文章编号：1007-4619 (2004) 05-0404-05

高光谱数据在数据库中的高效存储技术研究

张雄飞, 张　兵, 张　霞, 郑兰芬, 童庆禧
(中国科学院 遥感应用研究所, 北京　100101)

摘　要：　通过分析高光谱数据的特点, 结合数据库系统开发实践, 提出了高光谱数据集, 包括图像、光谱、属性等, 在关系数据库中的存储规范。结合 ORACLE 数据平台, 提出了 3 种高光谱数据的存储模式, 并结合应用实例进行对比, 分析了各自的优缺点以及适用条件。

关键词：　高光谱; 关系数据库; 存储; Oracle
中图分类号：　TP79/TP391　　文献标识码：　A

1　引　言

高光谱分辨率(简称为高光谱) 遥感或成像光谱遥感技术的发展是过去 20 多年中人类在对地观测方面所取得的重大技术突破之一, 是当前遥感的前沿技术[1]。它融合了空间成像技术和光谱分光技术, 其核心特点是图谱合一, 即能够获取目标的连续、窄波段的图像光谱数据。高光谱技术自从诞生以来, 就以其丰富的光谱维信息显著区别于传统的多光谱遥感技术, 在地质、植被生态、土壤, 以及城市应用等方面取得了重要的成果。高光谱分辨率遥感信息的分析处理集中于光谱维上进行图像信息的展开和定量分析。通过高光谱成像所获取的地球表面图像包含了丰富的空间、辐射和光谱三重信息, 因此, 自从 20 世纪 70 年代末, 美国喷气推进实验室(JPL) 在美国宇航局(NASA) 支持下首先对成像光谱仪进行概念设计和研究以来, 基于高光谱的地物目标识别和属性探测就成为了遥感应用领域的一个重要发展方向。航空成像光谱仪已经发展成熟, 现在运行的有美国的 WIS, AVIRIS; 加拿大 CASI, SFSI; 澳大利亚的 HYMAP 和中国的 PHI, OMIS 等。在航天领域中, 除人们所熟知的美国对地观测系统(EOS) 计划中的中分辨率成像光谱仪(MODIS) 和欧洲空间局的中分辨率成像光谱仪(MERIS) 之外, 搭载 220 通道高光谱仪 Hyperion 的 EO-1 卫星已经升空。可以预料, 高光谱成像卫星在地球资源开发利用及地球环境监测中将发挥越来越重要的作用。

如图 1, 图像立方体是高光谱数据存储的一个主要特征, 其中 x, y 轴表征了它们的几何坐标以及

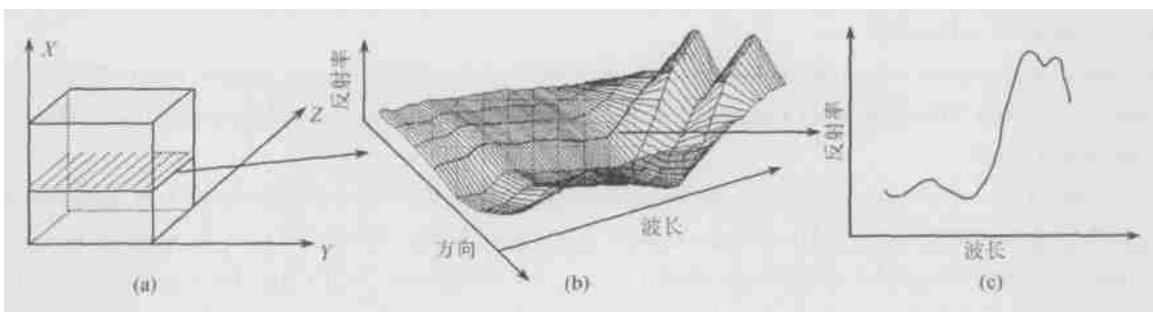

图 1　高光谱图像立方体数据结构示意图[2]
(a) 高光谱图像立方体; (b) 某个方向上的光谱剖面; (c) 像素的光谱
Fig. 1　Sketch map of hyperspectral image cube structure

收稿日期：2003-05-30; 修订日期：2003-07-24
基金项目：国家自然科学基金(编号：40271085); 国家 863-13"我国典型地物标准波谱库"项目支持; 北京市自然科学基金"面向精准农业的作物光谱数据库研究"支持。
作者简介：张雄飞(1977—　), 硕士研究生, 主要研究领域为网络化高光谱数据库系统。E-mail: tbsmooninwell@yahoo.com.cn。

