space had only three dimensions. By the nineteenth century the new mathematics of Riemann, Lobachevsky and Gauss unshackled the imagination and higher-dimensional geometries came into their own. This together with the abundance of multivariate problems has motivated the desire to augment our perception, limited as it is by the experience of our three-dimensional habitation, and spawned several visualization methodologies. The list of references is by no means exhaustive but indicative of the interest in this field. There is a short literature survey in [29], a beautiful historical review by Tufte in [30], and a more recent one in [31] emphasizing multi-dimensional (or equivalently multivariate) visualization. There are other surveys in [9] and [13]. The taxonomy of the methodologies, though important in its own right, is not our primary concern. Rather, since multidimensionality is not "natural" as a *perceptual experience*, we are interested in visualization methodologies which are:

- perceptually satisfying,
- preserve multidimensional information, and
- are intuitive.

Are any of these goals attainable? Are these criteria mutually exclusive or can we "have it all"? If not, what is attainable and how? These issues set the stage for the debate. We approach them by focusing on the foundations, **concepts before tools** rather than ad hoc methods or flashy demos. Starting with a review of the field, we consider visualization as a collection of transformations from "problem domains" to a perceptual domain, usually visual. This general and somewhat vague notion is given specific interpretations below. A variety rather than a single transformation is considered since various applications may require specialized notions of visualization. Though computer graphics is still it's underpinning, visualization is no longer simply another application but includes, as supporting disciplines, databases, real-time interaction, networking, supercomputing, multimedia, visual programming, systems theory, and human perception.

Extending Visualization from the Pixel to Icons

Position Statement: Georges Grinstein

We will be describing the various approaches taken by researchers over the last five years in extending the presentation of data from a pixel-driven to an icon-driven approach. Chernoff, Ellson and Cox, Beddow, Grinstein, Pickett, Smith, Levkowitz, Ward, Ribarski, all used a two or three dimensional iconographic perception-driven approach when representing multi-dimensional data. We will briefly discuss geometric and color icons and their generation of 2D

(See color plates, page CP-54)

1070-2385/95 $4.00 © 1995 IEEE

Proceedings of the 6th IEEE Visualization Conference (VISUALIZATION '95)
1070-2385/95 $10.00 © 1995 IEEE

textures, area and volume icons and their generations of symbolic objects that then produce rich displays, and further generalizations of this iconographic approach. The iconographic approach has been applied successfully in several areas such as medical imaging and computational fluid dynamics. It extends the concept of mapping data to a pixel to that of mapping data to an object, or icon, whose attributes - such as color, geometry, reflectivity, opacity, and sound - are under the control of the various fields of the data record. This technique allows the information content of the objects represented to have high dimensionality and allows for the fusion of multiple data sets, usually images, into a single integrated multi-modal display. This extension is novel when the selected object is symbolic in nature and when these icons are presented en masse on the screen to produce a texture that can harness human visual perceptual capabilities.

Ellson and Cox used an icon whose displayed values are individually interpretable with attentive focus. Pickett, Grinstein, and Levkowitz initially produced integrated displays harnessing preattentive vision using an icon that was an extension of the linear icons or glyphs often used in psychophysical studies and in plots of scientific data. Variations of this icon have been used to display vector fields, flow fields, and numerous other forms of application information. This icon in its many forms strikes at our ability to perceive line orientation "pre-attentively". Preattentive processing of visual elements is the ability to sense differences in shapes or patterns without having to focus attention on the specific characteristics that make them different. These iconographic displays translate possibly non-visible and undiscovered statistical structures in the data into potentially evident visual structures such as islands, streaks, or gradients. As humans discriminate textures very effectively and use variations in texture as important sources of information including the detection and recognition of objects, the observer sees these statistical properties in the data as qualities of the displayed texture. Work by Beck, Treisman and Gormican, and Enns documents the kinds of differences among elements that are discriminable preattentively.

