分类号_____ 密级_____

UDC _____ 编号_____

# 中国科学院研究生院

# 博士学位论文

## 高光谱数据库及数据挖掘研究

李　兴

指导教师　童庆禧 院　士，中国科学院遥感应用研究所 北京

郑兰芬 研究员，中国科学院遥感应用研究所 北京

张　兵 研究员，中国科学院遥感应用研究所 北京

申请学位级别 理学博士 学科专业名称 地图学与地理信息系统

论文提交日期 **2006** 年 **5** 月 **6** 日　论文答辩日期 **2006** 年 **5** 月 **30** 日

培 养 单 位 中国科学院遥感应用研究所

学位授予单位 中国科学院研究生院

答辩委员会主席：_____

# 摘　要

高光谱遥感技术是指具有$10^{-2}\lambda$的光谱分辨率，在可见光到短波红外波段其光谱分辨率高达纳米数量级的遥感技术。随着航空、航天遥感器技术的迅速发展和空间信息需求的日益扩大，遥感影像数据获取越加频繁，数据量也与日俱增。高光谱遥感影像更是因为其数据量巨大而对影像数据库的发展提出了新的挑战。在海量的光谱数据和影像数据被数据库有效的管理起来之后，从这些 TB 级的数据中挖掘有用的信息，成为了高光谱遥感应用的主要研究方向之一。这些数据并不完全结构化，数据质量也不尽完好，因此数据挖掘领域的相关技术方法被应用到高光谱遥感领域，在数据的海洋之中萃取知识的精华。

本篇论文在综合描述了高光谱遥感技术和数据库技术的发展背景前提下，提出了高光谱数据库和光谱数据挖掘的内涵和外延。并以此为切入点，对国内外地面光谱数据库、遥感影像数据库的发展和研究现状、数据挖掘技术的发展和研究现状，以及空间数据挖掘和影像数据挖掘技术做了深入探讨。通过对国内外研究进展的把握和自身项目研究知识积累，针对数据挖掘和应用方向提出了高光谱数据库的系统设计，并通过导入了六千多条数据，在 ORACLE 平台上完成了技术实现。以建立的高光谱数据库为基础，对光谱数据挖掘应用和高光谱影像数据挖掘应用做了一些开创性的研究工作。

本文取得的研究进展和创新点归纳如下：

- 在构建高光谱数据库基础之上，提出了针对数据库存储方案的光谱数据模型和影像数据模型，通过将必要的高光谱分析方法和分析模型整合到数据库之中，实现了数据、方法、模型在高光谱数据库中的集成与统一。
- 将多源数据统一到一个数据库平台之中，将原有的大表结构和表群结构作转换，设计了以二元数据为核心的星型数据库概念结构，建立了通用的高光谱数据库建库模式和模型设计线路。
- 将数据库数据挖掘技术应用在光谱数据模拟和光谱参量分析研究上，从光谱数据和属性数据这两个方向在数据库中实现了高光谱数据挖掘的应用。
- 将高光谱影像数据空间映射到数据库表空间，从而实现借助于数据库平台对高光谱影像进行数据分析与信息挖掘，实现了非负矩阵分解的逆变换高光谱数据压缩、最小长度模型辅助高光谱影像波段选择、非负矩阵分解高光谱影像特征提取、支持向量机的高光谱图像目标提取等高光谱数据自动分析模型。

**关键词：**高光谱数据库 数据挖掘 数据模型 光谱 影像

# Research on Hyperspectral Database and Hyperspectral Data Mining

**LI Xing**

Directed by Prof.TONG Qingxi, Prof.ZHENG Lanfen and Prof.ZHANG Bing

## Abstract

Hyperspectral remote sensing is a cutting-edge technology of acquiring land surface information, with a high resolution in spectrum. With the fast development of spaceborne and airborne sensors, hyperspectral data are more frequently and conveniently received. As a result, data storage, management and effective information extraction are becoming key challenges to hyperspectral remote sensing science and technology. Especially, automatic or semi-automatic information extraction is a predictabley trend for hyperspectral researches and applications.

