Dual Multiresolution HyperSlice for Multivariate Data Visualization

Pak Chung Wong *http://www.cs.unh.edu/∼pcw*

Andrew H. Crabb *http://www.cs.unh.edu/∼ahc*

R. Daniel Bergeron *http://www.cs.unh.edu/∼rdb*

Department of Computer Science, University of New Hampshire, Durham, NH 03824, USA

Abstract

We present a new multiresolution visualization design which allows a user to control the physical data resolution as well as the logical display resolution of multivariate data. A system prototype is described which uses the HyperSlice representation. The notion of space projection in multivariate data is introduced. This process is coupled with wavelets to form a powerful tool for very large data visualization.

1 Introduction

HyperSlice [3] is designed for visualization of multidimensional scalar functions. The main strength of the HyperSlice representation is its interactive environment in which only part of the data is displayed while the rest can be accessed via direct manipulation. We present an enhanced HyperSlice with a progressive refinement environment for data visualization. This environment allows a user to visualize massive amounts of data at a coarse resolution to identify areas that warrant investigation at finer resolutions. The coarse resolutions provide an accurate visual representation of the finer resolutions.

2 HyperSlice

HyperSlice is an extension of the scatterplot matrix which displays multivariate data as a set of bivariate plots. It defines a *point of interest* $c = \{c_i : i \in \mathbb{Z}\}$ which is the center of display, and *ranges* of display $r_i = \{(l_i, u_i) : i \in \mathbb{Z}\}$ which are the lower and upper limits of each dimension. Function values which fall out of the range (l_i, u_i) can be reached iteratively by panning across each display tile. A function of two spheres in the HyperSlice representation is depicted in Figure 1. We define the lower left tile of the HyperSlice as the origin which has the matrix coordinates of $(0, 0)$. So the tile above it has coordinates of $(1, 0)$, and the blank one to the right has $(0, 1)$. The cross hair lines shown in the three upper left tiles indicate the locations of the cross-section displays in the diagonal tiles (i.e., $\{(n, n) : n \in \mathbb{Z}\}$).

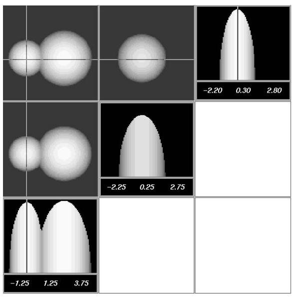

Figure 1: A HyperSlice plot.

3 Dual Multiresolution Exploration

Two approaches are applied in our design to reduce the size of the data and create a fine to coarse data hierarchy — *space projection* and *wavelet transform*. Our system combines these two processes and provides a dual multiresolution visualization environment to improve the browsing capability as well as the data navigation of the original HyperSlice.

3.1 Display Resolution Through Space Projection

Space projections are based on $\{\mathbb{R}^m \to \mathbb{R}^n : m > n\}$. Our definition of *projection* is stronger and more powerful than the similar term which simply describes the *view point projections* of different dimensions in the scatterplot matrix and prosection matrix [2] techniques. In our design, data from higher rank spaces are projected by different *norms* into data of lower rank spaces. By applying the projection to the data repeatedly, we generate a data hierarchy with multiple *display* resolutions.

3.2 Data Resolution Through Wavelets

Wavelets are based on *translation* $(W(t) \to W(t-1))$ and *dilation* $(W(t) \to W(2t))$. We use orthogonal wavelets because they provide non-redundant information. The ideal choice of wavelets is data and application dependent [1].

Without loss of generality, we describe a wavelet as a filter matrix that accepts a data stream with n items, and generates $n/2$ items of *approximations* and $n/2$ items of *details*. The approximation is a coarse summary of the original data, and the details contain the data loss during the decomposition. A hierarchy of coarse approximations is generated when this process is applied iteratively to the approximations to obtain increasingly coarse data.

