GLIDER: Mining and Analysis Tool based on Service Mash-ups

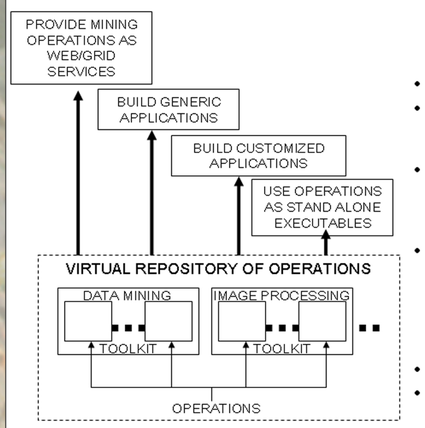
Sara Graves[1], Sundar Christopher[1], Rahul Ramachandran[1], Todd Berendes[1], Patrick Hogan[2]
[1]University of Alabama Huntsville
[2]NASA/JPL

1. Project Summary

NASA's strategic objectives state the need to improve the scientific and technological capabilities of the nation. Today's researchers and students have a difficult time fully exploiting the large volumes of NASA satellite imagery data, in part due to lack of tools that can provide the full spectrum of visually browsing remote data, analyzing a part or all of the imagery and extracting thematic information using advanced mining capabilities. To address this problem, *a NASA ACCESS project will integrate NASA World Wind, Interactive Visualizer and the Image Classifier for Satellites (IVICS) and the Algorithm Development and Mining (ADaM) toolkit using service mash-ups.* By leveraging existing mining and image processing services, the proposed integration will be both effective and efficient. The goal of this project is to integrate World Wind, IVICS and ADaM into a single seamless end user application (**GLIDER tool**), modifying the existing components as needed and integrating them via service mash-ups. The tool will be evaluated in research and academic settings using specific science case studies and student projects, focusing on improved visualization and classification of aerosols in combined MODIS/CALIPSO data. The team will refine and customize the tool based on feedback, and make it widely available for diverse applications with a wide variety of NASA imagery. World Wind, IVICS and ADaM are well established, mature tools with distinct, complementary capabilities. Synergistically integrating these individual tools will result in a complete, comprehensive, easy-to-use end user application for visualization, analysis and thematic information extraction from NASA imagery.

2. GLIDER Components

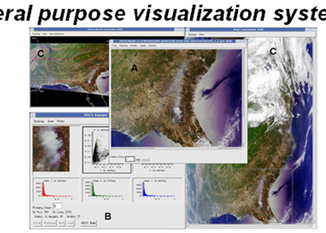
2.1 ADaM: Algorithm Development and Mining



- ADaM was originally developed in response to a NASA Research Announcement with the goal of mining large scientific data sets for geophysical phenomena detection and feature extraction
- It has evolved & expanded over time
- ***ADaM toolkit contains over 100+ different mining and image processing algorithms***
- ADaM components are lightweight and autonomous, and have been used successfully in a grid and web services environment
- ADaM has several translation components that provide data level interoperability with other mining systems (such as WEKA and Orange), and point tools (such as libSVM and svmLight)
- ADaM toolkit is available via the web
- Its used by many different institutions world wide

2.2 IVICS: Interactive Visualizer and Image Classifier for Satellites

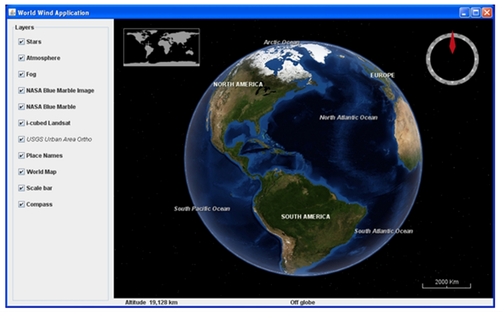
- Development of IVICS was initially driven by the requirements of the Earth Observing System (EOS) Clouds and the Earth's Radiant Energy System (CERES) to create a polar cloud mask using a neural network based classifier
- Supervised classification techniques require labeled training samples thus ***detailed visual examination capability***
- IVICS was developed to provide this functionality
- Evolved over time to a ***general purpose visualization system*** for satellite (and other) data



IVICS sample session with MODIS imagery. (A) IVICS main window (B) IVICS Sample editor for the selected region (C) IVICS map display window (D) Window displaying the entire satellite swath as a subsampled image.

2.3 World Wind

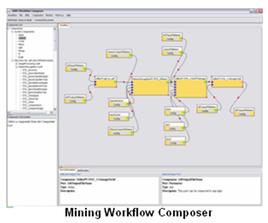
- World wind is a ***data browser/visualization toolkit***
- World Wind lets you easily visualize distributed satellite imagery and products
- World Wind lets you zoom from satellite altitude into any place on Earth.
- World Wind lets you explore data in a visually rich 3D, just as if you were really there.



