Procedural Shape Generation for Multi-dimensional Data Visualization

David S. Ebert[1], Randall M. Rohrer[2], Christopher D. Shaw[3], Pradyut Panda[1],
James M. Kukla[1], and D. Aaron Roberts[4]

[1] CSEE Department, University of Maryland Baltimore County
1000 Hilltop Circle, Baltimore, MD 21250, USA
{ebert,panda,jkukla1}@csee.umbc.edu

[2] Department of EECS, The George Washington University
Washington, DC, 20052, USA
rohrer@seas.gwu.edu

[3] Department of Computer Science, University of Regina
Regina, Saskatchewan, S4S 0A2, Canada
cdshaw@cs.uregina.ca

[4] NASA Goddard Space Flight Center, Mailstop 692.0
Greenbelt, MD, 20771, USA
roberts@vayu.gsfc.nasa.gov

Abstract. Visualization of multi-dimensional data is a challenging task. The goal is not the *display* of multiple data dimensions, but *user comprehension* of the multi-dimensional data. This paper explores several techniques for perceptually motivated procedural generation of shapes to increase the comprehension of multi-dimensional data. Our glyph-based system allows the visualization of both regular and irregular grids of volumetric data. A glyph's location, 3D size, color, and opacity encode up to 8 attributes of scalar data per glyph. We have extended the system's capabilities to explore shape variation as a visualization attribute. We use procedural shape generation techniques because they allow flexibility, data abstraction, and freedom from specification of detailed shapes. We have explored three procedural shape generation techniques: fractal detail generation, superquadrics, and implicit surfaces. These techniques allow from 1 to 14 additional data dimensions to be visualized using glyph shape.

1 Introduction

The simultaneous visualization of multi-dimensional data is a difficult task. The goal is not only the display of multi-dimensional data, but the *comprehensible display* of multi-dimensional data. Glyph, or iconic, visualization is an attempt to encode more information in a comprehensible format, allowing multiple values to be encoded in the parameters of the glyphs [10]. The shape, color, transparency, orientation, etc., of the glyph can be used to visualize data values. Glyph rendering [10, 11] is an extension to the use of glyphs and icons in numerous fields, including cartography, logic, and pictorial information systems.

In previous work, we explored the usefulness of stereo-viewing and two-handed interaction to increase the perceptual cues in glyph-based visualization. The Stereoscopic Field Analyzer (SFA) [4] allows the visualization of both regular and irregular grids of volumetric data. SFA combines glyph-based volume rendering with a two-handed minimally-immersive interaction metaphor to provide interactive visualization, manipulation, and exploration of multivariate, volumetric data. SFA uses a glyph's location, 3D size, color and opacity to encode up to 8 attributes of scalar data per glyph. These attributes are used when a vector visualization is not appropriate, such as when displaying temperature and pressure at each glyph. We are extending this work to combine glyph rendering with other visually salient features to increase the number of data dimensions simultaneously viewable.

2 Background

Our use of glyph attributes for visualization is based on human perceptual abilities. Color, transparency, position, and size are perceptually significant visualization attributes that are commonly used in visualization systems. Shape is a more challenging visualization attribute because three-dimensional shape perception is not as well understood as color, size, and spatialization perception. Most evidence suggests that shape variation is a valuable perceptual attribute that we can harness for visualization. Experiments show that humans can pre-attentively perceive three-dimensional shape [8]. Cleveland [3] cites experimental evidence that shows the most accurate method to visually decode a quantitative variable in 2D is to display position along a scale. This is followed in decreasing order of accuracy by interval length, slope angle, area, volume, and color. Bertin offers a similar hierarchy in his treatise on thematic cartography [2].

Our visualization system already utilizes glyph position in 3D, 3D scale (corresponding to Cleveland's length, area and volume) and color. Slope angle is a difficult dimension to use in an interactive system because of arbitrary orientation of the data volume. Therefore, the next opportunity for encoding a scalar value is shape.

One of the most difficult problems in glyph visualization is the design of meaningful glyphs. Glyph shape variation must be able to convey changes in associated data values in a comprehensible manner [10]. This difficulty is sometimes avoided by adopting a single base shape and scaling it non-uniformly in 3 dimensions. However, the lack of a more general shape interpolation method has precluded the use of shape beyond the signification of categorical values [2]. This paper describes three techniques we have explored for the procedural generation of glyph shapes for glyph-based volumetric visualization [7].