In this dissertation, the intension and extention of the concept hyperspectral database and spectral data mining has been brought forward after a background research on hyperspectral remote sensing and database. Based on this conception, ground-object spectra database, remote sensing image database, data mining technology have been studied, while spatial data mining and image data mining technology are focused. A conceptual design with data mining and application oriented is provided basetd upon project experience and international study. The hyperspectral database has been built up and more than 6,000 sets of spectral data have been uploaded. There are spectra of rocks, minerals, waters, concretes, trees, wheats, soils and hyperspectral images in this database. With this hyperspectral database, some researches on spectral data mining and hyperspectral image data mining have been carried out. In view of hyperspectral remote sensing technology and application, there are some advangates of this dissertation as follows:

1. Based on hyperspectral database, new storage models of spectra and hyperspectral image has been put forward. A new design has combined data, methods and models in a whole database platform.
2. To integrage different sources of spectra in one database platform, a new conceptual data structure, dualistic-core star structure, has been brought forward.
3. Forward (from spectra to attributes) and backward researches have been impleted based on hyperspectral database. Automatic optimizations have been applied in band combination and band selection.
4. From a totally different view of hyperspectral image, a projection from hyperspectral images to relation tables has been built up to improve analysis and information extraction. The following key techniques are implemented in hyeprspectral database: Band selection based on Minium Length Model, Feature extraction based on Non-negative Matrix Fraction, Data compression and decompression based on Non-negative Matrix Fraction, Object Decection based on Support Vector Machine.

**Key Words:** Hyperspectral Database, Data Mining, Data Model, Spectrum, Image

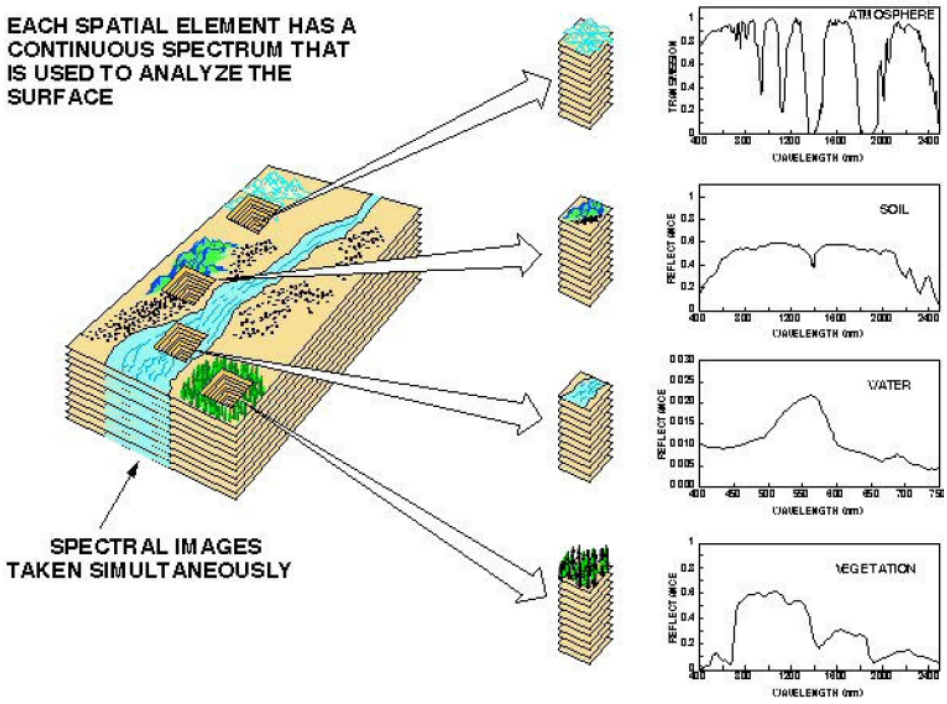
# 目 录

# 第一章 引言

## 1.1 高光谱遥感技术，数据库技术，数据挖掘技术的发展

### 1.1.1 高光谱遥感技术的发展

高光谱遥感技术（Hyperspectral Remote Sensing）是指具有 10-2λ的光谱分辨率，在可见光到短波红外波段的光谱分辨率高达纳米（nm）数量级的遥感技术。高光谱遥感通常具有波段多的特点，光谱通道数多达数十甚至数百个以上，而且各光谱通道间往往是连续的，因此高光谱又通常被称为成像光谱（Imaging Spectrometry）（童庆禧，1990,1999）。高光谱遥感影像的最显著特点在于每一个像元都可以提取出一条完整的连续的光谱曲线（Goetz, 1981,1985），如图 1.1，其光谱范围可以覆盖可见光、近红外和中红外。高光谱遥感技术的发展是过去二十多年中人类在对地观测方面所取得重大技术突破之一，是当前遥感的前沿技术（陈述彭，童庆禧等，1998）。




