Advanced Exploration of Large-Scale Scientific Image Databases

Banesh, Divya[1,5] Barnes, David[2] Klaassen, Sebastian[3] Shaikh, Uzma[4]

Mentors: Jim Ahrens[5], David Rogers[5], Li-Ta Lo[5], Chris Sewell[5]

[1] University of California, Davis
[2] Massachusetts Institute of Technology
[3] University of Vienna
[4] Purdue University
[5] LANL

Abstract

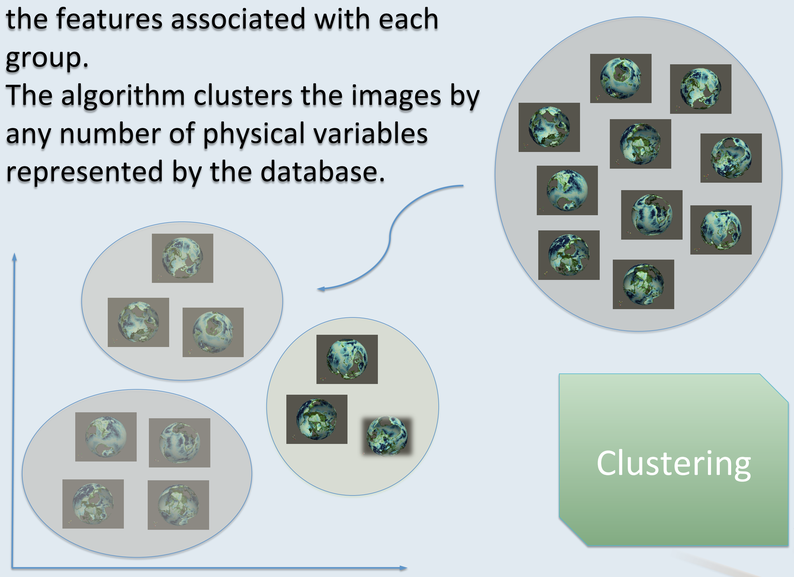
Large-scale scientific simulations generate massive amounts of data, which can be a bottleneck for scientific discovery. "Cinema", recently developed at Los Alamos National Laboratory and Kitware, Inc., achieves extreme scale in situ visualization by adopting an image-based approach for storing large-scale simulation data. Our application provides an interface to the Cinema database, allowing the scientist to view, cluster, query, and analyze images by leveraging the associated metadata. To get an overview of the complete space of images, the user is provided with a global view. The global view shows all images in a virtual 3D environment, where the alignment of images is defined by customizable context-dependent transformations of input dimensions such as time or viewpoint. A graphical interface allows the user to browse the Cinema database using metadata search as well as content-based image search. With metadata search, a scientist can look for images by querying the properties associated with the images. Additionally, with content-based search, given an image or part of image as input from the user, the program will return all images from the database that have similar features. The program will also provide existing trends or patterns that hold true over all images returned. Another way of grouping the images in the Cinema database based on physical variables such as temperature and pressure is the k-means clustering algorithm. This approach allows the scientist to recognize the underlying groups produced by the data set and the features associated with each group. The algorithm will be able to cluster the images on any number of physical variables represented by the database, as specified by the scientist. A web interface to the Cinema database completes the workflow by allowing the scientist to browse through the image database using the different techniques.

Background & Motivation

- Large-scale scientific simulations generate massive amounts of data, which can be a bottleneck for scientific discovery.
- Scientific simulations have evolved to extreme scale, making it increasingly difficult to perform post-processing due to limited data movement speeds and space.
- It is critical for scientists to be able to interact with the data produced so as to obtain meaningful insights.
- "Cinema" achieves extreme scale in situ visualization by adopting an image-based approach for storing large-scale simulation data.
- Our application provides an interface to the Cinema database, allowing the scientist to view, cluster, query, and analyze images by leveraging the associated metadata.