在地面上相对的空间位置关系, Z 轴则代表了光谱通道。在图像数据立方体中, 任意一个像素点都可以提出一条相应的光谱曲线, 每条曲线实际上是一系列的数据点, 对应于相应的光谱波段。可见, 这种新型的数据结构, 将几何空间与光谱空间很好地结合在一起, 以三维数据立方体的形式表现。利用高光谱遥感图像立方体, 一方面可以通过目标的空间几何形状信息对目标特征进行分析和定位; 另一方面, 可以通过目标的波谱特性来识别目标或揭示目标更丰富的物理和化学属性。

本文旨在探讨针对高光谱遥感光谱数据以及图像立方体为对象的高光谱数据库, 构造一种规范、合理的数据库存储规范与模式, 达到高光谱数据的高效管理、共享和传递的目的。

2　数据特点

如前所述, 高光谱遥感数据以图像立方体的形式存储, 从数据库数据结构分析的角度来说, 它具有一些显著不同的特点:

(1) 高光谱遥感数据一个对象对应几十乃至上千个波段的数据, 而且每个波段的数据都可能需要单独提出, 用作分析、计算。

(2) 数据库中必须兼容多种波段参数设置的光谱数据, 因为高光谱成像系统参数会有不同的波段数量、位置上的设置和光谱采样设置。

(3) 高光谱数据的图谱合一的特性, 需要将变化性很强的图像光谱维数据与图像空间维数据很好地结合在一起。

(4) 高光谱数据中图像和光谱是紧密联系的一个整体。通常, 图像的存储也有多种方式, 常见的是以文件方式存放于操作系统之中, 但如何针对高光谱数据的特点, 选择一种合适的方式保证图像与其他数据的整合性需要精心设计。

(5) 高光谱数据的辅助数据和属性复杂, 种类繁多, 一般为查询的检索关键字, 需要兼顾数据结构的简洁性与适度冗余性, 从而兼顾速度、安全性和存储空间。

3　高光谱数据库数据结构

3.1　关系数据库中高光谱数据的基本存储规范

从上述高光谱数据特点可以看出, 海量数据在数据库系统中数据结构的设计将直接决定数据库系统性能。在数据库中, 表是存储的基本单位, 也是本文关注的重点。综合前人的研究成果[3], 提出了在关系数据库中高光谱数据基本的存储规范为:

光谱数据表组＋属性数据表组。

光谱数据表组中的表主要存放对象的光谱数据; 属性数据表组中的表则主要存放对象的各种其他属性数据, 包括: 测量属性数据、地学属性数据、特征属性数据和图片属性数据。每个表组中表的数量根据数据量的大小而定, 两者通过能够唯一确定对象样本的字段进行连接, 即: 主关键字。如果在大型数据平台下, 例如: Oracle 中, 相对较独立且数据量比较大的字段, 如: 光谱数据、图片属性数据等, 一般需要单独开辟表空间存储, 以有利于提高查询效率。图像数据一般采用 Blob 方式存储于表中[4], 这样虽然在读取速度、开发难度上会有所不便, 但因为图像数据在这里和其他属性数据是一个整体, 为了保证全部数据的一致性、可迁移性以及完整性, 需要将图像数据以这种方式存入。

该存储规范有以下几个突出的优点:

(1) 高光谱数据除光谱外的其他属性数据繁多, 分为两个表组可以使结构更为清晰, 有利于系统的整理、扩展等等。

(2) 光谱数据将是高光谱数据应用的重点, 查询、读取等操作的频率远大于其他数据, 将光谱数据与其他数据分开, 有利于数据库的维护与安全性。

(3) 查询时的关键字都是其他属性的数据, 将光谱数据与之分离, 可以提高系统查询效率。

(4) 光谱数据的存储模式可能会根据需要发生变化, 将两者分离有利于光谱数据的操作、维护与拓展。

上述数据规范, 对高光谱数据在关系数据库中的存储与管理和各种高光谱数据库应用系统的研究、开发有着重要的意义。

3.2　在 Oracle 数据平台下的几种数据存储结构的比较

根据上述存储规范, 运用到比较常见的 Oracle 数据平台, 对光谱数据表可以实现的模式和几种存储结构对比研究如下:

波段独立顺列式　该模式是在光谱数据表中, 以每一个波段及其相应的反射率值作为一个记录, 相应的两个主要字段均以 Number(m, n) 作为字段类型。这种模式的优点是存储时波段相互独立, 存储、查询、处理等速度快, 便于分别提取, 特别有利于波段值单独操作, 可以对冗余的定位字段进行各种

