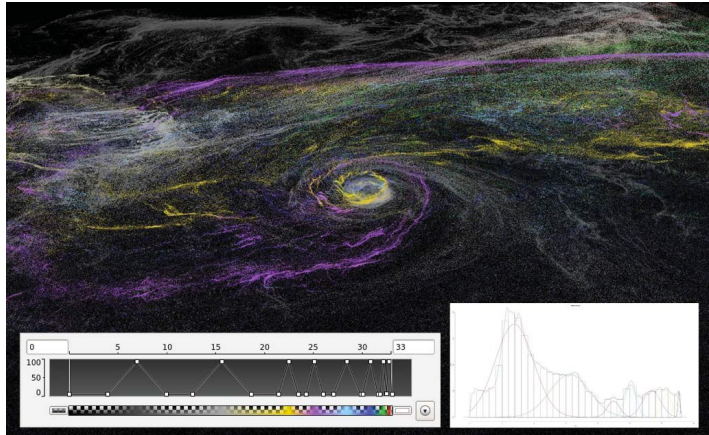


Machine Learning and Data Mining from Large-Scale Volumetric Data

Project Description

Large-scale simulations result in enormous amounts of data [2]. Visualization or even just transfer of the complete simulation data is a time consuming and tedious task. Extracting the essential part of the data is usually done after the entire simulation is finished. Since the storage capacity of the simulation servers is limited, not all data can be permanently



stored and is only available during the simulation itself. Domain scientists are typically scanning the stored results looking for specific features in the data. The essential data covers just a tiny amount of the entire data space. Our approach attempts to detect features concurrently to the simulation. Due to temporal as well as spatial data coherence, similar patterns can be detected and stored in dictionaries. In the optimal case the original data can be reconstructed by a linear combination of a small amount of dictionary entries [1].

Your Role in this Project

You will research and implement prototypes of methods to support domain scientist in detection of important features and characteristics in large simulation data. Since most of the simulation results are transient data, this should be performed in real-time concurrently to the run of the simulation.

Requirements

You need to have:

- prior experience in C++ or MATLAB programming
- fluent English language skills
- openness for a multicultural environment

References

[1] https://en.wikipedia.org/wiki/Sparse_dictionary_learning

[2] T. Theussl et al., Simulation and visualization of the cyclonic storm chapala over the arabian sea: a case study, KACSTIT, 2016