3 Perceptually-based Mapping of Shape Attributes

Glyph shape is a valuable visualization component because of the human visual system's pre-attentive ability to discern shape. Shapes can be distinguished at the pre-attentive stage [8] using curvature information of the 2D silhouette contour and, for 3D objects, curvature information from surface shading. Unlike an arbitrary collection of icons, curvature has a visual order, since a surface of higher curvature looks more jagged than a surface of low curvature. Therefore, generating glyph shapes by maintaining control of their curvature will maintain a visual order. This allows us to generate a range of glyphs which interpolate between extremes of curvature, thereby allowing the user to read scalar values from the glyph's shape. Pre-attentive shape recognition allows quick analysis of shapes and provides useful dimensions for comprehensible visualization.

Our use of glyphs is related to the idea of marks as the most primitive component that can encode useful information [2]. Senay points out that shape, size, texture, orientation, transparency, hue, saturation, brightness, and transparency are retinal properties of marks that can encode information [13, 14]. Bertin has studied the use of marks for two-dimensional thematic maps and gives examples of how shape can be misused in the rendering of these maps [2]. In his examples, shapes are used to represent purely categorical data and, for this reason, he uses a small collection of distinct icons such as star, cross, square, circle, triangle, and so on. Because each individual shape does not have any inherent meaning, the reader is forced to continually look up the shape's meaning in the map legend. The main difficulty is that a collection of arbitrary icons does not have any necessary visual order, and so any assignment of shape to meaning is equivalent.

4 Procedural Shape Visualization

We have explored three different procedural techniques for the generation of glyph shape: superquadrics, fractal detail, and implicit surfaces. All three techniques use a procedural approach for glyph design to solve the complex problem of meaningful glyph design. Procedural shape generation techniques provide flexibility, data abstraction, and freedom from specification of detailed shapes [5]. Procedural techniques allow the shape to be controlled by high-level control parameters. The user changes the glyph shape from a more directorial, indirect aspect, where he or she is unburdened from the full explicit specification of detailed shapes. Our goal for glyph design was to allow the automatic mapping of data to shape in a comprehensible, easily controllable manner.

4.1 Fractal shape detail

One simple procedural technique for shape visualization is to generate distorted versions of a basic shape, such as a cube, where the amount of deviation from the basic shape is proportional to the data dimension being visualized: low data values will map to the base shape and high data values will map to very perturbed shapes. Displacement mapping is used to create the deviation from the base shape. We used a fractional Brownian motion (fBM) [6] turbulence function to create the displacement from the original surface. The main idea with this technique is to add a high frequency component to the shape while not distracting from the perception of the low frequency base shape perception. This addition of high frequency information does not detract from the overall spatial pattern of the data, which can occur from the generation of non-related shapes. Figure 1 shows the results of this technique applied to a simple random test data set where the data values near zero are cuboid and the “fuzzy” cubes have data values near one.

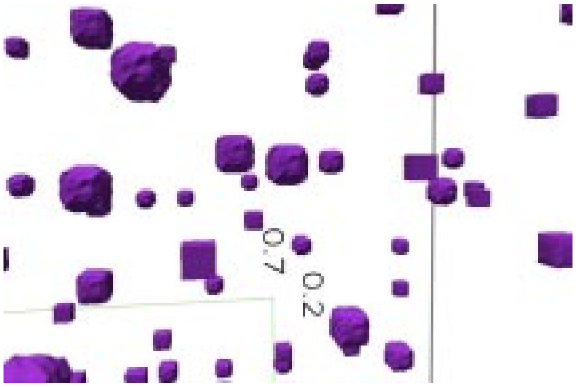

Fig. 1. Example fractal displacement of a base cube shape. The data values are a random data set, with values ranging from 0 (smooth cube) to 1 (fuzzy cube).

4.2 Procedural shape visualization using superquadrics

Superquadrics are a natural tool for automatic shape visualization that can allow from one to two data dimensions to be visualized with shape. Superquadrics, first introduced to computer graphics by Barr [1], are extensions of quadric surfaces where the trigonometric terms are each raised to exponents. Superquadrics come in four main families: hyperboloid of one sheet, hyperboloid of two sheets, ellipsoid, and toroid. For our initial implementation we have chosen superellipses due to their familiarity, but the system can be easily extended to use other types of superquadrics as well as combinations of types. For example, supertoroids could be used for negative values and superellipsoids for positive values.