数据库性能调整操作, 如: 加索引(index)、建立分区等等, 以提高效率。这种存储方式的缺点是记录数相对较多, 冗余字段会浪费一定的存储空间。

波段集中整合式 这个模式是在光谱数据表中, 以每一个样本为一条记录, 无论是波段值还是反射率值均以一个类似文件的方式存储, 相应的两个字段分别以 Clob, Blob 作为字段类型。这种模式的优点是结构性更好, 直观, 容易理解, 存储上也能节约空间, 但是在任何后续的读取、处理的过程中, 均需要以单独的程序对该字段进行操作, 如定位、跳跃等, 这将影响整个应用的速度。即便对于添加、读取、修改等基本数据库操作, 大对象类型数据仍然需要专门的包来操作, 增加了开发难度。

表单位式 这是最容易理解的一种方式。由于高光谱遥感数据量大, 类型繁多, 所以一种类型的数据对应一个表是最简单最容易理解的方式。既可以以某种高光谱仪器的波段为单位来建表, 也可以以对象为单位建表等, 共同点是会有一些比较固定的数据维, 或是波段数固定, 或是光谱数固定, 这样就可以根据该维来设定该表的结构。这样实施, 数据存储量不会冗余和浪费空间, 查询操作的效率高, 开发容易, 缺点是扩展性差。数据库开发不应该允许用户随着数据量的增大经常进行建表工作。所以这种方式仅限于一些特殊要求或者数据量比较固定的情况下使用。这种方式易实现开发, 画出曲线只要用 ORACLE 开发的前台工具 Developer 就可轻易的实现。

还有其他可能的存储模式, 不再赘述, 上述 3 种是其中比较典型的, 有各自的特点, 应视开发需求确定选择的方案。

4 数据结构设计实例

结合实际的数据库系统, 下面论述设计选择和实现。

4.1 针对点对象光谱数据的数据库系统

该类系统的数据一般是地面光谱仪采集的针对点对象的光谱数据。试验数据是北京农业科学院和中国科学院遥感应用研究所在 2001 年合作的小汤山精准农业项目[4] 中地面采集的光谱数据, 每个样本同时具有测量属性数据、农作物生化参量数据、图片数据和光谱数据。使用的光谱仪是美国的 ASD 光谱仪, 具有从 350—2500nm 共 2151 个波段。采用第一种——波段独立顺列式来存放。将全部数据存放于两个表, 又分别归属于 3 个表空间中, 表 1 是光谱数据表, 表名为 Wheatspectrum, 主要存放光谱数据; 表 2 是其他属性数据表, 表名为 Wheatnature, 两者通过统一的联合主关键字连接, 即 Sampleno+Sampledate

表 1 光谱数据表 **wheatspectrum** 结构

Table 1 Structure of wheatspectrum

字段	字段类型	长度	显示名称
Datano	Number	10	编号
Sampleno	Varchar2	16	样本名称
Wavelength	Number	4	波段
Wavedata	Number	11, 7	反射率值
Sampledate	Number	8	采样日期

表 2 属性数据表 **wheatnature** 结构

Table 2 Structure of wheatnature

字段	字段类型	长度及小数位	名称
Datano	Number	8	编号
Sampleno	Varchar2	16	样本名称
Sampledate	Number	8	采样日期
Instrument	Varchar2	16	测试仪器
Cloudine	Number	4, 2	云量/%
……等仪器属性参数			
LAI	Number	4, 2	叶面积指数
ltn	Number	4, 2	叶片全氮/%
……等作物生化参量属性参数			
picture	Blob		照片

两个字段, 该数据源是由采样时间和样本编号两者来确定样本唯一性的。3 个存储表空间分别对应于两张表和存储图片数据的 picture 字段。实现时存入了一天的 48 个采样点的数据, 每个样本有 350—2500nm, 共 2151 个波段数据, 即目前光谱数据表中总共有 1×48×2151= 103248 条数据。在其他属性表中, 共有 50 多个字段属性, 不再罗列。除图像数据字段外, 在仪器参数、生化参量参数方面各取了两个字段作为例子说明。

4.2 针对图像样本光谱的数据库系统

如前所述, 高光谱数据最明显的特点就是图谱合一。本数据库系统的目标就是要实现这个功能。此类系统数据源可以来自航空或者航天的图像光谱数据。试验数据采自 1999 年中国科学院遥感应用研究所在江苏常州的航空飞行数据, 仪器获取从 417—854nm 共 80 个波段的光谱数据。本系统普通属性表中的图片数据采用 650nm, 550nm, 450nm 3 个波段合成的真彩色图像, 并截取 300×300pixels 作为

