A Hierarchical Data Archiving and Processing System to Generate Custom Tailored Products From AVHRR Data

S. N. V. Kalluri[†], Z. Zhang[*], J. Jájá[*], D. A. Bader[✦], H. Song[†], N. El Saleous[†], E. Vermote[†], J. R. G. Townshend[†*],
Email: sk71@umail.umd.edu, Fax: (301) 314 9658, Phone: (301) 405 6733
[†] Department of Geography, University of Maryland, College Park, MD-20742, USA
[*] University of Maryland Institute for Advanced Computer Studies (UMIACS)
[✦]Department of Electrical and Computer Engineering, University of New Mexico

ABSTRACT

A novel indexing scheme is described here to catalogue satellite data on a pixel basis. The objective of this research is to develop an efficient methodology to archive, retrieve and process satellite data, so that data products can be generated to meet the specific needs of individual scientists. When requesting data, users can specify the spatial and temporal resolution, geographic projection, choice of atmospheric correction, and the data selection methodology. The data processing is done in two stages. Satellite data is calibrated, navigated and quality flags are appended in the initial processing. This processed data is then indexed and stored. Secondary processing such as atmospheric correction and projection are done after a user requests the data to create custom made products. By dividing the processing in to two stages saves time, since the basic processing tasks such as navigation and calibration which are common to all requests are not repeated when different users request satellite data. The indexing scheme described here can be extended to allow fusion of data sets from different sensors.

INTRODUCTION

Global Area Coverage (GAC) data from the Advanced Very High Resolution Radiometer (AVHRR) has become an important source of information for monitoring and characterizing the dynamics of the biosphere, atmosphere and the oceans at a global scale. This coarse resolution data set is the only data available globally on a daily basis over a time period of more than 16 years. Some typical applications of AVHRR data include land cover characterization (e.g. [1]), deriving variables to estimate surface energy balance (e.g. [2]), detecting forest fires (e.g. [3]) and monitoring sea surface temperature (e.g. [4]). Several global data sets derived from the AVHRR instrument have been produced to study land cover dynamics since 1981. These include several versions of the Global Vegetation Index (GVI) products, the continental NDVI data set produced by the Global Inventory Monitoring and Modeling Studies (GIMMS) group at NASA's Goddard Space Flight Center (GSFC), the Pathfinder AVHRR Land (PAL) data set, and the 1 km. Global land data produced by the Earth Resources Observation Systems (EROS) Data Center (EDC) [5]. Although these data sets have found widespread use in the earth system science community, they have several inherent limitations. Some of the limitations include availability of data in a fixed geographic projection, spatial and temporal resolution, with little or no capability for generating subsets based on user requirements. The compositing method and time interval are also static for all the previously mentioned AVHRR data sets. Some of these AVHRR data sets also have atmospheric correction applied to them. However, users who are interested in retrieving atmospheric properties from uncorrected satellite measurements and those users who want to experiment with new atmospheric correction algorithms may prefer to obtain uncorrected satellite data.

Because of the limitations of these data sets, the full potential capabilities of the AVHRR instrument cannot be exploited for special applications such as monitoring global Net Primary Production (NPP) [6]. Moreover, the requirements of users differ depending upon their specific application and usage of the data sets, and these requirements are expected to change as they gain experience in the use of these data sets and as their needs evolve [5]. Thus, there is an important need to design a processing system that can generate AVHRR data sets following the specifications of individual users. In this paper, we discuss the design and development of a hierarchical data archiving and retrieval system that allows the researchers to request and generate AVHRR data products according their needs by specifying: region of interest, map projection, spatial resolution, temporal resolution, land/ocean data, type of atmospheric correction, cloud screening, and compositing function.

SYSTEM DESIGN

Our processing system is called *Kronos* and consists of four major components: ingest and preprocessing, indexing and storage, search and processing engine, and a Java interface (Fig.1).

Ingest and Preprocessing

In this first step, AVHRR level 1B data is unpacked, calibrated, navigated, and quality and cloud flags are determined for each pixel using algorithms developed by [7]

0-7803-5207-6/99/$10.00 © 1999 IEEE.

