The Study and Implementation of Extraction HY-1B Level 1B Product Image Data Based on HDF Format

Shibin Liu [a], Wei Liu [ab], Hailong Peng [c]
[a]Center for Earth Observation and Digital Earth, Chinese Academy of Sciences, Beijing, China;
[b]Graduate University of Chinese Academy of Sciences, Beijing, China;
[c]National Satellite Ocean Application Service, Beijing, China

ABSTRACT

The Goal for development of Extraction HY-1B Level 1B (L1B) Product Image Data Procedure is to enhance the convenience of retrieving product data for users. The images extracted from L1B data are used as a quick look images for Product Archive and Distribution System of HY-1B (PADS) [1] and can be presented with a map while consumers search for data via Internet. And the procedure is integrated with Product Archive System (PAS) [1], which is mainly responsible for archiving and managing HY-1B product data. This paper mainly focuses on the implementation of this procedure as well as the technical problems encountered and resolved.

Keywords: extraction of Image, HDF4, SDS, geometric correction, thin-plate spline

Corresponding Author:
PhD; Wei Liu;
Center for Earth Observation and Digital Earth, Chinese Academy of Science (CEODE);
No.45, Bei San Huan Xi Road, Haidian District, Beijing, China, 100086;
phone: 0086 01081106983
email: liuweixyr@126.com;

1. INTRODUCTION

HY-1B satellite, which followed HY-1A satellite, is hosted by National Satellite Ocean Application Service of China. The oceanographic satellite is equipped with a 10-band ocean color scanner (COCTS), a 4-band CCD imager with 250-meter resolution (CZI), and an infrared water profile radiometer. Its major responsibilities are exploration of chlorophyll, suspended sediment, dissolved organic matter and sea surface temperature and other factors and monitoring for the dynamic changes of coastal zones.

HY-1B level 1B (L1B) Product data using HDF file format are through a series of processing from level 0 data, such as global positioning, band registration, radiometric calibration, and so on. In order to distribute the L1B Product data by Internet and enhance convenience for user to know about the data content instantly, PADS needs to extract product image data as quick look images from the HY-1B level 1B Product data based on HDF Format. Due to the fact that the demand was later than the development of PAS, the integration problem was laid before developers.

Hence, this paper first illustrates features of L1B data format--HDF and characteristics of data themselves and then comprehensively describes the key techniques and the work flow of the procedure to extract image. The procedural logic includes band data extraction, latitude and longitude data extraction, and geometric correction with thin-plate spline transformation. In the end, the paper discusses the design and consideration on how to integrate Extraction procedure with Product Archive System appropriately.

2. HY-1B L1B DATA

2.1 HDF File Format

HY-1B L1B data takes HDF (Hierarchical Data Format) 4.2 version as file format which was created at the National Center for Supercomputing Application in USA. HDF can let users store mixtures of data from different sources into a single file, such as image geophysical data, navigation data calibration data and so on [2].

The prerequisite to access HDF file is to understand the HDF software libraries. The HDF software libraries can be viewed as three interactive levels. The basic interface layer, or the low-level API, is reserved for software developers. It was designed for direct file I/O of data streams, error handling, memory management, and physical storage. The HDF application programming interfaces, or APIs, include several independent sets of routines, with each set specifically designed to simplify the process of storing and accessing one type of data. In most cases, all that one must do is to make the correct function call at the correct time, and the interface will take care of the rest. On the highest level, general applications, HDF includes various command-line utilities for managing and viewing HDF files, several research applications that support data visualization and analysis, and a variety of third-party applications [2].

This procedure just depends on the application programming interfaces to access the various types of HDF data structures which include raster image, palette, scientific data set, annotation, vdata, and vgroup. And HY-1B L1B data are constructed with scientific data set (SDS), which is a group of data structures used to store and describe multidimensional arrays of scientific data. Refer to Fig 1 for a graphical overview of the SD data set. Note that a scientific data set consists of required and optional components. SD API, as one of HDF application programming interfaces, provides routines that store, retrieve, and manipulate scientific data using SDS data model easily.