2.4 Mining Web Services

As part of a recent ACCESS project, UAH has teamed with the Goddard Earth Sciences Data and Information Services Center (GES DISC) to develop and demonstrate the use of distributed mining web service for science data analysis solutions. Many science data repositories are now providing online access to vast amounts of data. This project is demonstrating an approach for composing mining workflows of distributed web services to provide reusable analysis solutions for these online data resources. These services allow users to locally define analysis workflows that can be executed on data residing in online repositories. This project, called **Mining Web Services** (MWS) as a short name, enables ADaM toolkit capabilities for use in a distributed web service environment. This project has repackaged ADaM components as SOAP-based web services for easier and more dynamic integration in emerging distributed service oriented architectures. This suite of mining and related services has been packaged into deployable bundles to allow easy deployment in the service-oriented architectures at online data provider and other sites.

To solve complex science problems generally requires multi-step solution or workflow. A workflow describes how tasks are orchestrated, what component performs them, what their relative order is, how they are synchronized, how information flows to support the tasks and how tasks are being tracked. Currently, the industry standard for service orchestration is the Business Process Execution Language (BPEL). BPEL provides a standard XML schema for workflow composition of web services that are based on SOAP. This standardized composition description can be deployed on any BPEL engines which in turn parses the BPEL description and invokes the services in the appropriate sequence. To assist users in composing the workflows, UAH has adapted a graphical composition tool to work in this environment.



Mining Workflow Composer

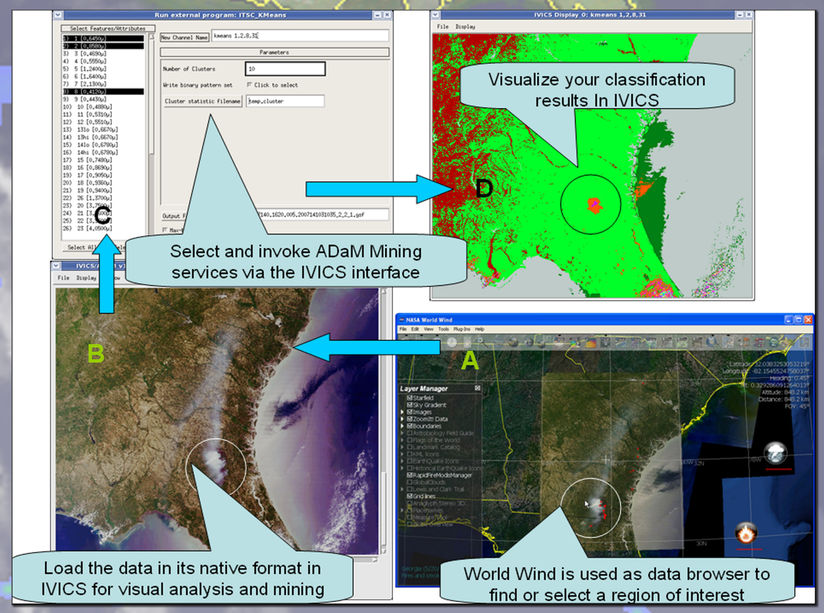
3. Earth Science Drivers

The GLIDER project will use two science scenarios as drivers using combined MODIS/CALIPSO data, both of which are of paramount importance to the aerosol-cloud-climate community. In each case interactive GLIDER tool will be used to improve classification.

- Enhanced visualization of Dust/Smoke/Clouds over Northern Africa to improve classification and study of aerosol direct and indirect radiative forcing using MODIS and CALIPSO data.
- Improved visualization and classification of biomass burning smoke and clouds over Africa for aerosol-cloud studies. World Wind already has interfaces for visualizing fires from the MODIS that will aid in our analysis.

4. Envisioned use of GLIDER

- GLIDER will allow scientists to
  - A. Browse and visualize distributed data from multiple perspectives
  - B. Pull in data in its native format for detail visual analysis
  - C. Perform complex image processing and mining (thematic extraction) on the data in its native format
  - D. View mining results and compare the results against other observations



5. Broader GLIDER Applicability

- Even though the science scenarios that will be used as drivers for this project are focused on Atmospheric science, the tool can be used in multiple domains in Earth science. There are numerous domains in Earth science that use remotely sensed imagery from different platforms to measure and observe both natural and manmade phenomena and processes. These domains all require tools to extract thematic information from the imagery – such as land-use change detection for urban planning; detection of weather hazards such as cyclones, convective storms; ecological impact of catastrophic natural events such as Hurricane Katrina; etc. The GLIDER tool will provide cross-cutting benefits across these diverse domains and applications in Earth science. Furthermore, in addition to NASA and academia, other federal agencies such as DHS, FEMA, and NOAA could potentially use this tool.