Graph-Based 3D Flow Field Visual Analysis

Pak Chung Wong (Technical Lead), Pacific Northwest National Laboratory
Han-Wei Shen, The Ohio State University
L. Ruby Leung, Daniel Chavarria, Pacific Northwest National Laboratory

Researchers at the Pacific Northwest National Laboratory (PNNL) and The Ohio State University are developing a viable visual analysis approach to prepare our scientists to address the exabyte 3D flow field data analysis problems in the very near future. To achieve this extreme size data analysis goal, we address *at least* four major scalability issues: *data* scalability, *algorithm* scalability, *visual* scalability, and *display* scalability in exabyte visual analytics.

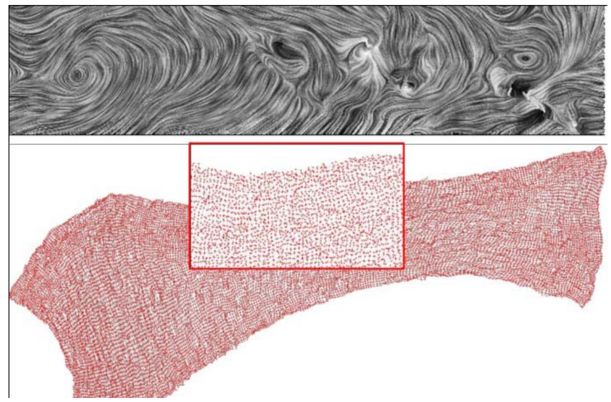

Figure 1: A climate flow field visualization and the corresponding graph visualization.

The research features a visual analytics pipeline that adaptively transfers an exabyte 3D flow field to a petabyte to terabyte graph structure that contains a network of salient information of the flow field, to a visualization with millions to tens of millions of data items (not bytes) that characterize the graph structures, and finally to a high-resolution display with over 15 million pixels for the scientist to interact with the data. Figure 1 shows an example of a climate flow field visualization and the corresponding graph visualization. All four scalability issues will be addressed together or separately along the proposed visual analysis pipeline. The ultimate goal is to bring the exabyte flow field data down to a manageable size and format so that scientists can use it to conduct analysis and experiments.

We collaborate with U.S. Department of Energy (DOE) domain scientists and apply the new extreme scale visual analytic pipeline approach to study 3D flow fields found in regional climate and combustion simulations. Much of the computation of this project is performed on Franklin at the National Energy Research Scientific Computing Center (NERSC). This three-year project has been funded by DOE Office of Science (SC) since August 2010.

Scope

The scope of this project includes:

- Designing the visual analytics pipeline concept for extreme scale scientific flow field analysis.
- Developing the visual analytics pipeline prototype on one of the DOE Leadership Computing Facility (LCF) computers.
- Collaborating with domain scientists and applying the visual analytics pipeline prototype to study extreme scale flow field simulation data.
- Deploying the extreme scale data analytics technology to DOE domain scientists.

Objectives

Flow field datasets represent one of the largest and most complex scientific data types facing DOE scientists today. Very large 3D flow field data can be found in many DOE SC modeling research areas such as climate, combustion, and high-energy physics. As suggested by a DOE workshop report on Challenges in Climate Change Science in 2008, “climate model data are growing faster than the data set size for any other scientific discipline, with collections of hundreds of exabytes expected by 2020….” The main objective of this project is to identify and deliver a viable visual analysis approach to prepare our scientists to address the exabyte 3D flow field data analysis problems in the very near future.

Our solution to the extreme scale data challenge is both innovative and non-traditional. It combines proven visualization techniques with newly established visual analytics approaches to address an extreme scale scientific data analytics problem. Figure 2 illustrates the concept of an extreme scale visual analytics pipeline for scientific data analysis. Given an exabyte (10^{18}) flow field dataset in Figure 2a, we first transform it to a graph-based representation known as FlowWeb in Figure 2b. A FlowWeb is a directed node-link graph that highlights the essential flow structures of a flow field dataset. Although the size of a FlowWeb varies, we want to see a FlowWeb graph in the petabyte (10^{15}) to terabyte (10^{12}) range, preferably the latter. As a graph of this size is still too big and complicated for direct visual interaction, we further extract the local features of individual graph nodes (or set of connecting nodes) of a FlowWeb and summarize these features as numerical *signature vectors* that characterize the FlowWeb graph nodes in Figure 2c. These signature vectors are then visualized (as a scatterplot in this example) in Figure 2d. We want to see a scatterplot that shows tens of millions (10^7) to millions (10^6) of graph vectors (not bytes) at this stage. Each graph vector will contain *salient* information that represents a portion of the original flow field data for analysis purposes, together with an index tag that allows scientists to trace back to the full resolution data.

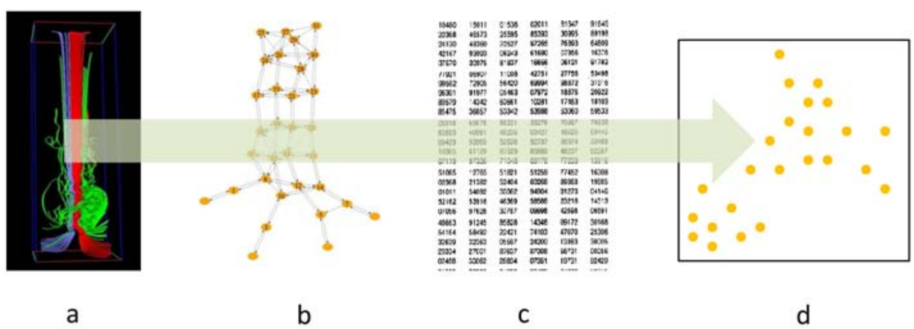

Figure 2 a) A 3D flow field. b) A FlowWeb of the flow field. c) A set of graph signature vectors generated from the FlowWeb. d) A mock sketch of a scatterplot visualization of the signature vectors.

Current Activities

In the first year of our research, we use output from climate simulations such as the National Center for Atmospheric Research (NCAR) Community Climate System Model (CCSM) and Weather Research and Forecasting (WRF) model to test our exabyte data analytics theories. The two models are running on Jaguar at Oak Ridge National Laboratory (ORNL).

Our initial challenge is to use our extreme scale visual analytics pipeline to identify a propagating feature known as the Madden Julian Oscillation (MJO). Most of the existing climate models that run at coarse resolution rely on convective parameterizations and they cannot capture MJO. Climate researchers will use our system to better understand what processes control MJO and its time scale, etc. Our investigation, if successful, will be the first effort to visually analyze MJO in regional climate studies.

Cutting-edge flow field to flow graph techniques are currently being developed. New flow graph signature designs customized for climate studies are being implemented. All our computation work is being done on Franklin at NERSC. It is our plan to perform testing at scale on Jaguar in the out years.

For more information, please contact Pak Chung Wong pak.wong@pnl.gov. (PNNL-SA-77951)