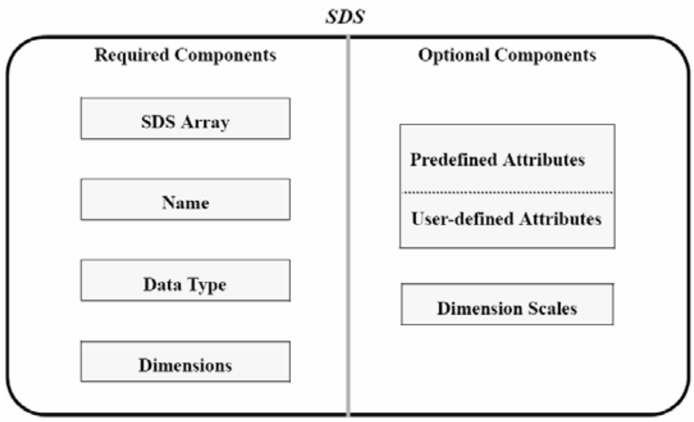


Figure 1. The Contents of a Scientific Data Set [2]

With this ability of SD API, the procedure could access SDS in the following manner:

1) Open a HDF file and initialize the SD interface.
2) Select an existing SDS using its index.
3) Read data from a data set, or retrieve information about a dimension, or read the values of an attribute specified by its index.
4) Terminate access to an SDS.
5) Terminate access to the SD interface and close the file.

2.2 The Characteristics of HY-1B L1B Data

For the moment, according to receiving and processing phases, the levels of HY-1B data are in turn termed L0, L1A, L1B, L2A, L2B, L2C and L3. L1B data, as the subsequent level of L1A, which applies radiometric correction to raw data, should be rectified in geometry in order to extract the image data with geographical projection for data distribution and data presentation based on map by Internet.

L1B data contain spectral bands data as well as corresponding longitude and latitude data. And the map that would be integrated with these image data is also described by longitude and latitude. Since the longitude and latitude data are related to the addresses of pixels in spectral bands data, it is possible to build up mapping functions between the spectral bands data and the longitude and latitude data, as will be described later. These longitude and latitude data can be referred to as GCPs that could locate a point in bands data knowing its position in the map. However, the rule of corresponding relationship between the two data sets is not one to one. To be specific, the bands data from COCTS with 1100m resolution contain 1664 pixels per scan line. Accordingly, every scan line contains 166 GCPs. The first GCP corresponds to the seventh pixel, and then, every 11 pixels in bands data will there be a control point pair. Similarly, the bands data from CZI with 250 m resolution contain 2048 pixels per scan line, and every scan line contains 102 GCPs. The first GCP is related to the fourteenth pixel, and then, every 21 pixels will there be a control point pair [3].

The L1B data from COCTS sensor explored by HDF Explorer are depicted in Fig 2. It shows that the bands data, longitude data, latitude data are all two dimensional. In addition, there are some key attributes, such as ‘Pixels Per Scan Line’, ‘Number of Scan Lines’, ‘Easternmost Longitude’, ‘Westernmost Longitude’, and so on.

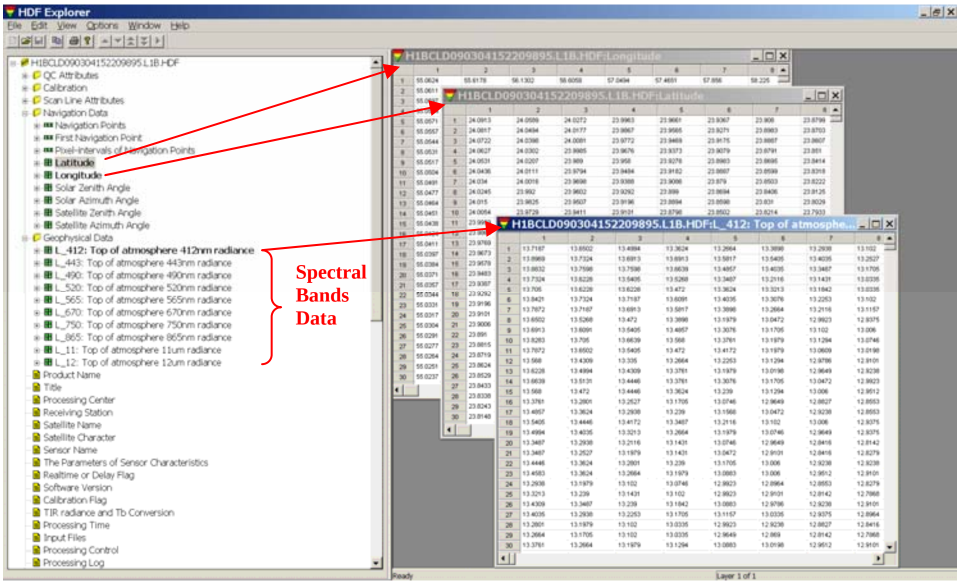


Figure 2. The L1B Data with HDF file format.

3. KEY TECHNIQUES AND IMPLEMENTATION FOR THE PROCEDURE

Fig.3 shows an activity diagram of the procedure. While the process is a straightforward one, there are some key techniques that are described in the following.

