Interactive Poster: 3D ThemeRiver

Peter Imrich[1] Klaus Mueller[1] Dan Imre[3] Alla Zelenyuk[3] Wei Zhu[2]

[1] Center for Visual Computing, Computer Science, Stony Brook University
[2] Applied Mathematics and Statistics, Stony Brook University
[3] Environmental Sciences, Brookhaven National Laboratory

Abstract

A limitation of the existing ThemeRiver [1] paradigm is that only one attribute can be displayed per theme. In this poster, we present a 3D extension, which enables us to display two attributes of each variable in the data stream. We further describe a technique to construct the Bezier surface that satisfies the ThemeRiver requirements, such as boundedness and preservation of local extrema.

1 Introduction

The ThemeRiver visualization traditionally displays different variables as distinctly colored data streams. The streams usually flow along the time axis and their width reflect the attribute of a particular stream at a particular point in time. This attribute can be anything worthwhile investigating, such as time fluctuations of different company stock values, ranging from simple distributions to more complex variables. The main advantage of a ThemeRiver visualization is that it portrays different data groups simultaneously, revealing their co-variance, showing how they behave together. An example of a 2D ThemeRiver visualization is shown in Fig. 1.

The 3D counterpart that we propose in this paper extends this idea and maps a second attribute, such as the revenue of the companies, as the height of the streams. Thus, the x-axis represents the time, the y-axis the stock price and the z-axis the company revenue. In short, 3D ThemeRiver is naturally suited to exhibit any sequential ternary covariate trends, mapping one quality as width and another as height. It is suited to correlate data episodes and environment.

2 Construction

Our 3D ThemeRiver is represented by a composite Bezier surface. Hence, the entire following discussion will revolve around the placement of the Bezier control points so that the resulting surface truly reflects the underlying data.

A very important property of the correct surface is that it needs to preserve the extreme points in the dataset. This constraint is also maintained for spline curves in the original 2D ThemeRiver application. In other words, it is undesirable to violate local maxima or minima, and it is important to control surface inflections. A rule of thumb in this situation is to find a surface that does not overshoot its four corner points of any of its Bezier patches. Similarly, the curvature of stream boundaries has to preserve the same extreme points. This is illustrated in Fig. 2.

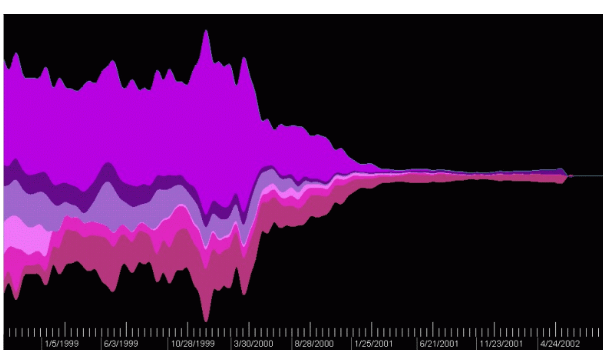

Figure 1: Traditional 2D ThemeRiver view on a few select dot.com company stocks in the period January 99 - April 2002.

