

HDF EOS 数据格式及其数据导入

景毅刚 李登科 张树誉

(陕西省气象局农业遥感信息中心, 西安 710015)

摘要 HDF-EOS 作为 NASA 存储和发布 EOS 数据的标准, 专门用于处理各种 EOS 产品。掌握 HDF-EOS 数据格式和数据导入是能够充分开发和应用 EOS 数据信息的关键。文中介绍 HDF 数据文件中图像模块、科学数据、Vdata 模块(表格数据)的结构, HDF-EOS 数据文件中点、条带、网格数据和元数据的数据格式, 以及打开 HDF 格式数据实现对 HDF-EOS 数据导入浏览的相关软件。

关键词 HDF HDF-EOS 数据格式 数据导入

引言

HDF(Hierarchy Data Format)数据格式是美国伊利诺伊大学国家超级计算应用中心(NCSA, National Central for Super computing Applications)于 1987 年研制开发的一种软件和函数库, 用于存储和分发科学数据的一种自我描述、多对象的层次数据格式, 主要用来存储由不同计算机平台产生的各种类型科学数据, 适用于多种计算机平台, 易于扩展^[1]。

1993 年美国国家航空航天局(NASA)把 HDF 格式作为存储和发布 EOS(Earth Observation System)数据的标准格式。在 HDF 标准基础上, 开发了另一种 HDF 格式即 HDF-EOS, 专门用于处理 EOS 产品^[2], 使用标准 HDF 数据类型定义了点、条带、栅格 3 种特殊数据类型, 并引入了元数据(Metadata)。

目前, 陕西省气象局农业遥感信息中心已接收 EOS/MODIS 数据。如何在不同遥感图像处理软件或 GIS 软件中实现 HDF-EOS 数据导入及其和其它数据格式之间相互转换, 是 EOS 数据信息开发、推广和应用的关键。当前, 随着 EOS 数据广泛应用, 与 HDF-EOS 相关的浏览、处理软件不断推出, 特别是一些遥感图像处理软件和 GIS 软件已逐步支持

HDF-EOS 数据格式。这不但推进了 HDF-EOS 数据的发布、推广和应用, 而且为遥感及 GIS 用户及研究人员提供了丰富的数据。

1 HDF 组织结构

HDF 文件由路径和数据对象构成, 每个数据对象包括指向该数据对象位置指针的指针域和定义该数据类型的信息域。HDF 为每一种特殊数据类型都提供专门应用程序接口, 这些应用程序接口组织在一起构成 HDF-EOS 库, HDF 库包括 3 个接口层^[2], 从上到下分别是 HDF 底层、HDF 应用层、HDF 顶层(图 1)。HDF 底层为软件开发者所应用,

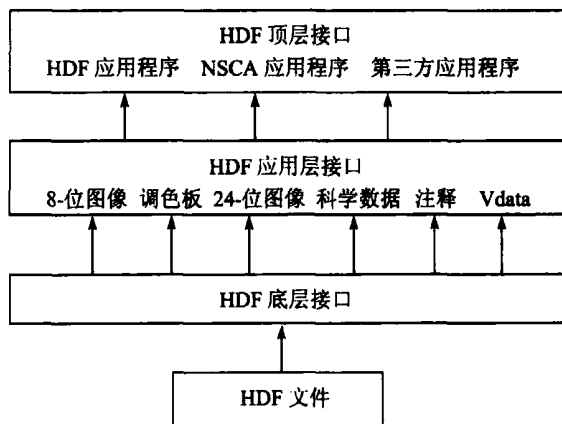


图 1 HDF 文件格式的构成

陕西省气象局“陕西省 EOS/MODIS 卫星遥感资料应用开发”课题资助

作者简介: 景毅刚, 男, 1968 年生, 高级工程师, 主要从事卫星遥感解译、应用开发, Email: Yigangjing@sina.com

收稿日期: 2004 年 1 月 9 日; 定稿日期: 2004 年 4 月 1 日

包括文件输入/输出、差错控制、内存管理、物理存储等应用程序接口。HDF 应用层接口包括 6 个独立模块分别用来简化 6 种数据类型(8 位图像、24 位图像、色彩、科学数据、注释、Vdata) 存储和访问过程。HDF 程序开发人员的工作主要是使用这些应用接口,而不必考虑底层细节。HDF 顶层包括 HDF 应用、NCSA 和第三方开发者制作应用程序。在底层和应用层中,HDF 库由可调用程序组成,这些程序被汇集在一起构成接口。每个接口都可访问特定

HDF 函数或 HDF 数据类型。HDF 子程序调用和数据类型访问通过接口实现。

1.1 图像模块(8 位图像和 24 位图像)