Other texture generating icons have been developed. The area-based icons are similar to the "acoustic icon" of Stettner and Greenberg and are capable of generating many rich structural textures. This family of icons is derived from an ecologically oriented view of the nature of perception which holds that the primary function of both the visual and auditory systems is to identify the sources of stimuli and their behavior. This view suggests that we should strive to create data displays which engage the automatic perceptual processes underlying the systems which identify the sources of visual and auditory stimuli. Such displays would relieve the user of much of the burden of deliberate, consciously steered analysis. Two recent articles support this point of view, one by Enns on an encoding that makes elements look like real shaded objects in depth, the other by Ramachandran on the use of shadowing in controlling the appearance of displays ([11] and [19]).

"Zoology", Interaction and searching for Gestalt

Position Statement: Andreas Buja

Data visualization is almost always part of a larger context of data analysis and problem solving. Data vis can play a role in data cleaning, data browsing, data reduction, statistical modeling, statistical testing, and in the presentation of data analytic results. With the exception of presentation, all these contexts are exploratory in nature and may require searching through vast numbers of views. The visual search space in data analysis is therefore not one individual view, but the totality of all possible views of the data. The task faced by the data analyst is to find his/her way to the meaningful and informative views in this search space.

Proceedings of the 6th IEEE Visualization Conference (VISUALIZATION '95)
1070-2385/95 $10.00 © 1995 IEEE

What does this mean for research in data visualization? It may mean that devising methods for searching the space of views may be as important as devising intricate rendering methods for individual views. Views are most useful when they are responsive objects that can be easily generated and manipulated. Manipulations on views should support the following three tasks that are part of most data analyses:

1. Searching for Gestalt in data
2. Posing queries about data
3. Comparing views of data

Each of these tasks can be approached in different ways, but there are some approaches that are more natural than others:

1. Searching for Gestalt is achieved by manipulating the focus of individual views, for example by animating projections, as in 3-D data rotations and grand tours.
2. Posing queries can be done graphically, for example, by painting multiple linked views. Painting is a graphical form of posing a query. Linked views are a mechanism for delivering graphical responses to queries.
3. Comparing views requires that many views are arranged in the visual field, for example, side by side, or in matrix layouts.

Existing interactive visualization systems typically provide two out of the three categories of manipulations, but few provide all. We note that the three manipulation categories

— ***focusing, linking, arranging views*** —

are useful independently of the rendering category: They are suitable for interactive computer implementations of scatterplots, parallel coordinates, Andrews curves, glyph presentations, to name a few rendering types that apply to multivariate data. ([11], [12], [13] and, [14]). As a polemic summary, we propose:

"Never mind how we draw pictures, how can we operate on them?"

The Grand Tour : Animating and Interacting with Projections of the Data

Position Statement: Daniel Asimov

The grand tour is a perhaps under-utilized methodology for visualization of multivariate data which is based on the following paradigm: Project the data to a 2-dimensional subspace and display it on the computer screen. Now quickly repeat this process by picking another nearby 2-dimensional subspace, and then another, ad infinitum (or at least as long as someone is willing to view the results). The sequence of subspaces should be chosen so that, eventually, it passes arbitrarily near to every possible subspace.

The result of this is an animation of 2-dimensional projections of the original data. A patient data analyst can watch this animation and often find patterns in the data that would have been undiscoverable by other methods. By recording which projection was used to obtain an "interesting" projection, the data analyst may be able to extract the reason for the interesting projection, and thereby learn something new, and possibly important, about the data.

This methodology has many variants which enable the grand tour to be used efficiently in different contexts with various goals. For example, the image space can be 3-dimensional rather than 2-dimensional. Whenever the 3D image appears to be worthy of further inspection, the tour can be frozen at that stage, and the interesting 3D image can be rotated around in order to understand it better. This offers the possibility of detecting patterns more complex than could be found using only 2D projections.

