VisTrails for Interactive Multiple-View Visualizations

Nathan Smith and Juliana Freire
Scientific Computing & Imaging Institute and School of Computing

Scientists are now faced with an incredible volume of data to analyze. To analyze and validate various hypotheses, it is necessary to create insightful visualizations of both the simulated processes and observed phenomena. This is far from being an interactive process and contains many error-prone and time-consuming tasks. Today, the generation and maintenance of visualization data products is a major bottleneck in the scientific process. VisTrails is a visualization system which streamlines the creation, execution and sharing of complex visualizations.

For example, CORIE is an environmental observation and forecasting system (EOFS) for the Columbia River. The goal of this multi-decade project is to predict complex circulation and mining processes in a system encompassing the lower river, the estuary, and the near-ocean. Real-time forecasts and simulations lead to tens of thousands of visualization products. Generating these visualizations is both error-prone and time-consuming. In particular, it makes comparing the results of different simulations, time-steps, or rendering algorithms difficult.

VisTrails manages the data, metadata and the exploration process, while scientists can focus on science. By capturing the provenance of both the visualization process and data, scientists manipulate, VisTrails enables reproducibility and simplifies the complex problem of creating and maintaining visualizations. The scientists can explore their visualizations by returning to previous versions of a workflow (or visualization pipeline); apply workflows instances to different data, explore the parameter space of the workflow; query the visualization history; and comparatively visualize different results.

The Radiation Physics Division of the Department of Radiation Oncology at the Massachusetts General Hospital uses advanced visualization algorithms to facilitate the diagnosis process and to locate pathological tissue in preparation for radiation therapy treatments. The visualizations process as currently deployed is very complex and time-consuming. Whereas a scanner can create a new data set in minutes, with advanced tools it takes from several hours to days to create appropriate visualizations. In addition, it is hard to reproduce many of the manual steps and a large number of files need to be maintained. This greatly limits the ability to explore the large volumes of data, hinder collaboration, and hampers the useful lifetime of data.

VisTrails streamlines the process of data exploration through visualization by keeping detailed provenance of both visualization workflows and associated data. The system allows a user to return to previous versions of a workflow; query the visualization history, annotate workflows and images; and visually compare workflows. In essence, VisTrails replaces the laboratory notebook. Because VisTrails unobtrusively captures detailed provenance information and also manages visualization data, it removes the need for manually creating and maintaining large directory structures of resulting datasets and images as well as the need for writing detailed notes.

This research is supported in part by the National Science Foundation (BS-0513692).