在 HDF 中,使用 8 位图像和 24 位图像两个模块实现图像数据存储^[2]。8 位图像模块中包括一个表示颜色图像二维数组,数组中每个值不是代表某个颜色值,而是一个单独调色板索引,调色板中每个条目代表一个含有红、绿、蓝 3 种颜色数值的颜色,数组中每个数据值是 8 位(图 2)。

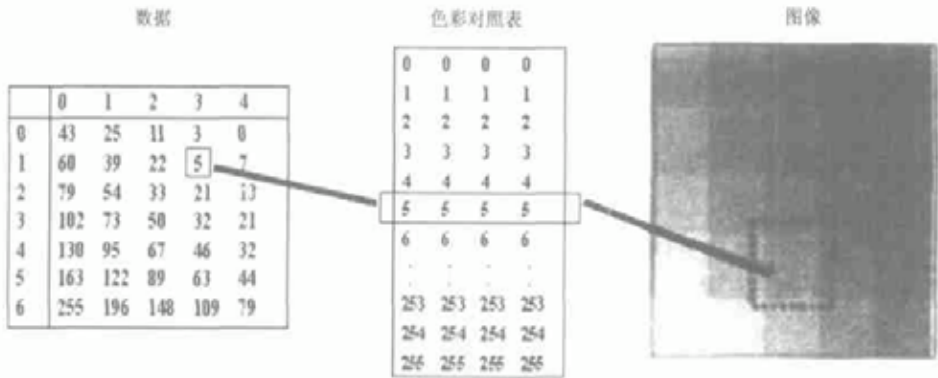


图 2 8 位图像 的表示

24 位图像与 8 位图像存储相似,只是它包括 3 个表示颜色图像二维数组,每个数组具有相同大小并分别代表某图像中每个色素红、绿、蓝值。

1.2 科学数据

HDF 定义的科学数据指图像数据之外其它对科学调查有用数据^[2]。这种数据使用多维矩阵存储。HDF 科学数据模块包括两个接口:一个是单文件接口即 DFSD 接口(因与此接口相关子程序以 DFSD 开头),这种接口一次只允许访问一个文件。另一个是多文件接口即 SD 接口,这种接口一次可以访问多个文件。SD 接口已被 ECS (EOSDIS (EOS Data and Information System) Core System) 用户广泛使用。DFSD 接口提供了一个程序集,用它可读、写任意行列和类型数组。通常把这种数组与相关信息统称为科学数据集(Scientific Data Set 简称 SDS)。SD 接口可以对多文件和数据对象进行同时操作,它的数据对象不仅包括标准 SDS 数据元素,还包括许多 DFSD 没有的附加属性。在 SDS 中,多维数组包括 8 位、16 位、32 位有符或无符整数以及 32 位、64 位浮点数。SDS 包括下列属性:①坐

标系统(Coordinate System),确定显示数据时所用坐标系统;②格式(Formatting),指定数据和属性值显示格式;③卷标(Label),描述一个独立变量和数据名字;④最大/最小值(Max/Min Value),存储由生产者提供的数据最大和最小值;⑤刻度(Scale),描述每一个坐标轴刻度;⑥单位(Unit),确定数据单位。在 HDF 中定义了两类科学数据表:①标准型,它把数据看成一张二维表,每一列中数据属于同一类型,每行大小固定(表 1);②带索引科学数据表(图 3)。

表 1 标准的科学数据表

编号	值	名称
1	2.34	Cygx1
2	- 89.43	Herx1
3	0.0023	Cygx3
4	1.115	Vela

1.3 Vdata 模块

在 HDF 文件中,Vdata 接口用于存储表格数据,每个表格由一系列记录组成,而每个记录又由一系列域组成。每个域的数据类型可以不同,但每个

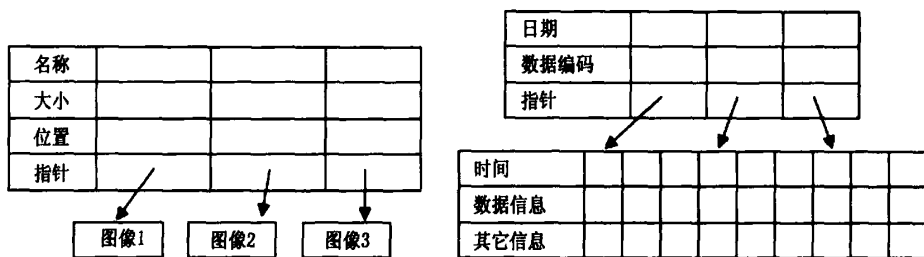


图3 带索引指针的科学数据表

记录必须包括相同的域。Vdata 表使用 3 种标识信息:表名、类和域名(图 4)。表名,实际上是一个卷标,用来描述该表来源和内容;类,用来表示数据意义以及通过数据意义区分其它表;域名,用来标识组成记录的每个域。