In the case the of superellipsoids, the trigonometric terms are assigned exponents as follows:

$$\underline{x}(\eta,\omega) = \begin{bmatrix} a_1 \ \cos^{\epsilon 1}\eta \ \cos^{\epsilon 2}\omega \\ a_2 \ \cos^{\epsilon 1}\eta \ \cos^{\epsilon 2}\omega \\ a_3 \ \sin^{\epsilon 1}\eta \end{bmatrix}, \quad \begin{array}{c} -\pi/2 \le \eta \le \pi/2 \\ -\pi \le \omega < pi \end{array}$$

These exponents allow continuous control over the characteristics (the concavity or convexity) of the shape in the two major planes which intersect to form the shape, allowing a very simple, intuitive, abstract schema of shape specification. For example, $\epsilon 1 < 1$ and $\epsilon 2 < 1$ produces cuboid shapes, $\epsilon 1 < 1$ and $\epsilon 2 \sim 1$ produces cylindroid shapes, $\epsilon 1 > 2$ or $\epsilon 2 > 2$ produces pinched shapes while $\epsilon 1 = 2$ or $\epsilon 2 = 2$ produces faceted shapes. As can be seen in Figure 2, varying the exponents achieves smooth, understandable transitions in shape. Therefore, mapping data values to the exponents provides not only a continuous, automatic control over the shape's overall flavor, but a comprehensible shape mapping as well.

To produce understandable, intuitive shapes, we rely on the ability of superquadrics to create graphically distinct [13, 14], yet related shapes. We encode two data dimensions in glyph shape in a manner that allows the easy separation of the shape characteristics.

4.3 Procedural shape visualization with implicit surfaces

Implicit surfaces are a powerful geometric modeling tool which use an underlying density field to represent volumetric information. Implicit techniques provide smooth blending functions for individual implicit field sources [5]. Isosurface generation techniques are then used to create a geometric representation of this volumetric density field. This natural, smooth blending of multiple implicit sources makes implicit surfaces a natural choice for shape visualization.

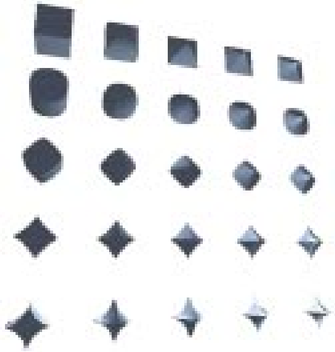

Fig. 2. Example superquadric shapes created by varying each exponent from 0 to 4.

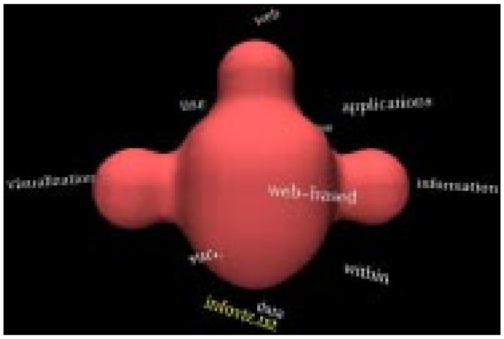

Fig. 3. Example implicit surface shape for a document on "Web-based Information Visualization." The three main bulges correspond to the frequency of the terms "web", "visualization", and "information".

For implicit shape visualization, we map up to 14 data dimensions to uniformly spaced vectors emanating from the center of a sphere. The length of each of these vectors is scaled based on the data value being visualized. At the end of each of these vectors, we place a uniform point source field function. We then generate an isosurface (typically, $F = 0.5$) from the resulting density field to create the glyph shape. This produces "blobby" shapes with bulges in a given direction indicating large data values for that dimension of the data. An example of the a single blobby glyph of this type can be seen in Figure 3.

Since size and spatial location are more significant cues than shape, the importance mapping of data values should be done in a corresponding order. In decreasing order of data importance, data values are mapped to location, size, color, and shape. In our experience, shape is very useful for local area comparisons among glyphs: seeing local patterns, rates of change, outliers, anomalies.

5 Implementation and Results

5.1 Information Visualization Results

We have applied procedurally-generated glyph shapes to the visualization of both scientific and information data. For information visualization, we have chosen an example of the visualization of "thematic" document similarities. Figure 4 shows a visualization of document similarities generated with the Telltale system [9]. The document corpus consists of 1883 articles from *The Wall Street Journal* from September and October 1989. Each glyph in Figure 4 represents a document in the corpus, and the document's X, Y, and Z position, color and shape each represent the similarity of the document to one of the 5 themes.