In another variant, the "guided tour", the user may interact with the tour as it is running to ask it to explore regions of greater interest. In yet another variant, the "automatic tour", it is the

Proceedings of the 6th IEEE Visualization Conference (VISUALIZATION '95)
1070-2385/95 $10.00 © 1995 IEEE

computer which makes the judgments about which projected views are "interesting", according to some criterion that the data analyst has provided in advance. This method, which requires no graphical output, is especially useful for exploring particularly high-dimensional data -- say greater than 8 dimensions -- because the "curse of dimensionality" causes the graphical-output methods to take an extremely long time to scan the enormous number of subspaces necessary to get reasonably close to every possible subspace ([4], [5], [6] and [7]).

Parallel Coordinates -- Visualization of N-Dimensional Geometry

Position Statement: Alfred Inselberg

Visualizing relations involving 2 (real) variables can achieved by mapping the information (data) into two dimensions and identifying the relations with the 2-dimensional patterns (regions, curves) obtained. Inductively then, the visualization of relations among N variables is equivalent to the visualization of N-dimensional Euclidean space.

Our premise is that understanding the underlying geometry of a multivariate problem can provide crucial insights into what is possible and what is not. For example, in 1917 the physicist Paul Ehrehfest proved the in N-space planetary orbits are stable if and only if N = 3, suggesting at least one reason why our physical space is 3-dimensional. In another dimensionality result it was shown that a rigid body rotating in N-space has an axis of rotation only when N is an odd number.

In the spirit of Descartes, we propose visualizing N-dimensional space by using a **coordinate system**. Since orthogonal axes "use up" the plane very fast, we construct a multidimensional system of *Parallel Coordinates* where, in principle, an arbitrary number of axes may appear. In turn, this induces a one-to-one mapping between subsets of Euclidean N-space and subsets of the plane and provides a systematic way for doing **analytic and synthetic N-dimensional geometry**. Relations among N variables (corresponding to hypersurfaces) are mapped into unique regions of the plane whose geometrical properties enable the visualization of the corresponding N-dimensional hypersurfaces. For example, linear relations correspond to finite sets of points, conics map into conics and in general smooth hypersurfaces into well defined planar regions. Further, images of an N-dimensional object(relation) under projective transformation (translations, rotations, scaling, perspective) correspond to 2-d projective transformations. That is, an N-dimensional object can be recognized even when transformed in this way. "Pencil and paper" constructions of all sorts are possible as well as efficient new geometrical algorithms.

The key "global" properties of Parallel Coordinates, relevant to the visualization issues at hand, are summarized below and contrasted with some other methodologies:

- There is no loss of N-Dimensional information. That is, the representation is unambiguous in the sense that no two distinct N-D relations have the same 2-D image (see the color plates at the end for examples of such 2-D patterns). By contrast, projections from N-D to lower dimensional subspaces lose information. For example, the three 2-D projections of a sphere of radius R and the surface which is the intersection of 3 cylinders of radius R, properly centered, are identical. Hence the two **different** surfaces can not be distinguished from their 2-D scatterplots which are identical.
- The representational complexity is $O(N)$. By contrast, the scatterplot matrix display has complexity $O(N^2)$ and poses severe limitations (of time and space) on the size of N that can be handled.
- Every variable is treated in the same way. A display where a different representation is used for each variable (e.g. as in the mapping of variables on attributes of an icon) requires that different "tricks" be learned for each variable.

Proceedings of the 6th IEEE Visualization Conference (VISUALIZATION '95)
1070-2385/95 $10.00 © 1995 IEEE

Further, if the order of the variables is changed (i.e. nose ← → eyes in Chernoff faces) the two displays for the **same** dataset look different and their equivalence may not be recognized.

- The representation is based on rigorous *geometrical* foundations. Hence, with some training, one can build intuition and accurately transform the search for relations among the variables into a 2-D pattern recognition problem.