本系统的数据标准。这里的数据表结构设计也采用了前面提到的数据存储规范, 即: 普通属性表组+光谱数据表组, 但是采用了第 2 种光谱数据表的存储模式——波段集中整合式: Imagespectrum 表为光谱数据表, Isnatures 表为光谱属性表, 在此表中应用了两种大对象数据类型 Clob 和 Blob, 即文本型大对象和二进制型大对象[5]。其中 imageno 相当于主关键字, Wlno 是为了系统能够有比较好的扩展性, 因为系统需求以及光谱图像可以从任意的高光谱仪器中得到的特点, 所以其波段数也是任意的, Wlno 就是用于解决这一问题。Wavelength 以文本方式存储了相应仪器的波段值, 每行一个波段, 然后再以 Clob 的形式存储于数据库中。反射率值是一个整体的形式出现的, 不像前面数据管理子系统那样分成每一波段一条记录存储。这里 Blob 字段中的光谱数据采用了 BIP 模式存储的 *.img 数据文件格式。该数据文件格式的数据结构为: 按照从第一像素点到最后一个像素点所有波段光谱反射率值以二进制的方式顺序排列在该文件中。具体的表结构见表 3 和表 4。

表 3　光谱数据表 **Imagespectrum** 结构

Table 3　Structure of Imagespectrum

字段	字段类型	长度	显示名称
Imageno	Varchar2	16	图像编号
Wlno	Number	4	波段个数
Wavelength	Clob		波段
Wavedata	Blob		反射率值

表 4　属性数据表 **Isnatures** 结构

Table 4　Structure of Isnatures

字段	字段类型	长度	名称
Datano	Number	8	编号
Imageno	Varchar2	16	图像编号
Imagedate	Number	8	日期
Instrument	Varchar2	16	测试仪器
Height	Nmber	4	离地高度/km
……等其他测量属性数据			
picture	Blob		照片

4.3　两系统采用的存储模式对比

在针对单对象光谱数据的数据库系统中, 之所以采用了顺列式存储光谱数据, 是因为在系统需求中不但需要为用户返回光谱曲线, 而且需要有光谱分析的功能, 该功能可能会需要查找任意一个采样的任意一个波段的数据, 采用这种存储方式, 每个波段设定为一条记录, 具有读取速度快, 易于查找, 结构简单, 扩展性强, 不受仪器变化, 即波段总数变化的影响等优点, 对于数据分析、处理时提取数据十分有利。同时考虑到这种存储方式的缺点: 记录条数众多, 如果系统中存放 100 天的数据, 则会有 1000 万条数据, 对于 Oracle 这种大型数据库, 仍然可以满足处理的需要; Sampleno+Sampledata 两个字段做的联合主关键字存储的冗余度比较大, 占据了一定的存储空间, 但是对于当前日益便宜的存储空间而言, 是可以接受的, 同时这种冗余可以通过在数据表上加索引(Index)、进行分区等数据库性能调整操作 x, y 两个数组绘图即可。

在针对图像光谱的数据库系统中, 之所以采用了整合式来存放光谱数据, 更多的考虑则是在数据量上。在图像光谱子系统中, 也曾经设计过使用第一种存储方式, 即顺列式来存储与图像对应的高光谱遥感数据。其设计思路如下: 假设光谱数据表为 Imagespe, 其结构如表 5。以 Imageno 为主关键字联系两个属性表, 像素编号作为冗余字段, 即第 imageno图片的第 Pixelno 点的各个波段及其值。该图像只有 80 个波段, 目前只有一副图片, 300×300pixel, 80 波段来进一步提高查询检索的效率。绘制曲线只要简单根据查询条件在普通属性表中得到主关键字, 再到光谱数据表中读取相应各字段, 即各波段数据及其反射率数据, 存入/pixel, 综上所述, 这个系统采用顺列式结构似乎利大于弊。但是计算下来总共 $1\times 90000\times 80=720\times 10^4$ 条记录。如果本系统录入 100 幅图片, 记录数就会暴增到 $7200000\times 100=7.2\times 10^8$ 条记录, 即 7 亿 2 千万条记录。如果该库中存入中国的 OMIS 高光谱成像仪, 128 个波段, 乃至更多的比如: USGS 的 HYDICE, 有 210 个波段, 那么记录数就会上 10 亿条量级了, 这即便对于 ORACLE 这样大型的数据库依然是一个很可怕的数字。所以在图像光谱数据库系统中, 采用了第 2 种波段集中整合式来存储光谱数据。这种模式的缺点是在数据处理时相比第 1 种模式, 开发难度有所增加, 相应的程序运行会影响整体效率。

