

Visualization Databases: Lossless Visualization of Large and Complex Data

Ryan P. Hafen, Saptarshi Guha, William S. Cleveland
Purdue University, RVAC COE

Project Scope: Large and complex datasets are ubiquitous in all fields and are the standard for homeland security. The common approach to analyzing such data is to resort to summary analysis. While summary analysis is essential, it is almost always not enough to successfully exploit the important information in the data. With homeland security data, we are looking for the needle in the haystack, and to do so we need to be analyzing the data at a level of detail that does not sacrifice information. We call this type of analysis *lossless analysis*. Achieving lossless analysis of large and complex data is immensely challenging and requires everything that enters in the analysis of data to be rethought.

Lossless analysis cannot be achieved without visualization throughout the process, from initial data cleaning to presentation of final results. Successful visualization of large and complex datasets requires visualization databases: a very large number of graphical displays, each with a very large number of pages. A single display can cover possibly hundreds of square feet. With a visualization database (VDB), an analyst constrained to small physical monitor space can explore many large displays by querying and viewing the pages of a display on an as needed basis.

Making VDBs successful in practice requires research in all areas of data visualization, tailoring current methods to VDBs and inventing wholly new ones.

Recent Progress: We have successfully employed visualization databases as part of three homeland security projects: the Indiana state syndromic disease surveillance program in collaboration with the Indiana State Department of Health (ISDH), disease surveillance in collaboration with the 30,000-patient Purdue University Student Health Clinic, and a network security project that involves the analysis of billions of network flows. We were awarded an NSF grant for analysis and visualization of large and complex datasets.

Future Plans: For VDBs we are working on a manager, a viewer, techniques that maximize use of physical screen space, new display methods, computational algorithms, hardware environments, and automation algorithms that make rendering decisions for the user.

Relevance to listed research areas: This research is relevant to the Advanced Data Analysis and Visualization research area. It is also related to any area of homeland security that requires the analysis of large and complex datasets.

Publications: Draft paper in progress for Journal of Computational and Graphical Statistics.