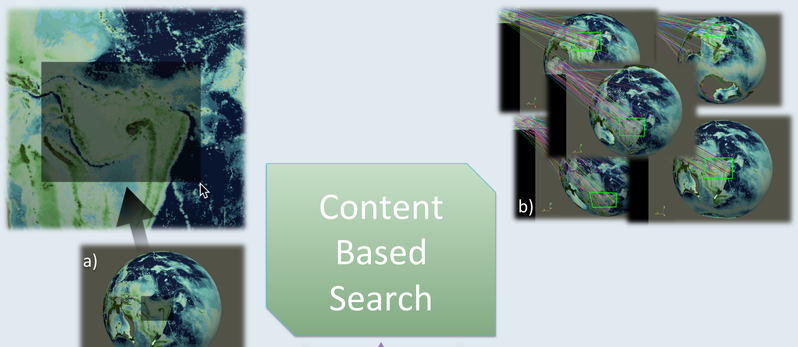
Clustering

- Images are grouped based on physical variables such as temperature and pressure using the k-means clustering algorithm.
- Clustering of the database allows the scientist to recognize the underlying groups produced by the data set and the features associated with each group.
- The algorithm clusters the images by any number of physical variables represented by the database.



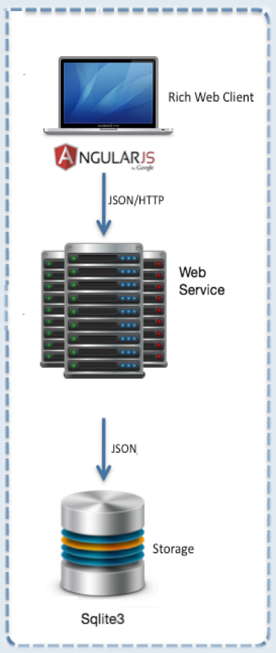
Content Based Search

- The Content-Based Image Retrieval algorithm returns images of interest using similarity search.
- The user selects an initial 'query image' from the Cinema database and marks a section of the image that contains the region of interest (a).
- The program will then return a set of images that contain the features initially selected with the area of interest delineated (b).
- Using this algorithm and given the series of time step-data available in the Cinema database, the user can track trends in lifetimes of features of interest.

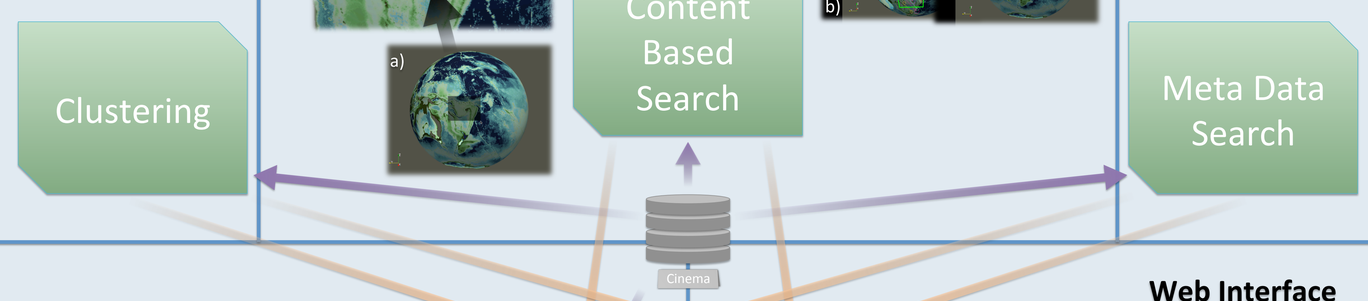


Meta Data Search

- Our framework provides a mechanism allowing the scientists to query the data by leveraging the metadata associated with the images.
- The scientist can look at the images with common features for the purpose of data analysis.
- The web interface is composed of
  - A front-end built with AngularJS
  - A back-end: a Restful web service using a SQLite database.

