Visualization of Large Geospatial Data Sets

Gordon Standart, Kelsey Stulken, Manuel Penaloza, and Ziliang Zong*

Department of Mathematics and Computer Science
South Dakota School of Mines, Rapid City, SD 57701

Abstract

The Earth Resources Observation and Science (EROS) Center of U.S. Geological Survey EROS is currently managing and maintaining the world largest satellite images distribution system, which provides 24/7 free download service for researchers all over the globe in many areas such as Geology, Hydrology, Climate Modeling, and Earth Sciences. A large amount of geospatial data contained in satellite images maintained by EROS is generated every day. However, this data is not well utilized due to the lack of efficient data visualization tools. Therefore, it will be beneficial to design a geo-visualization tool which allows EROS staff and global users to visualize the properties of download requests for the satellite images maintained on EROS web servers.

This paper describes a method for visualizing various characteristics of the satellite image download requests. More specifically, Keyhole Markup Language (KML) files are generated which can be loaded into an earth browser such as Google Earth. Colored rectangles associated with stored satellite scenes are painted onto the earth browser; and the color and opacity of each rectangle is varied as a function of the popularity of the corresponding satellite image. An analysis of the geospatial information obtained relative to specified time constraints provides an ability to relate image download requests to environmental, political, and social events.

* Corresponding author. Please direct questions concerning this article to Dr. Ziliang Zong at Ziliang.Zong@sdsmt.edu.
More detailed information about this project can be found at *http://www.mcs.sdsmt.edu/zzong/*.

1. Introduction

The "Massive Global Spatial Data Visualization project" is implemented to visualize the massive geospatial data provided by the Earth Resources Observation and Science (EROS) Center of U.S. Geological Survey (USGS). A large amount of spatial data contained in satellite images maintained by EROS is generated every day; however, this data is not well utilized by currently available data visualization tools. Therefore, it will be beneficial to design a geo-visualization tool which allows EROS staff to visualize the properties of download requests for the satellite images maintained on EROS web servers. The project will also provide a web portal through which global users may search and download the geo-visualization results.

The fairly recent decision by EROS to place all of their satellite imagery in the public domain has placed a significant load on existing resources. A massive number of satellite data download requests have subsequently been processed, and the volume of these requests is expected to increase steadily.

It is critical for EROS members to have the capability of quickly and easily evaluating relevant geographic information for all of the download requests made within specified dates. The purpose of this project is to visualize the nature of the download requests based on various geographic properties.

The nature of the proposed change is to create geo-visualization software that will radically improve the quality of information that can be discerned from the data contained in the log files. The results of these visualization efforts will be easily accessible via a web portal.

2. Visualization Method

Every download request for an EROS satellite image is stored in a log (text) file. The initial set of log files provided by EROS spanned approximately fifteen months – October 2008 through December 2009. Each log file entry contains a unique product ID along with the Landsat scene ID, row, path, acquisition date, and request date for a download request. Each and every one of these entries was parsed into the individual items of interest and inserted into a database table. This table currently contains over 544,000 records.

This table of download request records is supplemented by two additional tables: 1. a table of points which converts each row/path combination into the corresponding latitude and longitude of the center of a satellite image/scene, and 2. a table of large geographic areas (Europe, Asia, Africa, etc) and the set of row/path combinations defining a rectangle which approximates the area.

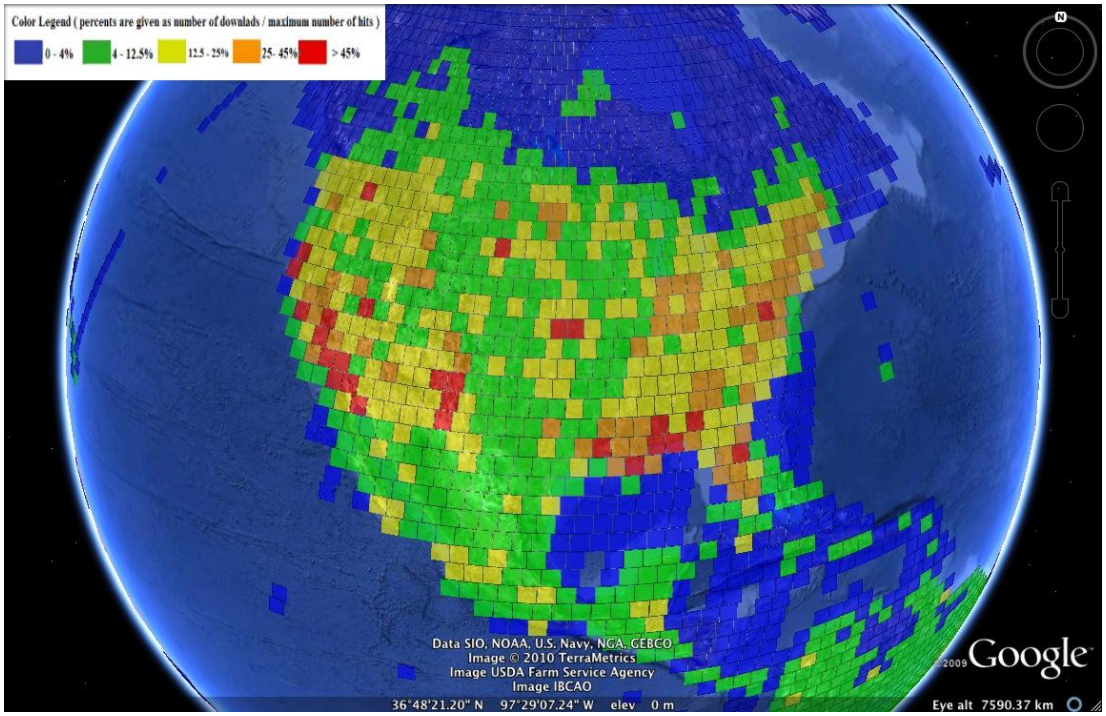
With this data in place, a user can request a visualization of a subset of the data set (or the entire data set) via a web portal. The portal will send the user's query to the database, and a KML file with the requested visualization will be generated for subsequent download. The user may then load the KML file into an earth browser (e.g., Google Earth) and perform a visual analysis of the applicable data set.