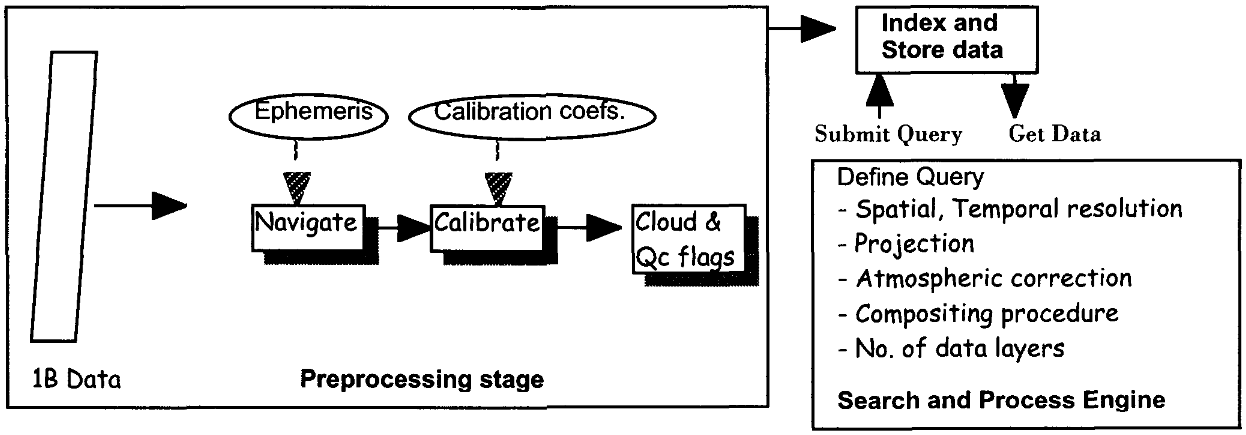

Figure 1. A schematic diagram of the kronos processing system

under NASA's Pathfinder Data Set Product Generation Algorithm (NASA NRA-94-MTPE-06). A brief outline of the preprocessing algorithms that were adapted in *Kronos* from the work by [7] is discussed here.

After unpacking the level 1B data, all the pixels are navigated to determine accurate geolocation using up to date ephemeris data and the orbital model of [8]. The visible and NIR bands are calibrated using the coefficients given by [9]. Brightness temperatures in bands 3 to 5 are calculated using the internal calibration targets. Zenith and azimuth angles for the view and illumination directions are computed for each pixel. Cloud condition is determined for each pixel using the methods described by [10]. Thus, for each Instantaneous Field of View (IFOV), several geophysical values are computed. These values are then indexed and stored.

Indexing and Storage

The objective of our indexing and storage scheme is two fold: to store the preprocessed GAC pixels without either any loss of information or resampling, and to efficiently retrieve the indexed data using a search engine. Storing and indexing multi-temporal satellite data is an important area of research which is being addressed by the scientific computing community [11]. Our data indexing scheme is novel, and can be briefly as follows. The processed data is indexed using a grid file combined with a *k-d* tree data structure. We first create a two dimensional grid with a spatial resolution of $1^0 \times 1^0$. Each $1^0 \times 1^0$ cell can be considered as a "bucket" containing all the IFOV's that have been navigated into that region. Within each bucket all the pixels are indexed using an optimal *k-d* tree spatial data structure. The *k-d* tree data structure is a hierarchical spatial data structure in which the *k* dimensional space is recursively divided [12]. In our case, *k* represents a two dimensional space denoted by latitudes and longitudes. Thus, the $1^0 \times 1^0$ space is recursively divided in to finer regions hierarchically, based on latitudes and longitudes, and a binary search tree is constructed. Note that this indexing ensures that all the navigated IFOVs are preserved absolutely with no loss in information.

The *k-d* tree for our case can be described as follows. The first branch within the $1^0 \times 1^0$ tree represents the median latitude of all the pixels within the cell. All the pixels within the $1^0 \times 1^0$ cell are divided into two regions by the median latitude. Within each region further division is made by the median longitude of the pixels. Thus, the latitude and the longitude are used alternatively to decompose the entire space. Finally, a global index is created which can be used to retrieve the pixels for a given Cartesian coordinate. A detailed explanation of the *k-d* data structure can be found in [13]. The *k-d* tree architecture is better for representing point data compared to other methods such as quadtrees [14]. After indexing, the data is stored on a disk array managed by IBM's High Performance Storage System (HPSS).

Search and Process Engine

A Java interface accepts queries from the users, who can specify the spatial and temporal resolution of the data along with the number of bands they want. The output images can be generated in 31 different projections, and the users can request either the calibrated data without any atmospheric correction or have Rayleigh, ozone, water vapor and stratospheric aerosol corrections applied to the reflective bands following the procedures described by [7].

The function of the search and process engine is to fetch the appropriate data from archive and apply the required processing tasks. First, an empty two dimensional grid is created, where each grid cell has a specific spatial resolution depending upon the geographic projection and spatial resolution specified by the user. Using the latitude and longitude coordinates, the *k-d* tree index is searched, and all the nearest IFOVs which can be potentially binned in to each output grid cell are retrieved. Atmospheric correction is done to each IFOV if specified by the user, and a single value is chosen to fill the grid cell following the compositing criteria defined by the user.