We conclude with the "anti-polemic": ***You can operate on the picture anyway you want but you are not going to find the information inherently lost by the display method!*** ([15], [22], [23], [24] and [25])

Panelists

Georges Grinstein is Professor of Computer Science at the University of Massachusetts Lowell, Director of the Graphics Research Laboratory, Director of the Institute for Visualization and Perception Research and Principal Engineer with The MITRE Corporation. He received his Ph.D. in Mathematics from the University of Rochester in 1978.

Andreas Buja received a Ph.D. in Mathematics in 1980 from ETH in Zurich. He has held several academic positions in Stanford, Univ. of Washington, then moving to Bellcore, New Jersey and joining in 1994 the AT&T Bell Laboratories.

Daniel Asimov is a Senior Computer Scientist at NASA Ames Research Center and has held academic positions at the Institute for Advanced Study in Princeton, Harvard University, Stanford and UC Berkeley. He received a PhD in Mathematics from UC Berkeley.

Alfred Inselberg has a Ph.D.in Applied Math (1965 Univ. of Illinois Champaign-Urbana), has held several academic positions, and was Senior Technical Corporate Staff Member at IBM Research. He is a consultant in Multidimensional Visualization and Professor at the USC Comp. Sc. Dept..

Proceedings of the 6th IEEE Visualization Conference (VISUALIZATION '95)
1070-2385/95 $10.00 © 1995 IEEE

Bibliography

[1] E. Anderson (1957), *A Semigraphical Method For The Analysis of Complex Problems*, Proc. Nat. Acad, Sci. USA **13**:923-927,

[2] D. F. Andrews (1972), *Plots of High-Dimensional Data*, Biometrics, **29**:125-136,

[3] D. Asimov (1985), *The Grand Tour: A Tool For Viewing Multidimensional Data*, SIAM J. of Sci. Stat. Comp. **6**:128-143,

[4] D. Asimov, A. Buja (1985), *Grand tour methods: an outline*, "Proc. of the 17th Symp. Interface of Comp. Sc. Stat.,

[5] D. Asimov, A. Buja, C. Hurley (1989) *Methods for Subspace Interpolation in Dynamic Graphics*, Bellcore TM,

[6] D. Asimov, A. Buja (1994), *The grand tour via geodesic interpolation of 2-frames, Visual Data Exploration and Analysis*, SPIE, NASA/Bellcore TM,

[7] D. Asimov, A. Buja, D. Cook, C. Hurley (1995) *Theory and Computational Methods for Dynamic Projections in High-dimensional Visualization* to appear,

[8] T. F. Banchoff (1990), **Beyond the Third Dimension**, Scientific American Library, NY,

[9] V. Barnett Edit. (1981), **Interpreting Multivariate Data**, Wiley, New York,

[10] J. Beddow (1990), *Shape Coding of Multidimensional Data on a Microcomputer display* Proc. of IEEE Conf. VISUALIZATION '90, 238-46,

[11] R. D. Bergeron, G. Grinstein (1989), *A Reference Model For The Visualization of Multi-Dimensional Data*, Proc. of Eurographics 89,

[12] J. Bertin (1967), **Semiologie Graphique**, Gauthier-Villars, Paris,

[13] D. Brissom Edit. (1979), **Hypergraphics: Visualizing Complex Relationships in Art, Science and Technology**, Amer. Assoc. for the Adv. of Sc., Westview Press, Boulder,

[11] A. Buja, D. Asimov, C. B. Hurley, J. A. McDonald (1988) *Elements of a Viewing Pipeline for Data Analysis* in **Dynamic Graphics for Statistics**, eds. W. S. Cleveland and M. E. McGill, Wadsworth,

[12] A. Buja, C.B. Hurley (1990), *Analyzing High-Dimensional Data with Motion Graphics* SIAM J. Sc. Stat. Comp. **11**, 1193-1211,