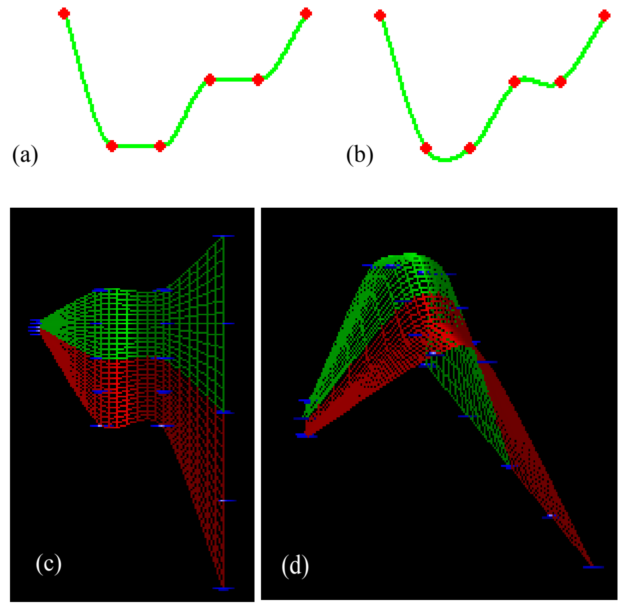

Figure 2: Bezier curves and surfaces around extreme points. (a) correctly interpolated datapoints, (b) the same curve with inflections and incorrect maxima, (c) top view of bezier surface that violates width extremeness, the lower boundary of the red stream contains two inflections and (d) the same surface viewed from a profile, here we see an incorrect peak.

Another more obvious requirement for the final surface is that it should be smooth. This can be achieved by placing the neighboring control points of adjacent patches into a co-planar configuration, forming preferable a parallelogram. This achieves C_2 continuity.

To satisfy the above, we represent each stream interval by two Bezier patches. The lower patch shares a boundary with the lower neighboring stream, and the upper patch shares a boundary with upper neighboring stream. The center of the stream lies on the edge shared by these two patches. This way, the stream boundaries as well as the stream troughs do not violate the previously stated ThemeRiver constraints. Both lie on edges of the two patches. (Recall that a Bezier patch passes through its four corner control points and interpolates the rest.)

Given a height field, the procedure that constructs the ThemeRiver bezier surface is as follows.

- Generate the boundary points - the height of a boundary can simply be a linear interpolation of heights of adjacent stream centers.
- Stack and center the data streams.
- Compute the placement of control points. The corner points are directly given by the data. Points along the edges can be determined solely from the positions among the corner control points along the same edges. Finally, the diagonal points only depend on the local slopes and their displacement from their closest corner point. The slopes at each corner point are designed either to pre-

email: {imrich, mueller}@cs.sunysb.edu, {imre, alla}@bnl.gov, zhu@ams.sunysb.edu

serve local extrema (slopes in these cases are zero in whatever direction the extrema occurs) or to blend the overall slopes of neighboring patches (overall slopes can be estimated by looking only at the positions of corner points of the involved patches).

3 Domain Application

Our particular application deals with the survey and analysis of a large collection of millions of digitized aerosol particle spectra. Our data comprise millions of 450-bin molecular mass spectrum for each individual particles, along with their total mass, a time stamp and score of environment variables like humidity, ozone concentrations and others. All of these make up a 500-D feature vector for each particle.

Our 3D ThemeRiver is part of a comprehensive data mining and data clustering package for aerosol data that we have developed at BNL. Atmospheric scientists use the 3D ThemeRiver application to visualize time-variant or other environmental trends in context to the data clusters. An example of such an interactive display is shown in Fig. 3. The display is linked to the classification engine and display, and scientists can interactively modify the variables and streams displayed.

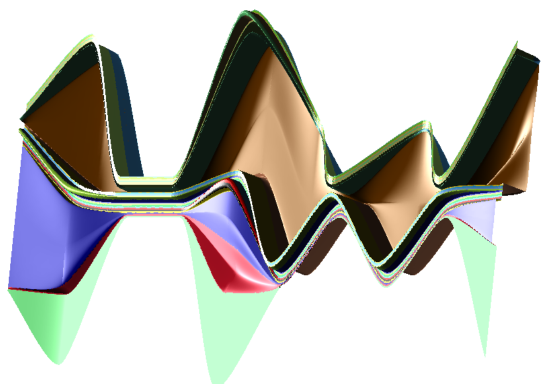

Figure 3: A 3D ThemeRiver visualization of 17 organic clusters. Width encodes overall cluster distributions (the magnitide of each cluster) and the height encodes incidence of zinc.