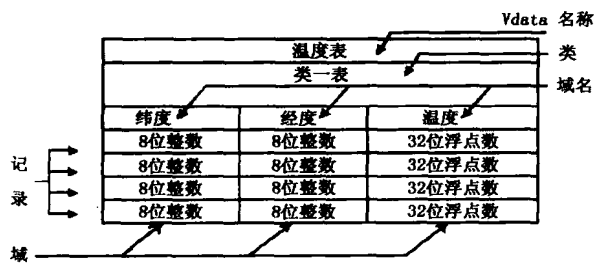


图4 Vdata 的存储方式

2 HDF EOS 数据格式

HDF-EOS 已成为 EOSDIS 数据生产和存档的标准格式,专门用于存储 EOS 数据^[2]。除了 HDF 的 6 种数据类型外,HDF-EOS 还支持另外 3 种数据类型:点、条带、网格。HDF-EOS 文件内容可以通过地理坐标和时间查询,每个 HDF-EOS 文件都包括元数据,它便于 EOS 数据访问。

2.1 点数据

点数据由在时间上间断的或空间上分散的一系列不规则数据记录构成,在某种意义上它构成了空间或时间上一系列间断点的状态或属性值。点数据模块适用于处理气象站数据、浮标数据、船载测量的海洋数据等。点数据模型是非常灵活的数据类型,它可以结合 HDF 的 Vdata、SDS、Vgroup 数据类型,对任意层空间数据进行存储、处理。通常有 3 种类型点数据:①由多个固定观测点提供的同一时刻的观测数据;②由某个观测站提供的不同时刻的观测数据,它由两个表构成,一个用来存储该站点的相关

信息(如经纬度、名称等),另一个用来存放观测数据;③由活动观测点获取的不同时刻的点数据。

2.2 条带数据

HDF-EOS 中的条带数据指卫星所携带的传感器沿一定路径对地表扫描,获得廓线范围内的各种有用信息(图 5)。条带数据获取有 2 种方式:①传感器沿轨迹方向获取数据;②传感器沿垂直轨迹方向获取数据。

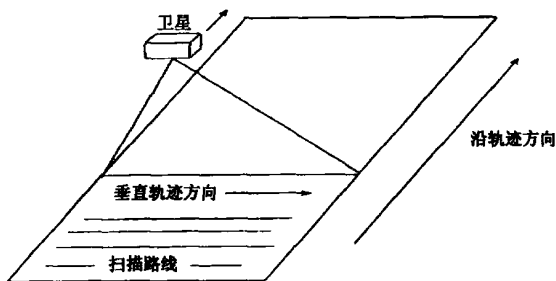


图5 典型的卫星条带数据

条带数据是 EOS 数据的主要组成部分,为空间信息研究提供了丰富原始数据。一个标准条带数据主要由数据域、地理坐标域、空间尺度和维映射构成。数据域是条带数据主要部分,通常以一维、二维或多维矩阵的结构存储来自传感器的原始数据或参数。空间尺度是在使用该库之前已被定义好,它通过给定数据和地理坐标域的名字和大小,确定整个数据的范围和属性特征;维映射定义了数据域和地理坐标域的关系,每个地理坐标域都有相应的数据域相对应。通常地理坐标域的大小与相应数据域的大小不同。条带数据存储方式见图 6。