[13] A. Buja, J.A. McDonald, W. Stuetzle (1991) *Viewing High Dimensional Data by Painting Multiple Views* Proc. of IEEE Conf. VISUALIZATION '91,

[14] A. Buja, D. Cook, D. F. Swayne (1995) *Interactive High-Dimensional, Data Visualization* J. of Comp.& Graph. Stat., to appear,

[15] A. Chatterjee, A. Inselberg (1995), **Visualizing Multidimensional Polytopes and Topologies for Tolerances**, Ph.D. Thesis, IBM Research Div. RC 20133,

[16] H. Chernoff (1973), *The Use of Faces to Represent Points in k-Dimensional Space Graphically*, J. Amer. Stat. Assoc. **68**:361-368,

[17] B. Flury, H. Riedwyl (1981) *Graphical Representation of Multivariate Data by Means of Asymmetrical Faces* J. Amer. Stat. Assoc. **76**:757-765,

[18] J. Friedman and J. Tukey (1974), *A Projection Pursuit Algorithm for Exploratory Data Analysis*, IEEE Trans. Comp., **C-23**:881-890,

[19] G. Grinstein and S. Smith (1990), *The Perceptualization of Scientific Data* in

Proceedings of the 6th IEEE Visualization Conference (VISUALIZATION '95)
1070-2385/95 $10.00 © 1995 IEEE

Proc. SPIE Proc. SPIE (Int. Soc. Opt. Eng.) Extracting Meaning from Complex Data: Processing, Display, Interaction, Santa Clara, CA, **1259**:190-9,

[20] H. Hinterberger (1987), **Data Density : A Powerful Abstraction to Manage and Analyze Multivariate Data**, Ph. D. Thesis, ETH , #4 VDF, Zurich,

[21] P. J. Huber (1985), *Projection Pursuit* Ann. Stat. **13**:435-475,

[22] A. Inselberg (1985), *The Plane with Parallel Coordinates*, Visual Computer, **1**:69-97,

[23] A. Inselberg, B. Dimsdale, A. Chatterjee, C.K. Hung (1993), *Parallel Coordinates : Survey of Recent Results*, Proc. of SPIE Conf. San Jose, CA, Jan. 1993, **1913**, 582-599,

[24] A. Inselberg, B. Dimsdale (1994), *Multi-Dimensional Lines I: Representations*, SIAM J. on Appl. Math., **54:2**, 559-577,

[25] A. Inselberg, B. Dimsdale (1994), *Multi-Dimensional Lines II: Proximity Applications* SIAM J. on Applied Math., **54:2**, 578-596,

[26] B. Kleiner, J. A. Hartigan (1981), *Representing Points in Many Dimensions by Trees and Castles* J. Amer. Stat. Assoc. **76**:260-269,

[27] J. LeBlanc, M. O. Ward, N. Wittels (1990), *Exploring N-Dimensional Databases* Proc. of IEEE Conf. VISUALIZATION '90, 230-7,

[28] T. Mihalisin, E. Gawlinski, J. Timlin, J. Schwegler (1990), *Visualizing a Scalar Field on an N-Dimensional Lattice* Proc. of IEEE Conf. VISUALIZATION '90, 255-62,

[29] S. J. Schwager (1986), *Applying Statistical Graphics to Multivariate Data* Proc. NCGA '86 Anaheim, CA 704-712,

[30] E. R. Tufte (1983), **The Visual Display of Quantitative Information**, Graphic Press, Cheshire, Conn.,

[31] E. R. Tufte (1990), **Envisioning Information**, Graphic Press, Cheshire, Conn.,

[32] J. W. Tukey, P.A. Tukey (1979), *Graphical Display for Data sets in 3 or more Dimensions* in [13],

Proceedings of the 6th IEEE Visualization Conference (VISUALIZATION '95)
1070-2385/95 $10.00 © 1995 IEEE