图表 1.1 高光谱遥感图像内涵

高光谱遥感起源于二十世纪八十年代。在地质研究领域，Hunt 在归纳各种地物光谱的基础上提出，如果能实现连续的窄波段成像，则地面矿物的直接识别就有可能实现（Hunt, 1980）； Goetz 利用航天飞机多光谱红外辐射仪（SMIRR）第一次以遥感方式实现了从空中直接鉴别粘土与碳酸岩矿物，并在美国国家宇航局（NASA）的喷气推进实验室（JPL）开始了成像光谱仪概念设计与研究计划（Goetz, 1982）。之后，在 1983 年，世界上第一台成像光谱仪 AIS-1 在美国喷气推进试验室研制成功，

并在矿物填图、植被化学成分、水色及大气的水分等方面进行试验应用并获得成功。之后，AIS-2、美国机载可见红外成像光谱仪（AVIRIS）、加拿大的荧光线成像光谱仪（FLI）、美国 Deadalus 公司的 MIVIS，GER 公司的 79 通道机载成像光谱仪（DAIS－7915）、芬兰的机载多用成像光谱仪（DAISA）、德国的反射式成像光谱仪（ROSIS－10 和 22）、美国海军研究所实验室的超光谱数字图像采集试验（HYDICE）、澳大利亚的 HyMap、美国的 Probe、加拿大 ITRES 公司的系列产品、以及由美国 GER 公司为德士古（TEXACO）石油公司专门研制的 TEEMS 系统等等（童庆禧，2003）。我国研制的新型模块化航空成像光谱仪 MAIS、推扫式成像光谱成像仪（PHI）和实用型模块成像光谱仪系统（OMIS）代表了亚洲高光谱遥感器的水平。

这些先后研制的高光谱遥感器性能逐渐稳定，探测效率逐步提高，在综合应用方面都获得了较大的成功。

F. A. Kruse, J. W. Boardman 和 J. H. Huntington 在美国内华达地区获取的 AVIRIS 数据进行了矿物填图（Kruse, 2003）。欧盟的 MINEO 计划利用高光谱技术对矿区进行环境影响评价与检测，联合英国、德国、葡萄牙、奥地利、芬兰五国在六个矿区建立试点，在 2000 年 1 月至 2003 年 6 月对矿区污染源、污染物搬运路线、污染症状填图、矿物风化强度填图、采矿影响的植被多样性和植被压力检测、污染矿物填图等等进行研究（CHEVREL S.，2001，2002）。Foudan Salem 等人通过高光谱影像对石油泄漏探测进行了系统研究（Foudan， 2002）。加拿大空间局利用 casi 数据对 Noranda 地区也进行了铜矿矿区识别研究，他们利用几种关键的变质矿物进行矿区潜力制图（Canadian Space Agency）。Ceccato 等（2001）利用 1600nm 和 820nm 波段的反射率比估算单位叶面积上的水分含量。Lênio 等利用 EO-1 上搭载的 Hyperion 光谱仪获得数据对巴西东南区的蔗糖种类进行精划分（Lênio et al，2005）。Martin 等结合不同森林树种之间特有的生化特性和已经在高光谱数据 AVIRIS 和簇叶化学成分之间建立的关系鉴别 11 种森林类型（Martin，1998）。Franklin 和 McDermid 运用逐步回归方法从 CASI 数据中选择三波段回归模型以估计平均树高，确定系数 $R^2$ 高达 0.67（Franklin，2003）。Goodenough 等（2003）引入了多角度信息，进行了树种的分类，并且用高光谱遥感器 HYPERION 数据、ALI 数据和 ETM+数据对同一个地区进行了分类。