3.3 Display Resolution versus Data Resolution

A major drawback of orthogonal wavelets is that the reduction rate is fixed at 50% in each dimension. Conventional aggregate functions, however, can generate more flexible resolutions. Data exploration using wavelet approximations count (both objectively and subjectively) on the shapes and the trends of the coarse approximations being preserved. For data mining, conventional aggregate functions are more natural and easier to understand and manipulate. We do not change the physical contents of the data during space projections, only the visualization. Wavelets, on the other hand, physically replace the data with smoother values.

Our system puts these two mechanisms into one powerful data visualization tool. As depicted in Figure 2, a data hierarchy with multiple data resolutions is first generated by wavelet decompositions. For data mining, space projections are applied to the data of the selected data resolution and data with multiple display resolutions are generated. Wavelet approximations provide a coarse view of a large dataset, which may not otherwise be able to be displayed on screen. Once the interesting patterns are located from this resolution of the data, the user can go to a higher data resolution to do the data mining using space projections.

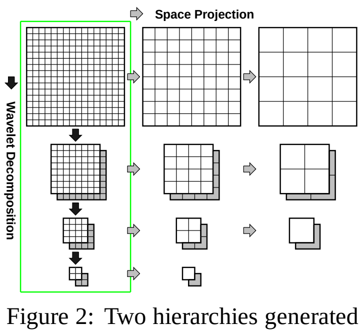

Figure 2: Two hierarchies generated by space projections and wavelets.

4 Visualization of Data and Error

Nearly all matrix-based visualization representations, including HyperSlice, duplicate mirror images (and/or movements) of the upper left half to the lower right half of the matrix. We, however, believe that this precious space (almost one half of the display area) can be used to provide another dimension of information. In additional to all the standard features to define a HyperSlice representation, our system also provides the *error* information generated by the wavelet decompositions. The approach of using wavelet details as a mean of *data authenticity analysis* is discussed in [4].

We present an application using a publicly accessible dataset[1] containing information about faculty at U.S. universities. We have selected six variates from this dataset – the number of faculty at each faculty rank and the average salaries at each rank. The data is displayed in Figure 3 and Color Plate 1a with 1024 (32^2) pixel blocks

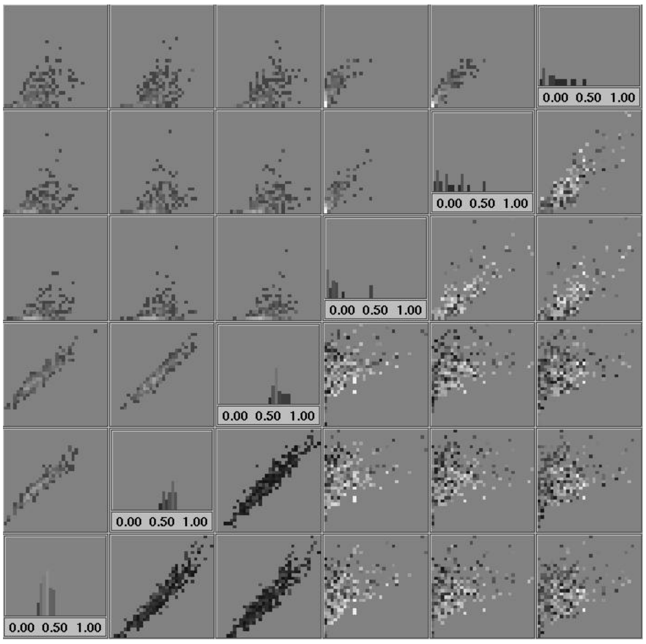

Figure 3: A fine projection of a coarse approximation.

used in each display tile. The display data, which is the second approximation generated from the wavelet transforms, represents well over 36K data values. The points of interest and the ranges of the display of the tiles are indicated in the corresponding diagonal tiles.