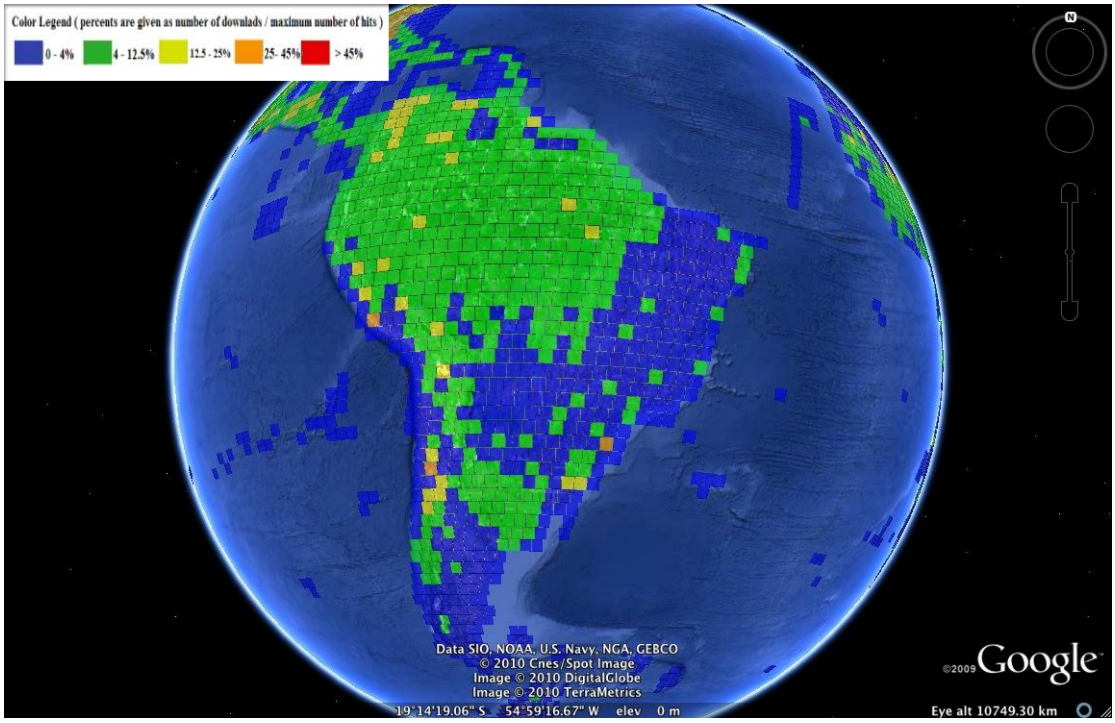
Key features of the KML schema are exploited to provide easy visualization in Google Earth. Each satellite image is embedded within the KML file with <Placemark> tags; and each placemark contains a colored rectangle with dimensions approximating the area of the image and a color corresponding to the popularity (number of download requests) of the scene. The disparity between the dimensions of each rectangle and the actual geometry of each stored image is negligible and actually desirable as it provides a more useful visualization of the download request data. Numerical data providing the row, path, latitude, longitude, and number of downloads are embedded within each placemark with <BalloonStyle> tags which enable the user to activate a data "balloon" over a rectangle of interest.

The nature of the color range at the heart of the visualization proved to be an elusive objective due to the highly subjective nature of human color perception. An EROS paper (Landsat Project Status – January 2009) provided a useful precedent that was subsequently implemented in the latest visualization software.

3. Visualization Results

Images from Google Earth displaying sample KML files generated by the visualization software are provided below.





4. Conclusions and Future Applications

The visualization tool provides a useful method to analyze massive amounts of data stored in a text file. The software and database design allow the end users to continue to utilize the tool with the large amounts of download request data which will continue to grow (possibly exponentially) as the availability of satellite imagery in the public domain becomes more popular. The ranges and types of colors used in the visualization will be an ongoing design issue, and better implementations will inevitably be discovered through experimentation and discovery of similar methods in other research efforts.

Another foreseeable application of this visualization tool will be to support data mining efforts related to EROS satellite images. The South Dakota School of Mines is currently pursuing research to provide predictive capabilities which will assist EROS personnel in storing/staging massive amounts of satellite imagery files in an effective manner. This visualization tool may ultimately provide the primary method of verifying the rules of association that these data mining efforts would provide.

5. Acknowledgements

We gratefully acknowledge the supports from the Earth Resources Observation and Science (EROS) Center of U.S. Geological Survey, the U.S. National Science Foundation under Grants No. CNS- 0915762 and the Nelson Research Grant under Grants No. 03988.