童庆禧等基于 MAIS 数据，经反射率定标，采用导数光谱波形匹配模型和光谱夹角填图方法，成功地对鄱阳湖湿地生态系统进行了分类（童庆禧等，1997）。张兵利用光谱数据海量数据的特点，提出特征优化的专家决策分类方法，完成了基于 PHI 高光谱数据的日本南牧农作物精细分类（张兵，2002）。王晋年利用 GERIS 数据在新疆阿克苏西部进行了矿物光谱识别、填图研究；利用 MAIS 数据在澳大利亚松谷铀矿区发现了铀矿的可能存在区（王晋年，1996）。宫鹏等利用成像光谱仪实测的光谱数据来识别美国加州的 6 种主要针叶树种（宫鹏，1998）。甘甫平等利用航天 Hyperion 高光谱数据在西藏驱龙地区识别出与斑岩铜矿密切相关的蚀变矿物组合，

并发现了矿化异常和若干较小的蚀变分布区，同时，他们也对德兴矿区污染进行探测（甘甫平，2000，2004）。

高光谱遥感正是随着传感器技术发展与遥感应用拓展不断相互促进的良性循环，不断发展壮大。光谱仪的观测平台由实验室、野外、高台发展到航空、航天；观测数据由单点光谱曲线向面阵像元波谱过渡；观测频率也由根据需求不定期向定时获取区域数据转变；高光谱遥感的应用也由定性转为定量，由实验研究向产业化运作转变；高光谱数据的处理手段也逐步软件化，高光谱数据的存储与管理也由文件系统管理逐步转为数据库系统管理。

高光谱遥感当前仍处在国际遥感科技发展的前沿，是人们关注的焦点之一。（童庆禧，2003）

### 1.1.2 数据库技术的发展

高光谱遥感最显著的特点就是海量的数据，而高光谱遥感技术的应用需要一系列的配套数据。高光谱遥感相对其他遥感手段而言，波段数量呈几倍、十几倍的增长；同时随着传感器的量化级数增加和几何分辨率的提升，同一地区的高光谱遥感影像的数据量往往是其他观测手段的几十倍。与此庞大的数据同等重要的是遥感影像数据获取时的配套参数，包括定标数据、地物光谱数据、控制点数据等等。高光谱遥感的特点和应用需求对其数据存储和管理提出了更高的要求。数据库技术的发展能够为高光谱遥感技术提供良好的应用平台。

数据库技术是研究数据库的结构、存储、设计和使用的一门软件科学，是数据管理和处理的最新技术。数据库技术在当今信息管理和处理中起着重要的作用，从某种意义上讲，数据库建设的规模，数据库信息的质量和数量，数据库的使用程度，是衡量一个国家信息化程度的标志。

以数据模型的进展作为主要的依据和标志，数据库发展阶段的划分为三段：

1) 第一代数据库系统：层次/网状数据库系统

第一代数据库的代表是 1968 年 IBM 公司研制的层次模型的数据库管理系统 IMS（IBM, 1968）和 70 年代美国数据库系统语言协商 CODASYL（Conference On Data System Language）下属数据库任务组 DBTG（Data Base Task Group）提议的网状模型（CODASYL， 1962，1969，1971）。

层次数据库的数据模型是有根的定向有序树，网状模型对应的是有向图。这两种数据库奠定了现代数据库发展的基础。这两种数据库具有如下共同点：支持三级模式（外模式、模式、内模式）。保证数据库系统具有数据与程序的物理独立性和一定的逻辑独立性；用存取路径来表示数据之间的联系；有独立的数据定义语言；导航式的数据操纵语言。

2) 第二代数据库系统：关系数据库系统

1970 年 IBM 公司 San Jose 研究室的研究员 E.F.Codd 博士首次提出了关系型数据库的数据模型和理论（E.F.Codd，1970），开创了数据库关系方法和关系数据理论研究，为关系数据库技术奠定了理论基础。70 年代时关系数据库理论研究和原型开发的年代，其中以 IBM 公司的 System R（MM Astrahan, 1976）和 Berkeley 大学研制的 INGRES（Michael Stonebraker，1976）为典型代表。经过大量的高层次的研究和开发取得了一系列的成果：

- 奠定了关系模型的理论基础，给出人们一致能接受的关系模型规范说明。
- 研究了关系数据语言，有关系代数、关系演算、SQL 语言、QBE 等等
- 研制了大量的 RDBMS 的原型，攻克了系统实现中查询优化、并发控制、故障恢复等一系列关键技术。

