社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2006－MPS－60 (9)
2006／6／26

Design and Implementation of Gateau API for HDF-EOS

Kazuko Yamauchi Eriko Touma Hiroyasu Kamo
Naoyuki Nide Sachiko Hayashida Kazuki Joe
Graduate School of Humanities and Sciences, Nara Women's University

Abstract

Gateau, which we have been developing, is a system that helps atmospheric scientists analyze data by visualizing earth observation data used in atmospheric science research. In this investigation, we have selected HDF-EOS as a data format for Gateau. The HDF-EOS format recently increases among data distributed as earth observation data. We implement an HDF-EOS read function to Gateau. Atmospheric scientists can perform 3D visualization without writing programs using the original API called the HDF-EOS library. In this paper, we describe the implementation of the HDF-EOS read function to Gateau.

HDF-EOSのためのGateau APIの設計と実装

山内 和子　当麻 英梨子　鴨 浩靖　新出 尚之　林田 佐智子　城 和貴
奈良女子大学大学院人間文化研究科

概要

我々が開発しているGateauは，大気科学研究に用いられている衛星観測データを３次元可視化することで大気科学研究者がデータを分析することを支援するためのシステムである．本研究ではGateauに読み込ませる衛星観測データフォーマットとして，近年配布が増加しているHDF-EOSを選択した．大気科学研究者はHDF-EOS読み込み機能を使用することで，HDF-EOSライブラリと呼ばれる独自のAPIを用いたプログラムを記述することなく３次元可視化を行うことができる．本稿ではGateauへのHDF-EOS読み込み機能の組み込みについて述べる．

1 Introduction

Earth observation data and meteorological data used in atmospheric science research are widely distributed by NASA and other research institutes. The data are distributed in various formats such as HDF-EOS [1], NetCDF [2] and GRIB [3]. The formats of distributed data are not unified. However, recently the distribution of data in the HDF-EOS format increases among any data formats. For this reason, we paid attention to HDF-EOS to choose it as a data format for a 3D visualization system Gateau [4][5]. Gateau is a 3D visualization system that we are developing.

In HDF-EOS, a single file has a hierarchical structure like typical file systems. To read data from an HDF-EOS file or to write data to an HDF-EOS file, it is necessary to use the original API called HDF-EOS library. However, it seems to be difficult for users inexperienced in data processing with APIs to use HDF-EOS. Therefore, we implement HDF-EOS read function for Gateau users. When an HDF-EOS file is opened with Gateau, the hierarchical data structure is plainly displayed by the tree view. In addition, when the dataset that users want to visualize is selected by a mouse click, a 3D visualization is performed.

The second section describes the exposition of HDF-EOS and HDF [6]. HDF is the origin of the concept of HDF-EOS. The third section presents development of the HDF-EOS read function. The fourth section presents the implementation of HDF-EOS as Gateau APIs.

2 HDF and HDF-EOS

This section describes HDF and HDF-EOS. The satellite data format HDF-EOS was developed based on the data format called HDF.

2.1 HDF

HDF, which stands for Hierarchical Data Format, is a scientific and technological data format. HDF was developed at the National Center for Supercomputing Applications (NCSA) of UIUC. Users can put a number of datasets hierarchically into a single HDF file. Moreover, the information about data and the format itself are also put into an HDF file with data. At the present time, two versions of HDF, HDF Ver.4 and HDF Ver.5 (hereinafter referred to as HDF5) are available. Hereafter, we describe the HDF5.

A single HDF5 file can hierarchically contain attribute informations called Global Attribute, Group and Scientific Data Set (SDS). Every file starts with the root. Group is a structure for organizing objects in an HDF5 file. Groups and SDSs can be hierarchically involved in a Group. The observed data are stored in an SDS.

2.2 HDF-EOS

HDF-EOS was defined by a collaboration of NASA and NCSA to store the data from the Earth Observing System (EOS). HDF-EOS supports three kinds of data types: Grid, Point, and Swath. It is necessary to use the original APIs called HDF-EOS library to access the data. The HDF-EOS library supports C and Fortran. Grid, Point, and Swath interfaces are prepared.