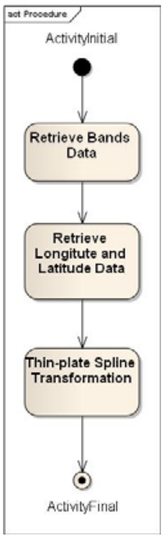


Figure 3. The activity diagram of the procedure.

3.1 Retrieve Bands Data

There are a series of spectral bands data in L1B data. However, considering the best effect of visual impact, extracting an image data as a quick look image just needs three bands data combined into a RGB false color image. Through a series of experiments, for COCTS, L_670 (Red Color), L_750 (Green Color), and L_490 bands (Blue Color) are selected to form a RGB false color image. Likewise, L_685, L_665, and L_443 bands are selected for CZI [3]. Moreover, because of the L1B band data with poor contrast, it is necessary to enhance the radiometric nature of the image to improve its visual impact.

The concrete steps to create images with original resolution from L1B HDF files are shown as follows.

(1) First, invoke SD API to open a L1B data file.

(2) Secondly, select the three SDSs of specific bands by their index.

(3) Thirdly, linear contrast modification, which is to expand the original brightness range to the range from 0 to 255.

(4) Finally, by Gdal library, add the three bands data into a dataset and export an image with original resolution.

3.2 Retrieve Longitude and Latitude Data

The process to access longitude and latitude data is similar to that for bands data. However, a L1B data file contains a huge number of longitude and latitude data that are GCPs. For instance, for COCTS, the number of scan lines is always more than 6000. So, the total number of GCPs is more than 166×6000. Thanks to the regular arrangement of GCPs, it is possible to construct a sampling matrix with 40 rows and 20 columns from the GCPs, which can reduce the number of GCPs involved in geometric correction [3]. The concrete process for constructing the sampling matrix is as follows. First, get the height and width of two-dimension arrays of longitude and latitude data. Then, divide the height by 40 and the width by 20, and then, round these two numbers to get *dy* and *dx* respectively. After that, by traversing the elements in the two-dimension arrays of longitude and latitude based on the steps of *dy* and *dx*, get the sampling matrix. Finally, add the sampling GCPs in the last row and the last column into the matrix if the matrix does not contain these sampling GCPs.

Nevertheless, for HY-1B satellite, frame loss occurs occasionally. In other words, some scan lines are missing due to some reasons. So, in order to make sure the accuracy of geometric correction for L1B data, it is indispensable to fill up these lost frames with black color rows. In L1B data, since one frame corresponds to four scan lines in bands data and 'Frame Number' SDS records the frame number of each line, so the procedure could traverse the SDS and compare the neighboring frame numbers and check whether the difference of the two frame numbers is more than 1. Once the difference is more than 1, it means that there exists lost frames and the number of the missed rows needs to be computed and filled with black color line at the right place in a new image, as is shown in Figure.4a. Then the procedure updates the addresses of pixels and related GCPs according to the new image. Besides, sometimes the swath width of HY-1B is able to span from eastern hemisphere to western hemisphere and the longitude values in western hemisphere are all defined as negative values, so that it is possible to lead the values of easternmost longitude that are less than the values of westernmost longitude. Care must be taken to avoid such problems. So if the difference between the easternmost longitude and the westernmost longitude is negative, it should add the negative value of longitude with 360.

3.3 Thin Plate Spline

With the list of GCPs which can also be related to the pixels in the new image, it is possible to build up explicit forms of mapping functions. There are a series of types of transformation generally applied in geometric correction, such as affine transformation, polynomial transformation, and radial basis functions.

Given a set of control points $\{w_i, i = 1,2,3...K\}$, a radial basis function basically defines a spatial mapping which maps any location x in space to a new location f(x), represented by

$$f(x) = \sum_{i=1}^{N} c_i \phi(\| x - w_i \|)$$

where ||.|| denotes the usual Euclidean norm and {ci} is a set of mapping coefficients. It is a real-valued function whose value depends only on the distance from the origin, or alternatively on the distance from some other points w_i, called a center. One possible choice for the kernel function φ is the thin plate spline,

$$\phi(r) = r^2 In(r)$$

Where $r = \| x - w_i \|$. The name "thin plate spline" (TPS) refers to a physical analogy involving the bending of a thin sheet of metal[3]. TPS is commonly used as the non-rigid transformation model in image alignment and shape matching. The popularity of TPS comes from a number of advantages:

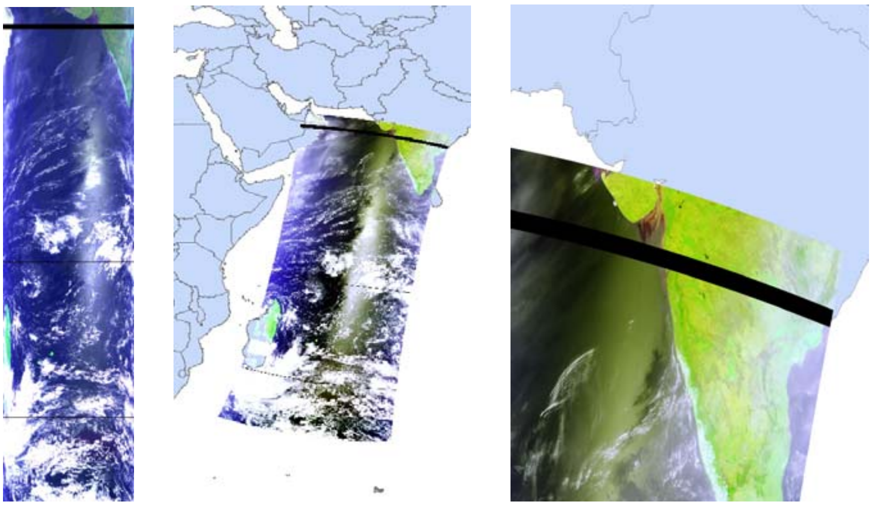
(1) It is the fundamental solution to the biharmonic equation, so the interpolation is smooth with derivatives of any order[4].

(2) A nice property of the TPS is that it can always be decomposed into a global affine and a local non-affine component. Consequently, the TPS smoothness term is solely dependent on the non-affine components. This is a desirable property, especially when compared to other splines, since the global pose parameters included in the affine transformation are not penalized [4].

(3) It has a more global nature than other kernel, so that a small perturbation of one of the control points would not affect the coefficients corresponding to all the other points [3].

With the GCPs and map projection information, the procedure uses the mapping functions above to find the corresponding pixel in the new image for each position located with longitude and latitude.

The procedure takes gdal_translate tool and gdal_warp tool to implement the warp steps. First, use gdal_translate to assign a coordinate system and GCPs. Since Longitude/latitude projection is in common use in oceanography, the procedure uses the projection with **WGS84** datum which is almost universally assumed to the datum used. Then, by gdal_warp to make TPS transformation, the image can be integrated with the map data in a Geographical Information System, or presented to consumers in a map-like form as shown in Figure 4b. Figure 4c depicts the local part of the image shown in Figure 4b.



a b c

Figure 4. The images produced from an experimental L1B COCTS data.

4. INTEGRATION WITH PAS

Since images extracted from L1B data would be presented to Internet, the extraction procedure should be completed during product archive processing. The procedure should focus on the product data that have been archived in product data storage area. In order to avoid unforeseen exceptions happening in the procedure that maybe lead inconsistency information in product data storage area and database, for instance, some intermediated data would remain, the procedure first scans the archived data to find out L1B data without a quick look image. Then, the procedure copies the L1B data to a temporary folder and creates a quick look image. Finally, the procedure copies the image back into the product data storage area.

5. SUMMARY AND FUTURE DIRECTION

Since the procedure merely uses the longitude and latitude data involved in L1B data, the accurate image registration mainly depends on the accuracy of device itself on the HY-1B satellite. Based on the experiments, the results are satisfactory to the requirements as quick look images integrated with a map by Internet. And the procedure also improves user experience and brings convenience for users to search their desired product data.

6. REFERENCES

[1] Liu Shibin, Liu Wei, Peng Hailong, "The Design and Implementation of the HY-1B Product Archive System", ISDE 6 (2009).
[2] HDF4 User's Guide (2010)
[3] Peng Hailong, Zou Bin, Liu Shibin, "Geometric correction for HY-1B data based on the thin-spline method" Marine Science Bulletin, Vol. 28 No.5 106-109 (2009).
[4] http://en.wikipedia.org/wiki/Thin_plate_spline

Shibin Liu;
Associate Director of Data Management Department;
Center for Earth Observation and Digital Earth, Chinese Academy of Science (CEODE);
No.45, Bei San Huan Xi Road, Haidian District, Beijing, China, 100086;
email: sbliu@ceode.ac.cn;

Research Interests:

- Data management
- Network Data Service System

Wei Liu;
PhD student;
Center for Earth Observation and Digital Earth, Chinese Academy of Science (CEODE);
Graduate University of Chinese Academy of Sciences;
No.45, Bei San Huan Xi Road, Haidian District, Beijing, China, 100086;
email: liuweixyr@126.com;

Research Interests:

- Data management
- GIS and remote sensing

Hailong Peng
Associate Researcher;
National Satellite Ocean Application Service, Beijing, China
8 Dahuisi Road, Haidian District, Beijing,GD 100081,China;
email: phl@mail.nsoas.gov.cn

Research Interests:

- Ocean remote sensing and Application of GIS
- Satellite Data archive and distribution