表 5　假设采用波段独立顺列式的光谱数据表 **Imagespe** 结构

Table 5　Structure of supposed table Imagespe used sequence wavelength

字段	字段类型	长度	显示名称
Datano	Number	10	编号
Imageno	Varchar2	16	图像编号
Wavelength	Number	6, 2	波段
Wavedata	Number	11, 7	值
Pixelno	Number	8	像素编号

在实现用户的光谱曲线需求时的思路如下：先显示出该图片的一般属性及显示该图片，然后接收鼠标点击的位置(x, y)，按照公式

$$k=300\times x+y$$

算出该像素在光谱数据文件中的像素序号。在数据库的光谱数据字段中，以 Wlno$\times k$ 得到该像素的起始读数的位置，向后读取 Wlno 个二进制数据放入 y 数组，得到该 pixel 的反射率数值。再以读行数据的方式，从 Wavelength 字段中读取波段值存入 x 数组，以 x, y 两个数组绘图返回给用户即可[6]。

由上述对比可以看出，采用哪种存储模式完全是根据具体的系统需求而定，顺列式在数据处理方面更为灵活、简便，但是相对数据冗余量大，存储的空间需要较大；整合式数据结构更为合理、紧凑，易于理解，但是在处理中效率不如顺列式高，同时在开发时需要考虑的问题以及开发难度都要大于顺列式存储模式。所以在统一的(光谱数据表＋普通属性表)存储规范下，采用哪种模式要视具体情况而定。

5 结 论

本文提出了具有特殊性的高光谱数据存储的规范，提出一种数据库存储规范，对于今后高光谱数据的整理和统一有着重要的作用。同时，本文首次提出了在 Oracle 数据库环境下高光谱遥感数据几种存储模式，对其各自的优缺点进行对比分析，并对于其中两种加以实现。Oracle 是当前比较流行的大型数据库软件，应用比较广泛，这为今后高光谱数据在数据存储、共享方面的开发打下了技术基础。

参 考 文 献 (Refereneces)

[1] Chen S P, Tong Q X, Guo H D. Mechanism of Remote Sensing Information [M]. Beijing: Science Press, 1998[陈述彭, 童庆禧, 郭华东. 遥感信息机理研究[M]. 北京: 科学出版社, 1998.]

[2] Zhang B, Zhang X, Liu T J. Dynamic Analysis of Hyperspectral Vegetation Indices[A]. Proceedings of SPIE[C]. 2001, 4548: 32—38.

[3] Michael Abbey, Michael J. Corey, Lan Abramson. Introduction to Oracle 8i. [M]. Beijing. China Machine Press 2001. [Michael Abbey, Michael J. Corey, Lan Abramson 著, 乐嘉锦, 王兰成等译, Oracle8i 初学者指南[M]. 北京: 机械工业出版社, 2001.]

[4] Jonathan Gennick, Carol McCullough-Dieter, Gerrit-Jan Linker. Oracle8i DBA Bible [M]. Publishing House of Electrocics Industry, 2000. [Jonathan Gennick, Carol McCullough-Dieter, Gerrit-Jan Linker. 赵艳勤, 刘冠英等译. Oracle8i DBA 宝典[M]. 北京: 电子工业出版社, 2000]

[5] Compute Center of Tsinghua University. Teaching Materials of Oracle8i Course[R]. 2001. [清华大学计算中心培训部. Oracle8i 数据库课程讲义[R]. 2001.]

[6] John Zukowski. The Way to Master Java2[M]. Beijing: Publishing House of Electrocics Industry, [John Zukowski 著, 邱仲潘等译. JAVA2 从入门到精通[M]. 北京: 电子工业出版社, 1999.]

Research in Effectively Storing the Hyperspectral Data Set in Database System

ZHANG Xiong-fei, ZHANG Bing, ZHANG Xia, ZHENG Lan-fen, TONG Qing-xi

(Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing 100101)

Abstract: The hyperspectral remote sensing plays more and more important roles in Remote Sensing area. Correspondingly its application becomes wider and wider. Some new hyperspectral airborne and space borne sensors provide us more choice to use such data. Hyperspectral database system is very important in improving the research in remote sensing theory, quantitative study, and applications. This paper aims to analyze the characteristic of hyperspectral data and provide a standard storing rule for the hyperspectral data set in the relation database system with practice of database system development. The hyperspectral data set include the spatial, spectral and other properties data. Three storing models in Oracle database platform are also displayed here, and compared with the practical application. Finally their advantages, shortcomings and the choosing conditions are analyzed.

Key words: hyperspectral data; database; store; Oracle