The data format of HDF-EOS Ver.2 is used by MODIS sensor on Terra/Aqua and GLI sensor on ADEOS-II. HDF-EOS Ver.2 is based on HDF Ver.4. Recently, the utilization of HDF-EOS Ver.5 based on HDF5 increases, too. We target the HDF-EOS Ver.5, because we believe that it will have precedence over HDF-EOS Ver.2 in the near future.

Fig.1 shows a summary of HDF-EOS file structures. If Grid, Point, or Swath data exists under the root, GRIDS, POINTS, or SWATHS Group is created, respectively. There is a structure named Group to contain datasets under the root. Datasets in which actually observed data are stored exist under a Group.

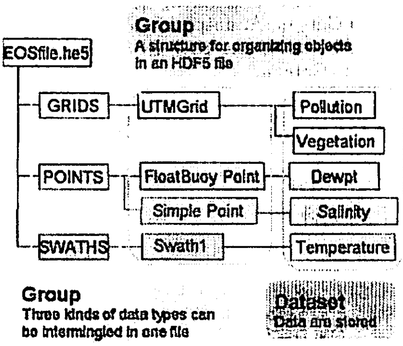

Figure1: HDF-EOS file structure

2.3 Data Models HDF-EOS Provides

This subsection briefly describes three kinds of data models (Grid, Point and Swath data) that HDF-EOS provides. A Grid dataset contains a known map projection, for example, Mercator or Homolosine. The observed value is stored in each grid. Eleven kinds of map projections are used in HDF-EOS.

A Point dataset is made up of a series of data records observed at irregular time intervals and at scattered geographic locations. Data are stored in a tabular form, and a relation between fields of the table can be defined with an ID.

Swath data is observed while a satellite moves along a track surface of the satellite. The satellite observes data by taking a scan perpendicular to the track surface. Axes perpendicular, vertical and parallel to the moving direction of the satellite are used to construct a multidimensional dataset. Swath data is stored in a field called Data Field.

3 HDF-EOS Read Function

This section describes the development of an HDF-EOS read function. The original HDF-EOS library is used for reading out data from HDF-EOS files. We implement the HDF-EOS read function in C so that atmospheric scientists can comprehend the HDF-EOS file structure without writing any program with the original library.

3.1 Development Language

As described in the section 2, HDF-EOS supports C and Fortran. Meanwhile, we develop Gateau in Java with VTK [7]. Therefore, when we implement the HDF-EOS read function for Gateau, it is necessary to call C programs that use the HDF-EOS library from the Java programs.

VTK is a C++ visualization library with a function of wrappers for other languages such as Java, Tcl or Python. Therefore, it is possible to write the program using VTK in Java. In our implementation, we connect a Java method and a C function via the visualization pipeline of VTK by wrappers. The implemented original module is called vtkEOSReader for reading out the data from HDF-EOS files. The next sub-section describes the implementation details of vtkEOSReader.

3.2 Implementation details of vtkEOSReader

We implement vtkEOSReader by the following procedures, which are the same process for creating VTK modules. The vtkEOSReader is a module to access an HDF-EOS file form Java. Note that the vtkEOSReader can be used as a VTK module. Using this module, VTK users can access HDF-EOS files from Tcl and Python by VTK wrapper. See Fig.2 for the composition.

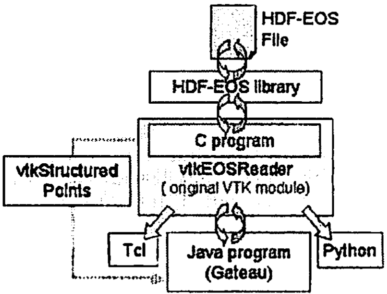

Figure2: vtkEOSReader composition

To create the original module, it is necessary to write a program to access an HDF-EOS file in C++ using the HDF-EOS library. Writing such a program, we implement several functions to access from HDF-EOS file in the C++ program: functions for opening an HDF-EOS file, obtaining the number of the HDF-EOS structure or the name of the structure. The original module makes the above functions written in C++ available from Java as the methods of a class called vtkEOSReader. The methods names that are used for accessing Grid data, Point data and Swath data, start with Gd, Pt and Sw, respectively. Table.1 gives some examples of vtkEOSReader methods for Grid data.

