

基于 .NET 的 HDF 光栅图像数据的读取与处理

HDF raster image data reading and processing based on .NET

焦 飞 (肇庆市工贸学校 广东肇庆 526000)

黄天文 (广东省肇庆市气象台 广东肇庆 526040)

摘要:论文对 HDF 数据格式和函数库进行研究,重点以栅格图像为例,详细论述如何利用 VC++ .net 和 VC#.net 对光栅数据进行读取与处理,然后根据所得到的象素矩阵用描点法显示图像。论文是以国家气象中心开发 Micaps3.0(气象信息综合分析处理系统)的课题研究为背景的。

关键词:HDF 光栅图像 图形图像 描点法 .NET

1 引言

目前国家气象中心正致力于开发 Micaps3.0,其中涉及到 HDF 格式数据的读取与处理,尤其是气象卫星接收的大部分数据存储格式为 HDF 栅格图像,研究如何对这些数据进行读取是非常必要的。国内专门介绍 HDF 格式的论文不多,在开发过程中需要参考国外有关的论文和网站介绍的资料,然后将这些理论应用到实践。

2 HDF 简介

HDF 是用于存储和分发科学数据的一种自我描述、多对象文件格式^[1]。该文件格式由美国伊利诺伊大学(the University of Illinois)的 NCSA 组织(The National Center for Supercomputing Applications)开发,英文全称为 Hierarchical data format,是一种超文本文件格式^[2],综合管理 2D、3D、矢量、属性、文本等多种信息,能够帮助科学家摆脱不同数据格式之间相互转换的繁琐,而将更多的时间和精力用于数据分析。HDF 能够存储不同种类的科学数据,包括图像、多维数组、指针及文本数据。HDF 格式还提供命令方式,分析现存 HDF 文件的结构,并即时显示图像内容。科学家可以用这种标准数据格式快速熟悉文件结构,并能立即着手对数据文件进行管理和分析。

在国内,国家卫星气象中心接收与处理的美国新一代地球观测卫星(EOS)中分辨率成像光谱仪(MODIS)资料可广泛用于气象、环境、林业、渔业、港口、交

通、自然灾害监测等领域^[3],其 MODIS 1B 数据就是基于 HDF 格式的。

HDF 文件格式的优势在于:可移植性强(独立于操作平台);属于超文本文件;可以存储并处理大数据量;一个文件集可以管理多种类型的数据结构;具有可扩展性。由于 HDF 的诸多优点,这种格式已经被广泛用于目前国外各种卫星传感器的标准数据格式,包括 NASA 已发射的 Landsat-7 号卫星。

HDF 文件由一个目录(directory)和一个数据对象集(collection)组成。在许多情况下,directory 接口和数据对象集是一一对应的。每个数据对象有一个目录接口,包含了一个指向对象地址的指针和定义数据类型信息。

HDF 库包括 3 个接口层^[4],从上到下分别是 HDF 底层、HDF 应用层和 HDF 顶层。该库支持以下数据类型:

- (1) 科学数据组:多维整型或浮点数组;
 - (2) 定点数据(Vdata&groups):以表来存储多值数据;
 - (3) 一般栅格(Gr):栅格影像;
 - (4) 注释——描述文件和部分文件(元数据)的文字;
 - (5) 8-bit 栅格影像;
 - (6) 24-bit 栅格影像;
 - (7) 调色板:8-bit 调色板(伴随影像);
- 为了便于理解 HDF 文件的格式,很多资料形象地

把 HDF 文件看成为一本有表格内容的多章节书。HDF 文件是“数据书”，其中每章都包含一个不同类型的数据内容。正如书籍用一个目录表列出它的章节一样，HDF 文件用“data index”（数据索引）列出其数据内容。

参考“国家 MODIS 数据中心”介绍的资料可以知道 HDF 文件结构如图 1 所示。

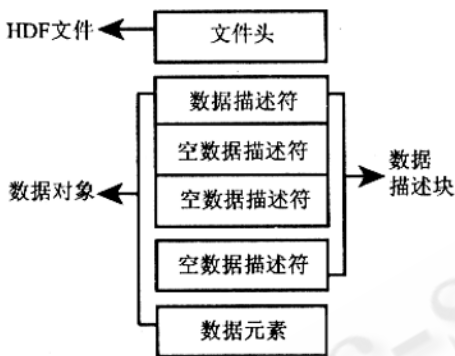


图 1 包含一个数据对象的 HDF 文件结构

其中文件头用来确定是否为 HDF 文件，是一个 32 比特的值，它由 4 个 ASCII 码形式的控制字符组成，每个字符占用一个字节，分别为 N、C、S、A。数据对象是 HDF 的基本结构。一个数据描述符和一个数据元素组成一个数据对象。数据描述符包含了数据元素的类型、位置、尺度等信息。数据元素则是实际的数据资料，是数据对象的原始数据部分，包含每个像素的值。所有的数据描述符都为 12 个字节长，包含个部分：一个 2 字节的标识符，一个 2 字节的参照数，一个 4 字节的数据偏移量和一个 4 字节的数据长度。如图 2。