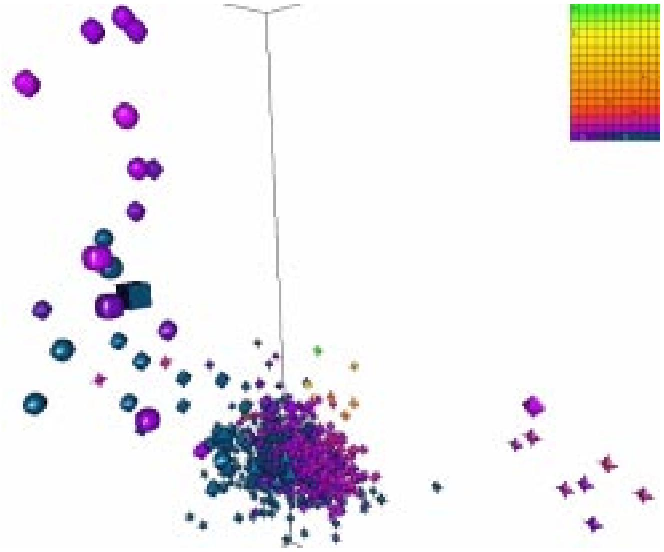

Fig. 4. Three-dimensional visualization of 1833 documents' relationship to gold prices, foreign exchange, the federal reserve, stock prices, and Manuel Noriega.

Document similarity to *gold prices*, the *foreign exchange rate of the U.S. dollar*, and *federal reserve* are respectively mapped to the X, Y, and Z axes. The Y axis is visually indicated in Figure 4 by the vertical line, with the X axis going to the right and the Z axis going to the left. The bulk of the documents have very low similarity to all of these 3 themes, so their glyphs are clustered near the origin at the bottom center.

The documents outside this cluster exhibit two spatial patterns: a cluster of 9 documents to the bottom right and a vertical branch on the left. The right cluster indicates the small number of documents in the corpus that discuss both *gold prices* and the *foreign exchange rate of the U.S. dollar*. The vertical branch depicts a larger collection of documents that discuss both *foreign exchange rate of the U.S. dollar* and the *federal reserve*.

A fourth attribute, similarity to *stock prices*, is inversely mapped to both superquadric exponents of the glyph shape, with highest similarity creating cuboids, then spheres, diamonds, and stars (lowest). Referring to the square array of sample glyphs in Figure 2, the similarity to *stock prices* maps to glyphs on the diagonal from the upper left to the lower right of Figure 2, with upper left indicating high similarity, and lower right indicating low similarity.

In Figure 4 the larger, rounder shapes along the vertical branch exhibit some significant relationship to *stock prices* while the more numerous star-shaped glyphs do not. Clearly the vertical branch contains articles relating *foreign exchange*, *federal reserve* and *stock prices*.

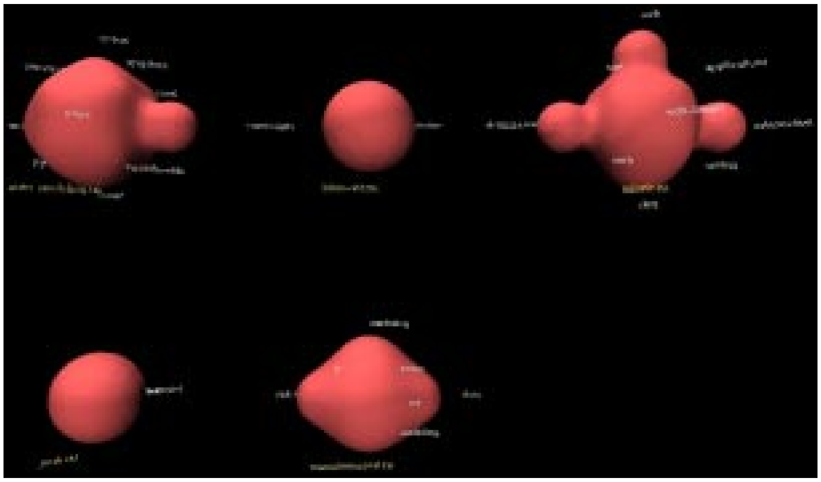

Fig. 5. Multiple documents term frequency visualized as implicit surface shapes. The document in the upper left and the upper right both have a high frequency of the term “information” (bulge to the right). The upper right blob is the same as that in figure 3