The data read from an HDF-EOS file are stored in the array declared in a C program. In order to visualize the data by Gateau, it is necessary to convert it into the data type that Gateau can read, namely, the data type for VTK. The HDF-EOS data stored in an array in the C program are converted into the vtkStructuredPoints data type. The converted data are passed to the Java program (Gateau), and the data are visualized by Gateau. The vtkStructuredPoints consists of three dimensional grids parallel to x-axis, y-axis, and z-axis. Each grid interval is fixed-sized.

Table 1: vtkEOSReader method examples (Grid)

Method Summary	
int	GdFileOpen(String *filename*) Opens a file and retrieves a file ID.
String	GdGetNameOfGrid() Retrieves a list of Grid structure names.
int	GdGetNumberOfFields(int *grid*) Retrieves the number of datasets under the Grid structure that is specified by *grid*.
vtkStructuredPoints	GdGetOutput(String *fieldname*) Converts the data in dataset specified by *fieldname* into vtkStructuredPoints data and returns those data.
void	GdClose() Closes the file opened by GdFileOpen().

3.3 File Structure and Data

When users read out data from HDF-EOS files, the names and the IDs of structures are used. It is necessary to comprehend the data structure in the file beforehand. A tool called h5dump is provided by NCSA for displaying an HDF file structure. Users generally comprehend the file structure by displaying the file structure with h5dump to know how the file structure is as well as the names of structures and the datasets in advance. After understanding the file structure, users write a program for reading data using the HDF-EOS library.

However, in this implementation, users expect that they open HDF-EOS files, which they do not know the data structure beforehand, to read out data directly. Therefore, we need to perform a query about the number of structures, the names of structures, and datasets to understand the file structure beforehand. Especially, it is necessary to inquire about Grid, Point and Swath because the entire picture of the file structure can be comprehended by retrieving each structure of Grid, Point and Swath. Then the retrieved information about the file structure is displayed by a tree view. If a dataset is chosen, the number of dimensions, the number of grids and other information about the selected dataset are inquired, and data are read out.

We take Grid as an example to describe a brief overview of retrieving the file structure and reading out data using vtkEOSReader. The names of the methods of vtkEOSReader are shown in parentheses. Steps 1 to 6 are the procedures to display the file structure by a tree view. Steps 7 to 10 are necessary after choosing the dataset that users want to read out from the tree view by mouse click.

1. Open a file and retrieve the file ID (GdFileOpen())
2. Retrieve the number of Grid structures (GdGetNumberOfGrid())
3. Retrieve the name of Grid structures (GdGetNameOfGrid())
4. Retrieve the ID of the Grid structure that users want to access (GdGetGridId())
5. Retrieve a list of Grid structure names (GdGetNumberOfFields())
6. Retrieve the names of datasets under the specified Grid structure (GdGetNameOfFields())
7. Retrieve the number of dimensions of the specified dataset (GdGetRank())
8. Retrieve the number of grids of each dimension of the specified dataset (GdGetDims())
9. Retrieve the data type of the specified dataset (GdGetDataType())
10. Convert the data in specified dataset into vtkStructuredPoints data and return the converted data (GdGetOutput())

It is necessary to retrieve the structure and read out data ditto with Point and Swath. An inquiry suitable for each structure is necessary, since Grid, Point and Swath have structures different to each other.

4 Implementation to Gateau

When atmospheric scientists analyze their data, they may want to observe the data as numerical values before visualizing them. We implemented an HDF-EOS browser that displays data as numerical values. When an HDF-EOS file is opened by Gateau, the numerical display of data or the 3D visualization can be selected by a mouse click. We explain the HDF-EOS browser and the visualization of HDF-EOS data in this section.

4.1 HDF-EOS Browser

We implement an HDF-EOS browser using vtkEOSReader. The HDF-EOS browser shows the file structure by a tree view to display data in the selected dataset, when an HDF-EOS file is accessed by Gateau. Gateau discriminates files by file type of ".h5" or ".he5" whether the file is an

HDF-EOS file. If the selected file is an HDF-EOS file, the file structure is displayed by a tree view on the left corner of Gateau screen. If a file with file type ".h5" or ".he5" which is not actually an HDF-EOS file is selected, the file ID is not retrieved when Gateau opens the file. Therefore, Gateau determines that the selected file is not an HDF-EOS file. In such a case, a message dialog, saying that the selected file is not an HDF-EOS file, is displayed.