We notice that the overall pixel intensity (see also Color Plate 1c) of the $(0,1)$, $(0,2)$, and $(1,2)$ tiles generated by the *average* norm in Figure 3 are much lower than the rest of the tiles in the lower right half of the matrix. This indicates that the approximations depicted in the $(1,0)$, $(2,0)$, and $(2,1)$ tiles (which are the pairwise scatterplots of the salary figures of full, associate, and assistant professors) are more accurate representations than the rest of the approximations in the upper left half of the matrix. Suppose we want to see the distribution of the error in order to show that all (i.e., not just some of) the error values in the mentioned tiles are small. We decrease the display resolution and apply the *maximum* norm space projection. As we can see, the pixel intensity of the $(0,1)$, $(0,2)$, and $(1,2)$ tiles stays almost the same in Figure 4 and Color Plate 1b. This shows that *all* the error data values of the three tiles are indeed very low.

[1] http://lib.stat.cmu.edu/datasets/colleges/aaup.data

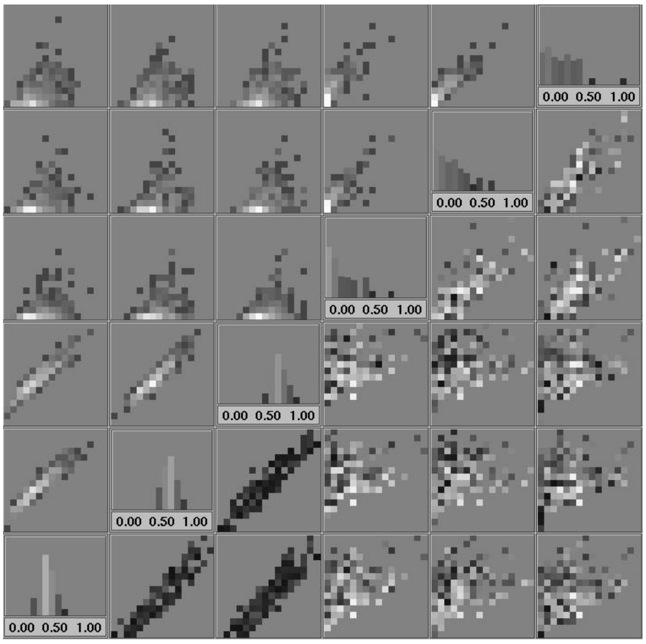

Figure 4: A coarse projection of a coarse approximation.

5 Conclusion

We present a system prototype of an enhanced HyperSlice which supports multivariate data visualization. The concept of *display resolution* supported by *space projection* is introduced. This notion is coupled with the concept of *data resolution* provided by *wavelets* to form a powerful multiresolution visualization system. This paper is part of our on-going efforts on very large data visualization using wavelets [4, 5].

Acknowledgements

This work has been supported in part by the National Science Foundation under grant IRI-9117153. The authors wish to thank Jarke van Wijk and Robert van Liere, who offered helpful discussions and support of many forms.

References

[1] Ingrid Daubechies. *Ten Lectures on Wavelets*. SIAM, Philadelphia, Pennsylvania, 1992.

[2] Bob Spence, Lisa Tweedie, Huw Dawkes, and Hua Su. Visualization for functional design. In *Proceedings of IEEE Information Visualization '95*, pages 4–9, Los Alamitos, California, Oct 1995. IEEE Computer Society Press.

[3] Jarke J. van Wijk and Robert van Liere. HyperSlice. In *Proceedings IEEE Visualization '93*, pages 119–125, Los Alamitos, California, October 1993. IEEE Computer Society Press.

[4] Pak Chung Wong and R. Daniel Bergeron. Authenticity analysis of wavelet approximations in visualization. In *Proceedings of IEEE Visualization '95*, pages 184–191, Los Alamitos, California, October 1995. IEEE Computer Society Press.

[5] Pak Chung Wong and R. Daniel Bergeron. Multiresolution multidimensional wavelet brushing. In *Proceedings of IEEE Visualization '96*, October 1996.