关系数据模型（数据结构、关系操作和数据完整性）：关系模型的概念单一，实体以及实体之间的联系都用关系来表示；以关系代数为基础，数据形式化基础好；数据独立性强，数据的物理存储和存取路径对用户隐蔽；关系数据库语言是非过程化的，减少用户编程的难度。

80 年代是 RDBMS 产品发展和竞争的时代。RDBMS 产品经历了从集中到分布，从单机环境到网络，从支持信息管理、辅助决策到联机事务处理的发展过程，对关系模型的支持也逐步完善，系统的功能也不断增强。RDMBS 的实现技术已研究得十分透彻。我们已经知道如何在外部存储设备上存储数据、如何使用各种复杂的存取方法、缓冲策略和索引技术访问外部存储设备上的数据。数据库恢复、并发控制、事务管理、完整性和安全性的实施、查询处理、优化等技术也已经得到深入了解，并在大多数商用的集中式和分布式 RDBMS 中得已实现。

3) 第三代数据库：以面向对象为主要特征的数据库系统

自 1990 年高级 DBMS 功能委员会（The Committee for Advanced DBMS Function）发表“第三代数据库系统之言”（Third-Generation Database System Manifesto）一文以来（The Committee for Advanced DBMS Function，1990），经过几年来的研究和讨论，对第三代数据库系统的基本特征已有三点共识：

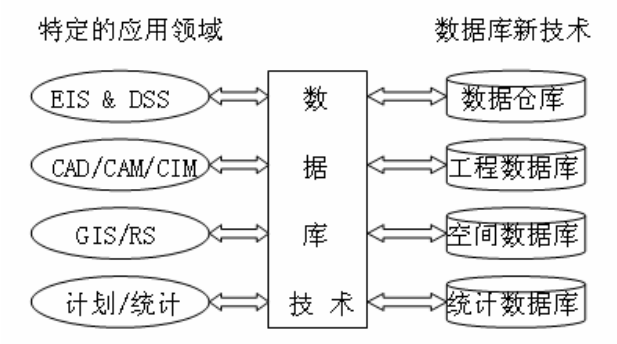
第三代数据库系统应支持数据管理、对象管理和知识管理。

第三代数据库系统必须保持或继承第二代数据库系统的技术，如非过程化数据存储方式和数据独立性。

第三代数据库系统必须对其它系统开放：支持数据库标准语言，支持标准网络协议，系统应具有良好的可移植性，可连接性，可扩展性和可互操作性等的高性能特征。

第三代数据库支持多种数据模型（比如关系模型和面向对象的模型），并和诸多新技术相结合（比如分布处理技术、并行计算技术、人工智能技术、多媒体技术、模糊技术），广泛应用于多个领域（商业管理、GIS、计划统计等），由此也衍生出多种新的数据库技术。

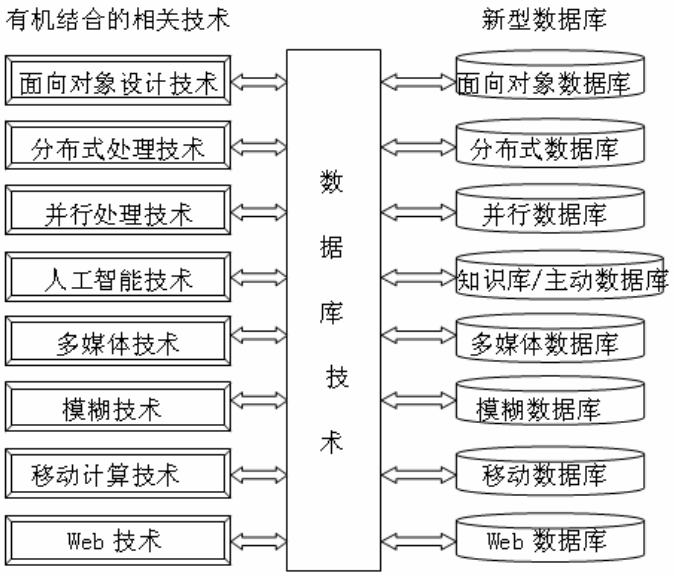
数据库技术在商业领域中的巨大成功，极大促进了其他领域对数据库技术的需求，而随着网络应用的丰富和深入，数据对象从简单的数字、字符等文本类型向诸如图形、图像、音视频、Web 网页等各种类型和事件拓展，特定的应用领域对数据库技术提出了新的要求和挑战。如图 1.2 所示，数据库技术与其它相关应用领域的结合，直接推动了数据库技术的研究和发展。