2.3 网格数据

网格数据是最常用的一种处理空间数据的数据格式,被广泛应用于 GIS 和 RS 领域。HDF-EOS 网格数据接口与地图投影紧密结合,提供了相关的投

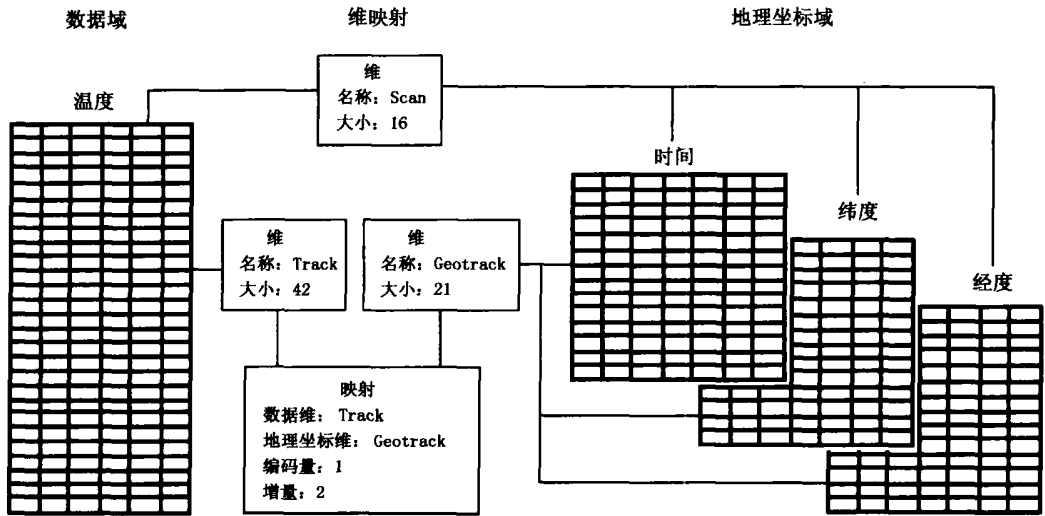


图 6 条带数据的存储方式

影参数和方程,使用户很容易获得网格中任意点经纬度,这更加适用于地理空间数据处理。HDF-EOS 网格数据集有 3 个主要特征:数据域、空间尺度、投影。其中数据域是网格最重要部分,它由二维或多维数组构成,数据域中每个数据项具有相同数据类型,并且它与一定的地理坐标系相对应。投影是网格的核心,它能够利用数学方程方便地将经纬坐标转换成 X-Y 坐标。投影种类很多,但 HDF-EOS 只支持 6 种:Geographic, Interrupted Goode's Homolosine, Polar Stereographic, Universal Transverse Mercator, Spaced Oblique, Lambert Azimuthal Equal Area, 这些投影都存放在专门程序库中,称为通用坐标转换软件(GCTP, General Coordinate Transformation Package)。HDF-EOS 提供了全部程序的源代码,可以根据各自需要指定投影参数,并进一步开发自己需要的投影类型。

2.4 元数据

元数据常被称为“关于数据的数据”,它主要用于表现源数据的特征、描述源数据的关系、实现源数据的快速查找和有效使用。元数据对于数据的存储和发布很重要,因为它记录数据集的主要特征,包括数据结构和内容描述,以及数据集开发的历史等。在 EOS 中定义了 3 种元数据:①核心元数据(core metadata),它能够满足所有标准数据产品需要;②具体产品元数据(product specific metadata),它只能满足特定数据产品需要;③结构化元数据(structural

metadata),用于描述 HDF-EOS 文件中数据域具体细节包括空间尺度信息、数据域信息和地理位置信息,它是 HDF-EOS 所特有元数据,不同数据类型结构化元数据的作用不同。

3 HDF-EOS 数据导入浏览

目前,NOAA/AVHRR 卫星数据应用最广泛,而新一代极轨卫星 EOS/MODIS 不但保留了 AVHRR 功能,而且在数据波段数目和数据应用范围、数据分辨率、数据接收和数据格式等方面都作了相当大改进,这些改进使 MODIS 将成为 AVHRR 的换代产品^[3]。因此,随着 EOS 数据广泛应用,HDF-EOS 数据格式将逐步被遥感、地理信息系统用户和软件开发商所接受,随之出现许多针对 HDF-EOS 数据的导入浏览软件,通过它们可以对 HDF-EOS 数据的导入浏览并可转换其存储格式。

3.1 HDF-EOS 版本

随着 EOS 数据广泛应用,NASA 以 HDF 为基础对 HDF-EOS 的功能不断扩充,版本升级。在 HDF4 基础上开发了 HDF-EOSv2.x(x=0,1,2,3,4,5,6,7);进一步又根据 NCSA 在 1998 年推出的 HDF5 开发了 HDF-EOSv3.0 版本。HDF5 与 HDF4 相比,增添了许多功能,如文件存储容量可超过 2 Gb,数据对象的数目不再局限于 200 000 个以内,输入/输出功能更灵活和完善,数据模式更简单等。

3.2 HDF EOS 浏览软件

目前,已出现许多 HDF-EOS 的浏览软件,如 EOS View, view-hdf, Web Winds, HDF Explorer, HDF Look, JHV, DIAL, HDF Inspector 等。下面简要介绍 EOS View 和 view-hdf 软件。