Glyph color is mapped inversely to similarity to *Manuel Noriega*. Most of the documents fall in the turquoise and purple range, indicating no significant relationship. However, the documents in the orange, red, and yellow-green range represent documents with a significant relationship to *Manuel Noriega*. Many of these documents mention the effect of the coup attempt against *Manuel Noriega* and its effect on *foreign exchange rate of the U.S. dollar* (vertical axis). The fact that these orange, red, and yellow-green documents are not in either of the branches indicates that these articles did not relate heavily to either *federal reserve* or *gold prices*, and their star shape indicates no relationship with *stock prices*.

We have also used the implicit surface technique for text-based information visualization. The input data for this visualization was word frequency from a collection of text documents [12]. The frequency of the 14 most frequent words were mapped to vector length in positioning the density field sources: documents with common word usage exhibit similar shapes. The results of this process can be seen in Figure 5. From quick examination of this figure, documents with similar topics and word usage are easily distinguishable.

5.2 Scientific Visualization Results

We have used this system to examine several scientific visualization data sets. Figure 6 shows the visualization of a magnetohydrodynamics simulation of the solar wind in the distant heliosphere (20 times the distance of the Earth to the Sun). The simulation data is a 64 × 64 × 64 grid containing the vector vorticity and velocity for the simulation.

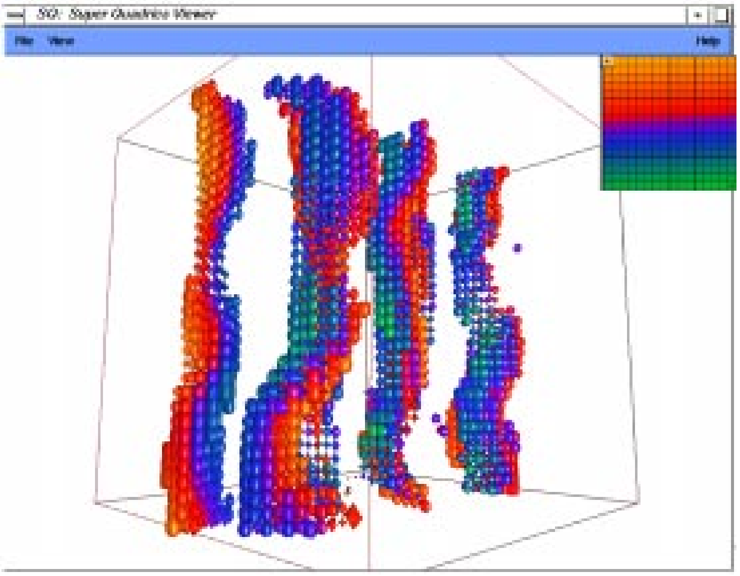

Fig. 6. Visualization of a magnetohydrodynamics simulation of the solar wind in the distant heliosphere showing both velocity components and vorticity components of 6 vortex tubes.

Opacity is used to represent vorticity in the j direction, so that the 6 vortex tubes (only 4 are visible) represent zones in space where this vorticity is somewhat larger than zero. Glyph shape is based inversely on the velocity in the j direction. Positive velocities are displayed as larger, rounder to cuboid shapes and negative velocities are displayed as spiky, star-like shapes. Zero velocity is represented by the diamond shape. The overall columnar pattern of the data is not disturbed by the introduction of the shape mapping, but the velocity variation can still be seen as we traverse the lengths of the tubes. In this case, values close to zero in terms of j vorticity (still fluid) have been masked out.

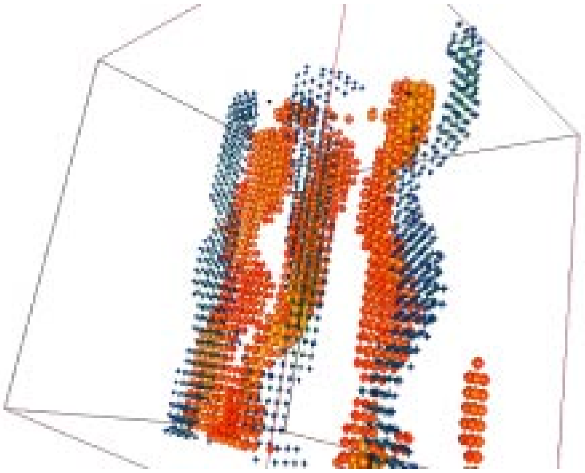

Fig. 7. Visualization of a magnetohydrodynamics simulation of the solar wind in the distant heliosphere displaying 3 vortex tubes with positive j vorticity (cuboids and ellipsoids) and 3 vortex tubes with negative j vorticity (stars).