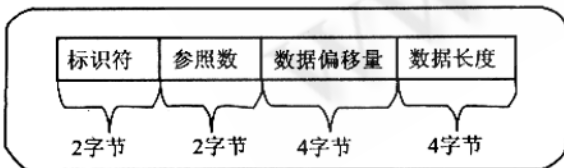


图 2 数据描述符内容

其中，标识符是数据描述符域，表示存于相应数据内容的数据类型。例如 306 是光栅图像对象的识别符。

参照数是一个 16 比特无符号整型数。HDF 文件中的每一个对象，由 HDF 库和数据描述符中的标签确定一个唯一的引用字。在引用字确定的数据对象间，标签和引用字不能改变。标签和引用字的结合可以唯一确定文件中对应的数据对象。

引用字没有必要连续指定，因此，在一个具有相同标签区分对象的方法后，不能假设引用字的值有任何意义。有时应用程序员也会发现在他们的程序把一些另外的信息加到引用字中是很方便的，但必须强调的是，HDF 库本身并不识别这些含义。

数据偏移量是一个 32 比特无符号整型字。通过存储文件开始时的字节数和数据内容开始时的字节数，指明文件中数据内容的位置。

数据长度是一个 32 比特无符号整型字。它表示整个数据内容的字节大小。数据内容增加，其长度也要增加。

3 获取和安装 HDF 库

HDF4.1r3 可以从 NCSA 的网站免费下载。这个版本的 HDF 库包括 C 和 FORTRAN 库文件、id 文件、HDF 命令行实用程序以及 HDF 文档。源代码和编译后的库都可以从 NCSA 那里得到。编译成的 HDF 库可用于 Windows 9x/NT、Macintosh 和 Unix 平台。下面以 Visual Studio .net2003 为例，介绍如何进行设置以使用 HDF 库。

首先把下载 HDF 库压缩包解压。在 Visual Studio .net 环境中新建一个 VC++ 项目，点击菜单“工具”下的“选项”，选择“项目”下的“VC++ 目录”，“包含文件”，指定包含文件所在的目录。目录就是刚才解压的 szip20\include, zlib122\include 和 5-165-win-net\include 三个目录，它们下面有 89 个 h 文件；“库文件”，指定为 \zip20\lib、\zip20\dll、\zlib122\lib、5-165-win-net\dll 和 5-165-win-net\lib 五个目录，这些目录里共有 8 个 lib 文件；然后再把 -165-win-net 目录里的 hdf5.dll、hdf5_cppdll.dll、\zlib122 目录里的 zlib1.dll 和 \zip20 目录里的 szlibdll.dll，共四个文件复制到 windows\system32 目录，或者复制到 windows\system 目录；接着采用 Dlls 编译：选择“文件”，“新建”，“项目”，“VC++ 项目”，“win32 项目”，在应用设置里选“控制台应用”，“空项目”，“添加存在

项”，“C/C++ 程序”；再新建一个项目，选择菜单“项目”，“属性”，“配置管理器”，选择“release 链接器”，“调试”，“生成调试信息”，选择“No”，在“链接器”中选择“输入”，“附加依赖项”，添加 hdf5dll.lib、hdf5.cppdll.lib、zdll.lib 和 szlibdll.lib 四个文件名；最后在“C/C++ - 预处理器”的“预处理器定义”里添加“_HDF5USEDLL_；HDF5Cpp_USEDLL_”，在“C/C++ 代码生成”的“运行时库”选择“multi-threadedDLL (/MD)”。这样就可以在.NET 环境下利用 VC++ .net 调用 HDF 库进行编程了。

4 光栅图像的读取

常规光栅 (GR) 数据模型，即 GR API。利用 GR API (应用编程接口) 可创建、写入、读取和查询光栅图像。常规光栅图像是一个数据结构，用于存储和描述光栅图像和与之关联的调色板，如图 3 所示。