(1) EOS View 是 NASA 为 ECS 项目开发的检查和校正 HDF 和 HDF-EOS 数据文件的文件浏览工具。使用 EOS 数据产品的用户可以通过所显示的元数据域和数据对象浏览 HDF 文件和单个对象内容。EOS View 支持图像、光栅图像、多维数组、文本、表格、Vdata、Vgroup。属性和注释也可以被浏览。EOS View 应用软件能够将 HDF 和 HDF-EOS 数据文件转换成 ASCII、二进制等特定类型,还能够提供将 HDF 和 HDF-EOS 文件加入其它软件系统的手段。

(2) view-hdf 是由 NASA 的 Langley 研究中心的 CERES 数据管理组开发并由 Langley 大气科学数据中心发布的分析工具,用来访问以 HDF 和 HDF-EOS 格式存储的数据。view-hdf 工具能够选择和设置 HDF 文件的 SDS 和 Vdata 结构中变量、渲染二维和三维图像以及在多种世界地图投影上绘制有地理坐标的数据。此外 view-hdf 还有其它一些特征,如小块地图能够以 PostScript、压缩的 PostScript 和 GIF 格式保存,还可以直接发送打印,提取的子集和统计结果可以被写入 ASCII 文件中。

3.3 遥感图像处理和 GIS 软件对 HDF EOS 支持

RS(Research System)是著名的遥感图像处理软件开发公司,该公司的 IDL、ENVI、Noesys 系列产品软件都支持 HDF,其中 IDL(Interactive Data Language)是程序员操作和显示 HDF 数据的工具,ENVI 可以容易地实现以 HDF 格式存储的遥感图像处理^[4],Noesys 是访问、浏览、组织和管理 HDF 文件的低成本应用软件。对于使用 HDF 数据来说,IDL、ENVI、Noesys 系列产品是很好的 EOS 数据分析和虚拟现实的软件。ESRI 并没有开发 HDF-EOS 格式与 ARC/INFO 之间转换的工具软件。但是 NASA 针对 GIS 用户对 EOS 数据的需求开发了 EOSGIS 数据格式转换器,用来实现 HDF-EOS 数据格式和 ARC/INFO 以及 ERDAS IMAGINE 所支持的 Geo TIFF 数据格式之间的转换。EOSGIS 将逐步实现 HDF-EOS 向其它 GIS 软件所支持的数据格

式转换。EOSGIS 软件的出现将促进 EOS 数据在 GIS 领域的应用。此外 EOSGIS 将 HDF-EOS 数据格式转换成 Geo TIFF 格式之后,可以在 ERDAS IMAGINE 中使用。

3.4 DIAL 工具

DIAL(Data and Information Access Link)是 NASA 开发的基于 WWW 的科学数据发布工具。它支持 HDF、HDF-EOS、文本、二进制、HTML、TIF 等格式,主要用来通过网络发布 EOS 数据。用户通过网络浏览器连接到 DIAL 网址,可以实现:基于空间、时间、参数的搜索;查看数据和元数据;浏览、提取数据,并且可以指定数据范围;在线下载多种格式的数据;根据 X-Y 坐标,自动生成表格数据。

3.5 Imapp2daac.exe 工具

由于 HDF-EOS 数据格式是针对 EOS 卫星信息存储的数据格式,它和标准 HDF 数据格式有区别,为了把 HDF-EOS 转换成标准的 HDF 数据格式,可在 DOS 环境下利用 Imapp2daac.exe 应用程序把 HDF-EOS 格式转换成标准的 HDF 格式。

4 结语

本文对 HDF、HDF-EOS 的数据格式及 HDF-EOS 数据导入浏览进行了介绍。可以看出该数据格式在存储和处理空间数据,特别是在存储和处理遥感、地理信息系统数据方面有其自身的优势,所以它已逐步受到广大遥感、地理信息系统用户和研究人员的重视。HDF-EOS 有文件模式多样性的特点,针对每种数据模式,HDF-EOS 都提供了相应的读、写、访问、查询接口;HDF-EOS 支持多种计算机平台,使得该格式更容易实现多平台操作和共享,在许多软件支持下可进行数据导入或格式转换,其中 IDL、ENVI 是较理想的软件,Imapp2daac.exe 可把 HDF-EOS 格式转换成标准 HDF 格式。

参考文献

- 1 张璞,崔彩霞.遥感数据解析平台建设初探.新疆气象,2003,(6):26-27
- 2 黄春林,李新.HDF-EOS 数据格式在处理空间数据中的应用.遥感技术与应用,2001,16(4):252-259
- 3 刘闯,葛成辉.美国对地观测系统(EOS)中分辨率成像光谱仪(MODIS)遥感数据的特点与应用.遥感信息,2000,(3):45-48
- 4 张京红,景毅刚.遥感图像处理系统 ENVI 及其在 MODIS 数据处理中的应用.陕西气象,2004,(1):27-29