XmdvTool: Visual Interactive Data Exploration and Trend Discovery of High-dimensional Data Sets[*]

Elke A. Rundensteiner, Matthew O. Ward, Jing Yang and Punit R. Doshi
Department of Computer Science, Worcester Polytechnic Institute, Worcester, MA 01609
{rundenst|matt|yangjing|punitd}@cs.wpi.edu

1. VISUAL DATA EXPLORATION

XmdvTool is a public-domain software, initiated by M. O. Ward in 1994, is being developed at WPI for the interactive visual exploration of multi-variate data sets [4, 1]. XmdvTool supports an active process of discovery of patterns, trends and outliers in large-scale high-dimensional data sets. The major hurdles we overcome are the problems of display clutter, intuitive navigation and efficient access for visual exploration over large data sets.

XmdvTool incorporates several distinct display methods for multivariate data visualization that allow the users to view data from different perspectives [4]. The displays are tightly linked, such that visual interactions via one display can be seamlessly refined via other displays. XmdvTool supports a variety of advanced visual interaction tools, including brushing in screen space, data space, and structure space, panning, zooming and distortion [1].

Visual Exploration Scale-Up. Conventional multivariate visualization techniques do not scale well with respect to the number of objects in the data set, resulting in a display with an unacceptable level of clutter. To overcome this limitation, XmdvTool adopts a hierarchical approach by providing a suite of navigation and filtering tools to facilitate the viewing of the data at the desired level of detail [2] allowing systematic discovery of data trends.

Efficient Data Access. Efficient database transcriptions of operations defined in the visualization context, such as visual hierarchical drill-down and roll-up, are critical to allow for the near real-time behavior required for interactive tools. Organizing the hierarchical structure into a MinMax tree [3] reduces recursive hierarchy navigation into range queries. MinMax has allowed us to achieve performance levels required for interactive visualization even when connecting to large persistent data sets on Oracle.

High-Dimensionality Reduction. When visualizing data sets with a large number of dimensions, existing multi-dimensional visualization techniques become cluttered and thus ineffective. We thus reduce the dimensionality of the data while maintaining the relationships between data points using a *visual hierarchical dimension reduction* approach. We construct hierarchical dimension cluster trees based on clustering the dimensions, instead of the data points. Thereafter, we construct low dimensional data spaces guided by user interaction of the hierarchical dimension cluster tree.

Prefetching-Driven Caching. Exploration via visual interaction tools typically results in predictable traversal patterns of the data sets and thus effectively continuous query refinements. Hence customized caching and prefetching techniques have been shown to be effective in our tool [3]. In particular, we employ semantic caching principles to maintain relevant results of the previously executed queries indexed by their query specifications in the local client buffer.

Our *speculative prefetcher* brings data into memory when the system is idle. The prefetcher is based on the property of exploratory systems that queries remains "local". Xmdv Tool applies a variety of prefetching techniques based on both the analysis of current user interactions as well as the archived user session history.

XmdvTool Implementation. XmdvTool 5.0 is implemented in C/C++ with TclTk and OpenGL primitives. Interaction to Oracle8i is written in Pro*C/C++ embedded SQL primitives. The XmdvTool Home Page at *http://davis.wpi.edu/~xmdv* provides downloads of yearly releases of our software.

2. REFERENCES

[1] Y. Fua, M. Ward, and E. Rundensteiner. Structure-based brushes: A mechanism for navigating hierarchically organized data and information spaces. *IEEE Viz. and Computer Graphics*, pp.150-159, 2000.

[2] Y. Fua, M. Ward, and E. Rundensteiner. Hierarchical parallel coordinates for exploration of large datasets. *IEEE Proc. of Visualization*, pp.43–50, 1999.

[3] I. Stroe, E. Rundensteiner, and M. Ward. Scalable visual hierarchy exploration. In *Database and Expert Systems Applications*, pp.784–793, 2000.

[4] M. Ward, J. Yang, and E. Rundensteiner. Hierarchical exploration of large multivariate data sets. *Proceedings Dagstuhl '00: Scientific Visualization*, 2001.