IMPLEMENTATION

The *Kronos* system is written in C and has been bench marked on an IBM 160 MHz power2 processor with 512 MB RAM connected to an eight way disk array with a storage capacity of 72 GB. To ingest one days GAC data for the entire globe including land and oceans, it takes 66 minutes. Indexing this data is done in 14 minutes. After the data is indexed, a global image file consisting of five calibrated bands, the Normalized Difference Vegetation Index, view and solar geometry, cloud and QC flags can be generated in 14 minutes in Goodes Homoloscine projection at 8km. spatial resolution containing both land and ocean data.

DISCUSSION

Storing satellite data in a hierarchical data structure has several distinct advantages compared to storing the data as individual two dimensional images. Since all the geophysical values for each IFOVs are indexed along with precise navigated earth location information, there is no loss of information caused by resampling. For generating custom tailored AVHRR products quickly and efficiently, the methodology implemented here can be considered as a two fold approach:

- Initial processing, which is common to all queries (i.e. unpack, navigate, and calibrate); This processing is done only once to the GAC data, and then indexed and stored in the archive.
- Secondary processing, which is unique to specific queries (i.e. choice of compositing, atmospheric correction, and projection). This is done on the fly in a real time mode following user specifications.

By dividing the processing into these two components, we can save a lot of time since the data does not have to be navigated and calibrated for each request. At the same time, the data is not tied to a specific projection or correction scheme. Although *Kronos* has been designed to index and archive AVHRR data, this system can be extended to process and store geo-spatial raster data from other satellite sensors as well. We are currently working on implementing the system in a parallel computing environment. The methodological innovations from this research allow the fusion of data from multiple sensors for earth system science.

REFERENCES

[1] J. R. G. Townshend, C. O. Justice, and V. T. Kalb, "Characterization and classification of South American land cover types using satellite data," *International Journal of Remote Sensing*, vol. 8, pp. 1189-1207, 1987.

[2] B. Seguin and B. Itier, "Using midday surface temperature to estimate daily evaporation from satellite thermal IR data," *International Journal of Remote Sensing*, vol. 4, pp. 371-383, 1983.

[3] M. Matson and B. N. Holben, "Satellite detection of tropical burning in Brazil," *International Journal of Remote Sensing*, vol. 8, pp. 509-516, 1987.

[4] L. M. McMillin, "Estimation of sea surface temperatures from two infrared window measurements with different absorption," *Journal of Geophysical Research*, vol. 80, pp. 5113-5117, 1975.

[5] J. R. G. Townshend, "Global data sets for land applications from the Advanced Very High Resolution Radiometer: An introduction," *International Journal of Remote Sensing*, vol. 15, pp. 3319-3332, 1994.

[6] S. D. Prince and S. N. Goward, "Evaluation of the NOAA/NASA Pathfinder AVHRR Land Data Set for global primary production," *International Journal of Remote Sensing*, vol. 17, pp. 217-221, 1995.

[7] N. El Saleous, E. F. Vermote, C. O. Justice, J. R. G. Townshend, J. Tucker, and S. N. Goward, "Improvements in the global biospheric record from the Advanced Very High Resolution Radiometer (AVHRR)," *International Journal of Remote Sensing*, vol. Accepted for publication, 1999.

[8] G. Rosborough, D. Baldwin, and W. Emery, "Precise AVHRR image navigation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, pp. 644-657, 1994.

[9] E. Vermote and Y. J. Kaufman, "Absolute calibration of AVHRR visible and near infrared channels using ocean and cloud views," *International Journal of Remote Sensing*, vol. 16, pp. 2317-2340, 1995.

[10] L. L. Stowe, E. P. McClain, R. Carey, P. Pellegrino, G. G. Gutman, P. Davis, C. Long, and S. Hart, "Global distribution of cloud cover derived from NOAA/AVHRR operational satellite data," *Advances in Space Research*, vol. 3, pp. 51-54, 1991.

[11] R. E. Wolfe, D. P. Roy, and E. Vermote, "MODIS land data storage, gridding, and compositing methodology: Level 2 grid," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, pp. 1342-1338, 1998.

[12] H. Samet, *Application of Spatial Data Structures*. Reading: Addison-Wesley, 1990.

[13] J. L. Bentley, "Multidimensional binary search trees for associative searching," *Communications of the ACM*, vol. 18, pp. 509-517, 1975.

[14] H. Samet, *The Design and Analysis of Spatial Data Structure*. New York: Addison Wesley, 1989.