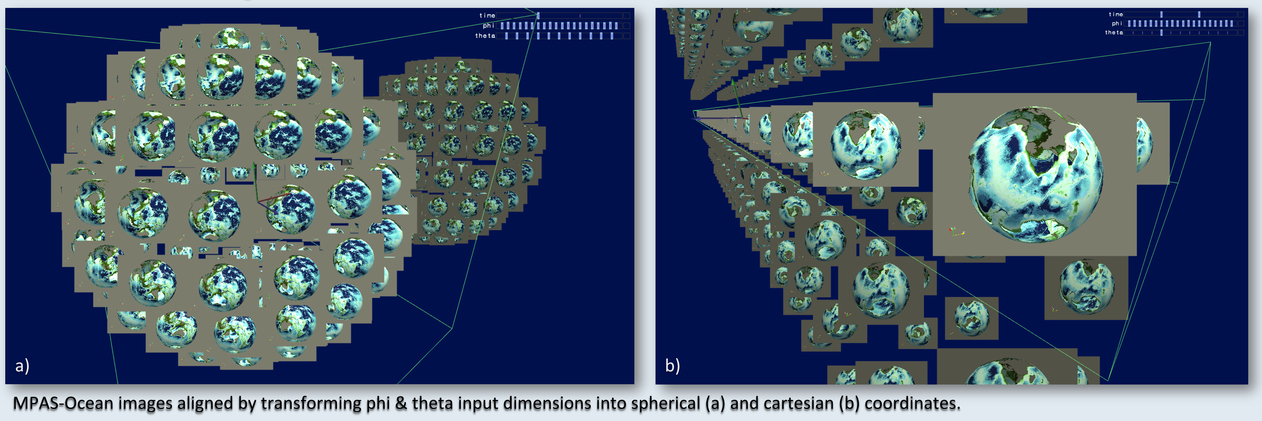


Components of the image search engine



Global View

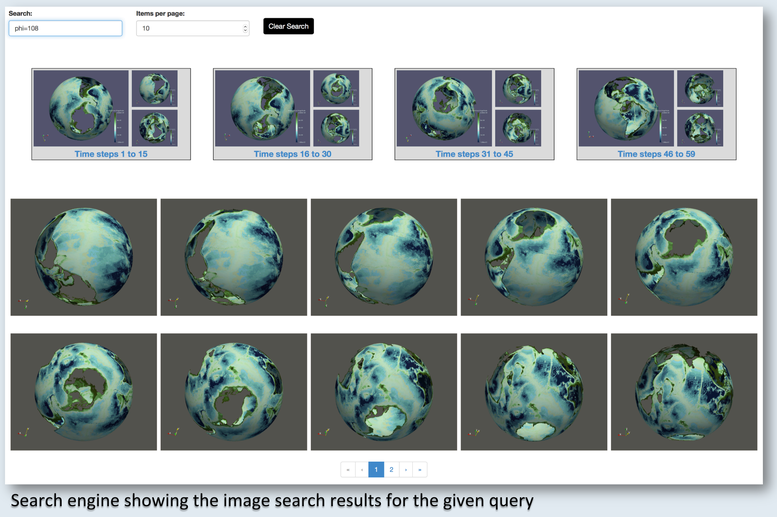
- The global view provides the scientist with an overview of all images in the database.
- By flying through the virtual image space and trying out different image alignments, the database can be explored in an intuitive way. Images area aligned by customizable context-dependent transformations of input dimensions such as time or viewpoint.
- Results of image classification and reduction operations can directly affect how images are viewed.



MPAS-Ocean images aligned by transforming phi & theta input dimensions into spherical (a) and cartesian (b) coordinates.

Web Interface

Scientist



Search engine showing the image search results for the given query

ACKNOWLEDGEMENTS:

We would like to thank Francesca Samsel, Jonathan Woodring and Arnold Eatmon.

Los Alamos NATIONAL LABORATORY EST. 1943