EOS/MODIS 数据^[6]，在不同遥感图像处理软件或 GIS 软件中实现 HDF2EOS 数据导入及其和其它数据格式之间相互转换的方法值得借鉴。

现在 HDF 最新版本为 5.0，关于 HDF5 的数据格式及读写操作的例子可查阅 The HDF Group 网站^[7]。虽然 HDF4 和 HDF5 在数据结构上大相径庭，但它们具有一个共同的特点是分层式数据管理结构。

在 HDF4 中与影像有关的是 DFR8API 和 DF24API，专门用来提取 8 位和 24 位栅格影像。例如：

Status = DFR8getdims (filename, &width, &height, &haspal) 用来获取影像的高、宽和调色板个数。

Status = DFR8getimage (filename, image2data, width, height, pal2data) 得到影像数组和调色板数组。

在 HDF5 中，NCSA 发布的 hdf2hAPI 提供了丰富的函数帮助用户提取影像信息。例如用 H5Imget - image - info 函数输出影像的宽、高、面板个数、交错模式、调色板个数等参数，根据宽、高、面板个数可以为影像数据分配内存空间，定义一个数组来存放。再调用 H5Imread - image 函数读出这个数组。对于每个调色板都可以用 H5Imget - palette - info 得到调色板的维数信息，它是一个由 2 个元素组成的数组，如一个 256 色的 RGB 调色板可表示为 pal_dim[2] = {256, 3}。根据输出的维数信息参数可为调色板数据分配内存空间，定义一个数组来存放。再调用 H5Imget - palette 函数读出这个数组^[8]。

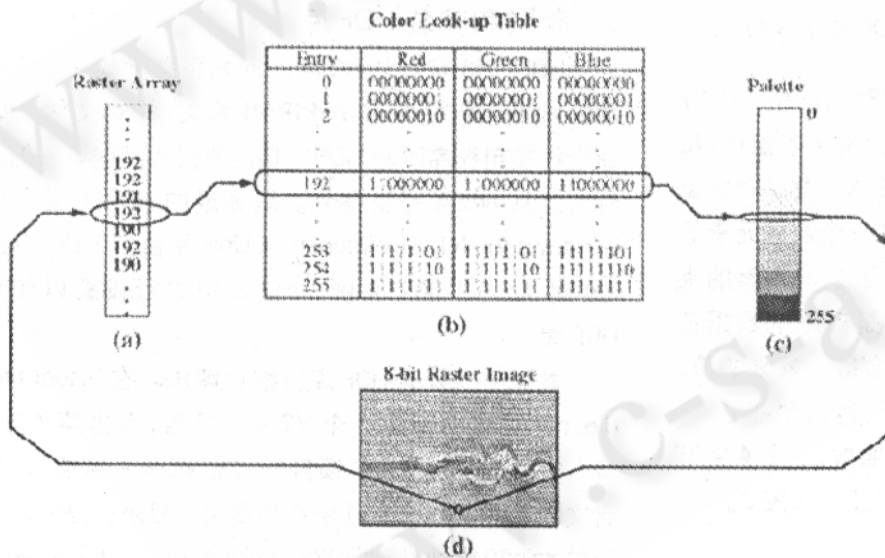


图 3 光栅图像和调色板

每一个 GR 对象必须包含如下内容：name (名字)，dimensions (维数)，image array (图像数组) 和 pixel type (象素类型)。创建 GR 对象时，要求用户必须提供这些参数^[5]。

GR 对象中有两类可选内容：palettes (调色板) 和 attributes (属性)。

目前，陕西省气象局农业遥感信息中心已接收

5 对图像的处理与显示

数字图像在计算机上以位图 (bitmap) 的形式存在，位图是一个矩形点阵，其中每一点称为像素 (pixel)，像素是数字图像中的基本单位。一幅 m × n 大小的图像，是由 m × n 个明暗度不等的像素组成的。数字图像中各个像素所具有的明暗程度由灰度值 (gray level) 所标识。一般将白色的灰度值定义为 255，黑色灰度值定义为 0，而由黑到白之间的明

暗度均匀地划分为 256 个等级。对于黑白图像,每个像素用一个字节数据来表示,而在彩色图像中,每个像素需用三个字节数据来表述。彩色图像可以分解成红(R)、绿(G)、蓝(B)三个单色图像,任何一种颜色都可以由这三种颜色混合构成。在图像处理中,彩色图像的处理通常是通过对三个单色图像分别处理而得到的。对于位图的相关概念这里就不再详细讲述。

如果要将彩色图像转换为灰度图像,只要将图像中的每个像素取出来,然后取像素的 R、G、B 颜色分量。

6 以黑白方式显示图像

在气象应用中经常要把彩色图像用黑白方式显示,其黑白化处理通常有三种方法:最大值法、平均值法和加权平均值法。这三种方法可以根据对图像黑白化处理的具体要求进行选择,通常采用平均值法。三种方法的原理如下:

最大值法。最大值法使每个像素点的 R、G、B 值等于原像素点的 R_0 、 G_0 、 B_0 值中最大的一个,即 $R = G = B = \text{MAX}(R_0, G_0, B_0)$ 。最大值法会形成亮度很高的黑白图像。

平均值法。平均值法使每个像素点的 R、G、B 值等于原像素点的 R_0 、 G_0 、 B_0 值的平均值,即 $R = G = B = (R_0 + G_0 + B_0) / 3$ 。平均值法会形成比较柔和的黑白图像。

加权平均值法。加权平均值法根据需要指定每个像素点 R、G、B 值的权数,并取其加权平均值,即 $R = G = B = (W_r * R_0 + W_g * G_0 + W_b * B_0) / 3$ 。其中: W_r 代表该像素点 R 的权数,大于 0; W_g 代表该像素点 G 的权数,大于 0; W_b 代表该像素点 B 的权数,大于 0。通过对 W_r 、 W_g 、 W_b 取不同的值可以获得不同的效果。

根据像素值就可以用描点法进行作图了。由于在开发过程中主要使用 C#.net 语言,下面就以 C# 编写的代码为例讲解。主要是通过 Bitmap 类的 SetPixel() 方法和 GetPixel() 方法以描点的方式实现效果。假设 pictureBox1 控件显示的是原始的卫星云图,赋给位图对象 MyBitmap,则可以利用平均值法转为黑白图像,再描点显示。

```
for (int x = 0; x < Width; x++) //Width 为图
像的宽度
```

```
for (int y = 0; y < Height; y++) //Height
为图像的长度
```

```
{ pixel = MyBitmap. GetPixel(x, y);
int r, g, b, Result = 0;
r = pixel. R;
g = pixel. G;
b = pixel. B;
Result = ((r + g + b) / 3);
bitmap. SetPixel(x, y, Color. FromArgb(Re-
sult, Result, Result));
```

```
}
pictureBox1. Image = bitmap;
```

其中,GetPixel() 方法用于获取 Bitmap 对象中指定像素点的颜色,SetPixel() 用于设置 Bitmap 对象中指定像素点的颜色。描点法必须调用像素点经过处理后的红、绿、蓝值,因此使用了 Color 结构的静态方法 FromArgb()。

7 以反色方式显示图像

图像的反色实际上就是取每一个像素点的相对颜色值,例如,如果图像的某一个像素点的 RGB 值为 (r, g, b) ,则它的反色值就为 $(255 - r, 255 - g, 255 - b)$ 。其它处理与实现方法可参考上面的“以黑白方式显示图像”。这些方法涉及到 Windows 平台下的 .Net Framework GDI+ 类,在开发过程中主要参考了“基于 GDI+ 的通用图形平台设计”一文^[9]。

8 结束语

论文是以国家气象中心开发 Micaps3.0 的课题研究为背景的。在开发 HDF 数据处理的模块中参考了国内外相关资料,研究了 HDF 的组织结构,并使用 VC++ .net 和 VC#.net 实现了数据的读取,栅格图像的显示和变换。这些方法对于处理和管理地理信息系统的大量图形数据和属性数据具有一定的借鉴作用,掌握和运用 NC-SA 提供的 API 提取影像数据,可以节省时间,提高程序编写效率。希望该文能对基于 .NET 的有关 HDF 数据处理的软件开发提供参考。

(下转第 56 页)

参考文献

- 1 张璞、崔彩霞,遥感数据解析平台建设初探[J],新疆气象,2003,26(6):26-27.
- 2 刘玉洁、杨忠东,MODIS 遥感信息处理原理与算法,北京:科学出版社,2001.
- 3 李登科、张树誉,EOS/MODIS 遥感数据与应用前景,陕西气象,2003,1(2):37-40.
- 4 黄春林、李新,HDF2EOS 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.
- 5 HDF 使用简介[OL]. <http://satellite.cma.gov.cn/eos/hdfc3.html>,2007-1-23.
- 6 景毅刚、李登科、张树誉,HDF-EOS 数据格式及其数据导入,气象科技,2005,33(3):278-282.
- 7 HDF5 Tutorials and Examples[OL],<http://www.hdfgroup.org/tutorial5.html>,2007-1-23.
- 8 王玲、龚健雅,基于 HDF 文件的组织方式与影像提取,测绘通报,2003,4:35-37.
- 9 张玲、陈元春、孙勇,基于 GDI+ 的通用图形平台设计,计算机工程,2005,31(12):218-220.