4 Comparison of 3D ThemeRiver with other approaches

As an experiment we compared the performance of our 3D ThemeRiver approach with a modified 2D version that also attempts to incorporate a second variable. Basically, we wanted to study if we can actually gain from the 3D extension, or if a modified 2D version would have performed just as well. In the 2D version, we assigned each stream a constant hue and saturation, but varied stream brightness in the same way we raised and lowered the landscape in the 3D version. The results of this experiment are shown in Fig. 4. There, 12 clusters are mapped across time, with width being mapped to their overall distribution and with height or brightness tracking the incidence of iron. The modified 2D version (Fig. 4a) highlights well the regions abundant with iron, however it looses visual separation of streams in zones lacking particles with this element. The stream distinctions are slightly improved when brightness ranges are clipped between.25 and 1 (see Fig. 4b). Nonetheless, this modification compromises the strength of highlights. The 3D ThemeRiver preserves colors and reflects the changes of iron occurrence on the z-axis. Interactive navigation of this scene is able to accentuate the depth diversity even more. In this respect a fully 3D extension to ThemeRiver appears superior to these other approaches.

5 Navigation

3D Navigation greatly enhances the visual understanding. Our 3D ThemeRiver can be rotated, translated, scaled and box-zoomed

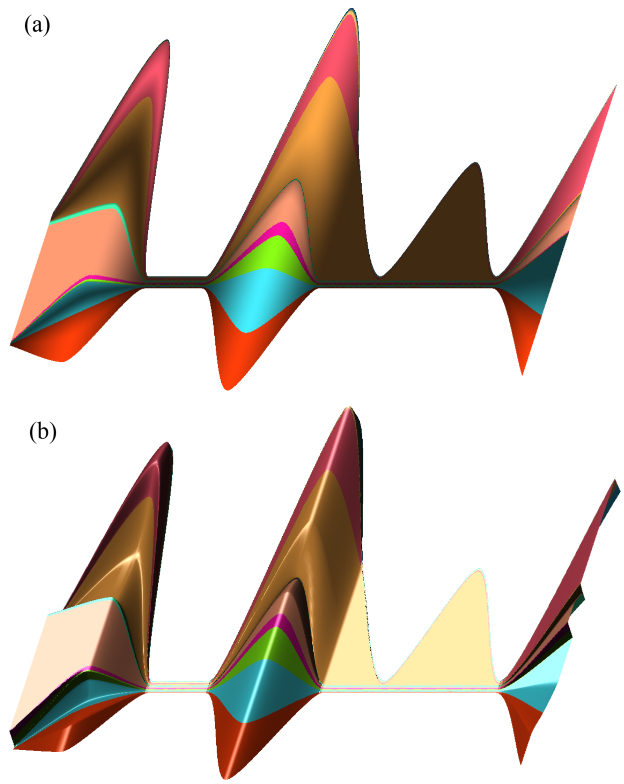

Figure 4: Comparison of the 3D approach to a 2D HSV approach: (a) 2D HSV ThemeRiver, (b) 3D ThemeRiver.

all in real time, facilitated by commodity graphics hardware. In addition, the user has also the flexibility to move the light source around in the scene, to emphasize different geometric aspects of the flow. Shadows add additional depth cues. Fig. 3 shows a navigable 3D ThemeRiver visualization. There 17 organic streams depict their overall distribution (the width) and occurrence of zinc, (height of streams).

6 Future Work

There are several potential areas for further research and improvements of this prototype tool. We would like to investigate ways to enrich 3D ThemeRiver to visualize more attributes per stream. One way would be provide a CD player-like interface to animate over the set of attributes. Another, quite intriguing, strategy would be to employ the concept of spectral volume rendering [2] to provide a set of "metameric lamps" to be used for highlighting different combinations of stream attributes on the fly.

Acknowledgments

We thank the Center for Data Intensive Computing (CDIC) at Brookhaven National Lab for their generous support of part of this work.

References

[1] S. Havre, E. Hetzler, P. Whitney, and L. Nowell, "ThemeRiver: Visualizing Thematic Changes in Large Document Collections," *IEEE Trans. Visualization and Computer Graphics*, vol. 8, no. 1, pp. 9-20, 2002.

[2] S. Bergner, T. Möller, M. Drew, G. Finlayson, "Interactive spectral volume rendering," *IEEE Visualization 2002*, pp. 101-108, 2002.