图表 1.2 特定应用领域中的数据库新技术

数据库技术的应用领域主要包括：计算机辅助设计和计算机辅助制造（CAD/CAM）、计算机集成制造（CIM）、计算机辅助软件工程（CASE）、办公信息系统（OIS）、地理信息系统（GIS）、知识库系统、专家系统、决策支持系统、实时信息处理系统等等。这些新的应用领域都需要数据库的支持，但它们的数据管理功能和性能中有相当一部分是传统的数据库系统所不能支持的。

计算机领域中其它新兴技术的发展对数据库技术产生了重大影响。传统的数据库技术和其它计算机技术的互相结合、互相渗透（如图 1.3），使数据库中新的技术内容层出不究。如分布式数据库、并行数据库、演绎数据库、知识库、多媒体数据库等等，它们共同构成了数据库大家族。




图表 1.3 数据库技术与相关技术的结合

为了适应数据库应用多元化的要求，研究适合某种应用领域的数据库技术，产

生了面向特定应用领域的数据库，如工程数据库、统计数据库、科学数据库、地理数据库、遥感影像数据库等，这是数据库技术发展的又一重要特征。研究和开发面向专门应用领域的数据库系统的基本方法是以传统数据库技术为基础，针对专门应用领域的数据对象的特点，建立特定的数据模型，它们或者是关系模型的扩展和修改，或者是具有某些面向对象特征的数据模型。

Internet Web、自然科学研究和电子商务领域，现在已经成为了信息与信息处理的巨大源泉，而其中自然科学研究产生的大量的复杂的数据集，需要安全和信息集成的有力手段，需要对有序数据进行检索与查询（如：时间序列、图像分析、网格计算和地理信息等）。通过对信息集成、数据流管理、传感器数据库技术、半结构化数据与 XML 数据管理、网格数据管理、DBMS 自适应管理、移动数据管理、微小型数据库、数据库用户界面等这些目前的数据库热点问题的研究和探讨，数据库技术发展体现在不断的与新技术和新应用融合（孟小峰等，2004）。Jim Gray 在 SIGMOD2004 年会的主题发言中提到，数据库体系结构面临革命性变革.新的应用和需要将促使这一变革的到来（Jim Gray, 2004）。

### 1.1.3 高光谱数据库概念的提出

高光谱遥感技术与应用的发展对其数据存储与管理提出了新的要求，而数据库技术在面对新领域的需求时，需要做出新的变化。高光谱数据有着其与众不同的特点，所以决定了在数据库开发方面也有其与众不同的特点，高光谱数据库的概念便是由此而来。高光谱数据库系统是专门面向高光谱数据，体现图谱合一特性，综合了光谱数据库、光谱分析功能和数据挖掘功能于一体的专用数据库系统（张雄飞，2004）。它与通常意义上的光谱数据库或者影像数据库不同在于：高光谱数据库不仅存储室内或野外光谱辐射计获取的目标光谱数据，它还存储以图像立方体形式存在的高光谱图像光谱数据。它能够将高光谱遥感应用的图像、光谱有机的融合起来，形成对高光谱影像的综合分析与应用。

高光谱数据库的核心是空间数据库技术和图像数据库技术。

空间数据库是描述、存储和处理空间数据及其属性数据的数据库系统。空间数据是用于表示空间物体的位置、形状、大小和分布特征等诸方面信息的数据，适用于描述所有二维、三维和多维分布的关于区域的现象。属性数据为非空间数据，用于描述空间物体的性质，对空间物体进行语义定义。空间数据库的主要功能是提供对空间数据和空间关系的定义和描述；提供空间数据查询语言，实现对空间数据的高效查询和操作；提供对空间数据的存储和组织；提供对空间数据的直观显示等。它涉及到计算机科学、地理学、地图制图学、摄影测量、遥感等多门学科，在数据查询、操作、存储和显示等方面比较复杂。以空间数据库为核心的地理信息系统应用已经从解决道路、输电线路等基础设施的规划和管理发展到更加复杂的领域，已