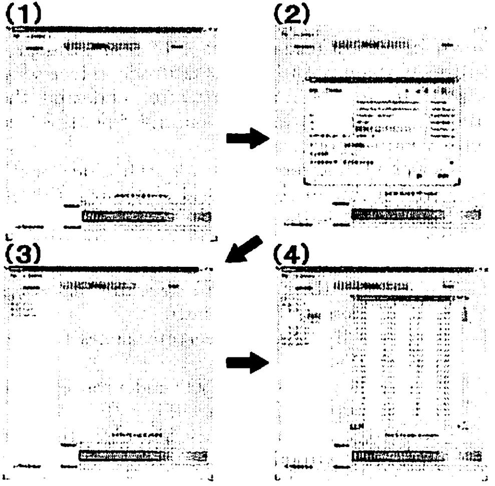

Figure3: The execution of the HDF-EOS browser

Fig.3 shows the execution of the implemented HDF-EOS browser. (1) in Fig.3 is the initial screen of Gateau. When the file selection button on the upper left corner of the screen is pushed, a file selection dialog appears. Atmospheric scientists select an HDF-EOS file that they want to visualize among the files displayed on the dialog (2). When an HDF-EOS file is selected, the structure of the selected file is retrieved. Then the file structure is displayed by a tree view on the left corner of the screen (3). Atmospheric scientists can comprehend the file structure by expanding nodes of the tree. When the node with the dataset name that they want to display is selected, the HDF-EOS file is accessed to read out data. Then a new frame that displays data numerically appears (4).

4.2 Visualization of HDF-EOS Data

We implement an HDF-EOS read function to Gateau with Java and vtkEOSReader. When a file is selected, the structure is displayed by a tree view as in the case of the HDF-EOS browser. When a dataset that atmospheric scientists want to visualize is selected, the selected HDF-EOS file is accessed. The data that are read out from the HDF-EOS file are stored in an array of C program. Then the data are converted into the vtkStructuredPoints data. The converted data are passed to Gateau, and 3D visualization is performed. The 3D visualization and the numerical display of the data can be performed simultaneously using the function of the HDF-EOS browser (see Fig.4).

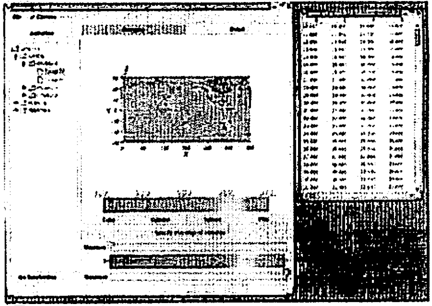

Figure4: Gateau incorporating an HDF-EOS reading function

5 Conclusions

In this paper, we reported the implementation of the HDF-EOS read function into Gateau, which is a 3D visualization system for atmospheric science research. By using Gateau, atmospheric scientists can achieve a 3D visualization without writing any program for accessing HDF-EOS data.

As a future work, we plan to develop a GUI that helps atmospheric scientists to analyze data more effectively. Gateau shows beforehand information about datasets to atmospheric scientists. Moreover, data preprocessing before visualizing data and selectable techniques for visualization should be displayed. We want to display simplified results after executing suitable processing of data by thumbnail when the mouse is put on a dataset that is displayed by the tree view.

6 References

[1] HDF-EOS Information Center, http://hdfeos.gsfc.nasa.gov/

[2] NetCDF, http://www.unidata.ucar.edu/software/netcdf/

[3] Guide to GRIB, http://www.wmo.ch/web/www/WDM/Guides/Guide-binary-2.html

[4] C.Watanabe, E.Touma, K.Yamauchi, S.Hayashida And K.Joe: Development of an Interactive Visual Data Mining System for Atmospheric Sciences, Proceedings of International Symposium on High Performance Computing VI (2005)

[5] E.Touma, K.Yamauchi, H.Kamo, N.Nide, S.Hayashida, K.Joe: Gateau: A 3D Visualization System for Intuitive Analysis of Atmospheric Science, PDPTA'06 (2006)

[6] The NCSA HDF Home Page, http://hdf.ncsa.uiuc.edu/

[7] The Visualization ToolKit, http://public.kitware.com/VTK/