Figure 7 is a visualization of the same magnetohydrodynamics data, but with the opacity, color, and glyph shape all mapped to the j component of vorticity. Negative vorticity components produce concave shapes (blue stars), while positive values produce convex shapes (orange cuboids and ellipsoids). The use of this data mapping clearly shows three tubes with negative j vorticity and three tubes with positive j vorticity.

6 Conclusions

We have developed several new techniques for intuitive, comprehensible creation of glyph shapes. These techniques are based on procedural shape generation and increase the number of dimensions of data that can be comprehensibly visualized in a glyph-based visualization system. Superquadric functions, fractal surface displacement, and implicit surfaces have been shown to be useful techniques for the automatic generation of glyph shapes and the visualization of multi-dimensional data. Superquadrics are a very natural technique for shape visualization of one to two data dimensions. Fractal surface displacement seems useful for small categorizations (4 to 6 easily discernible shapes) of a single data dimension and can actually be added to the other two techniques as a secondary shape cue. Implicit surface techniques show great promise in visualizing between eight and twenty data dimensions with shape. All of these procedural shapes allow the intuitive understanding of data variation among glyphs, while preserving the global data patterns. We have shown the value of these techniques for both multi-dimensional information and scientific visualization.

References

[1] A. Barr. Superquadrics and angle-preserving transformations. *IEEE Computer Graphics and Applications*, 1(1):11–23, 1981.

[2] J. Bertin. Semiology of graphics, 1983.

[3] William S. Cleveland. *The Elements of Graphing Data.* Wadsworth Advanced Books and Software, Monterey, Ca., 1985.

[4] David Ebert, Chris Shaw, Amen Zwa, and Cindy Starr. Two-handed interactive stereoscopic visualization. *Proc. IEEE Visualization '96*, Oct. 1996.

[5] David S. Ebert. Advanced geometric modeling. *The Computer Science and Engineering Handbook*, Allen Tucker Jr., ed., chap. 56. CRC Press, 1997.

[6] David S. Ebert, F. Kenton Musgrave, Darwyn Peachey, Ken Perlin, and Steven Worley. *Texturing and Modeling: A Procedural Approach, Second Edition.* AP Professional, 1998.

[7] J. D. Foley and C. F. McMath. Dynamic process visualization. *IEEE Computer Graphics and Applications*, 6(3):16–25, March 1986.

[8] Andrew Parker, Chris Cristou, Bruce Cumming, Elizabeth Johnson, Michael Hawken, and Andrew Zisserman. The Analysis of 3D Shape: Psychological principles and neural mechanisms. In Glyn Humphreys, editor, *Understanding Vision*, chapter 8. Blackwell, 1992.

[9] Claudia E. Pearce and Charles Nicholas. TELLTALE: Experiments in a Dynamic Hypertext Environment for Degraded and Multilingual Data. *Journal of the American Society for Information Science (JASIS)*, 47, April 1996.

[10] Frank J. Post, Theo van Walsum, Frits H. Post, and Deborah Silver. Iconic techniques for feature visualization. In *Proceedings Visualization '95*, pages 288–295, October 1995.

[11] W. Ribarsky, E. Ayers, J. Eble, and S. Mukherja. Glyphmaker: creating customized visualizations of complex data. *IEEE Computer*, 27(7):57–64.

[12] Randall M. Rohrer, David S. Ebert, and John L. Sibert. The Shape of Shakespeare: Visualizing Text using Implict Surfaces. In *Proceedings Information Visualization 1998*, pages 121–129. IEEE Press, 1998.

[13] H. Senay and E. Ignatius. A knowledge-based system for visualization design. *IEEE Computer Graphics and Applications*, 14(6):36–47, November 1994.

[14] H. Senay and E. Ignatius. Rules and principles of scientific data visualization. *ACM SIGGRAPH HyperVis Project, 1996.*