经广泛应用于环境和资源管理、土地利用、城市规划、森林保护、人口调查、交通、商业网络等各个方面的管理与决策。

图像数据库是这样一个系统，它能够将一大批图像及其有关信息存储在一起，并对他们进行有效的管理，保证数据的一致性、完整性，支持各种应用。根据图像的特殊性，一般将图像数据库结构分为五级模式：用户视图、语义特征视图、图像特征视图、特征表示、特征组织。图像数据库包含了对图像信息检索、图像解释以及图像信息的识别与处理的支持，因此，图像数据模型的建立便显得重要，一般包括：图像数据存储结构，图像特征描述（图像实体的形状、颜色、纹理和空间关系等），和辅助信息（图像说明以及各种相关因素）等等（杨宇艇，潘云鹤，1996）。图像数据库已经广泛应用于医学教学与研究、遥感应用、人脸识别等众多领域。

高光谱数据库是将高光谱遥感应用的数据进行集成整合，通过空间数据库技术和图像数据库技术来实现对高光谱遥感数据的描述、存储和处理。它涉及到高光谱图像数据建模、数据集成、图像显示、分析应用等各方面知识。高光谱数据的特殊性以及其应用的广泛性，使得高光谱数据库的建设方兴未艾。

### 1.1.4 光谱数据挖掘的内涵与外延

数据库技术虽然可以高效地实现海量数据的录入、修改、统计、查询等功能，但无法发现数据中存在的关系和规则，无法理解数据中包含的信息,无法根据现有的数据预测未来的发展趋势。缺乏挖掘数据背后隐藏的知识的手段，导致了“数据爆炸但知识贫乏”的现象。数据挖掘是从数据仓库中提取出可信的、新颖的、有效的并能被人理解的模式的高级处理过程。所谓模式,可以看作是我们所说的知识,它给出了数据的特性或数据之间的关系,是对数据包含的信息更抽象的描述（Han, 2001）。

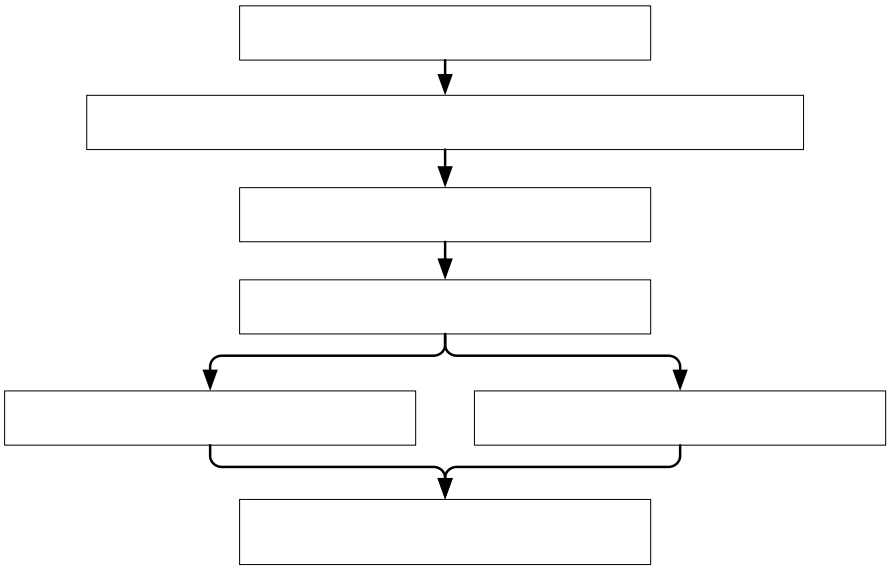
高光谱遥感技术的应用离不开光谱数据挖掘。高光谱数据挖掘涉及三个层次的内容，一是高光谱遥感的理论与技术基础，它涉及电磁波理论与成像光谱遥感的技术基础。二是高光谱数据的定量化和参量化、以及在此基础上的高光谱图像分类与地物识别。三是多源数据的理论化与模型化结合，以获取理想的结果（张兵，2003）。由于光谱对地物化学成分和结构的微细变化非常敏感，地物微细的化学成分和结构的变化常常导致吸收位置和吸收形态的变化。因此，地物的光谱特性在现实世界中是非常复杂的。在早期研究中，这成为一大劣势。随着技术的发展和对地物光谱特性和迁移原因的深入探究及知识的增加，这一劣势正转变为探究地物化学成分和结构及自然环境微细变化的强大优势（燕守勋，张兵等，2003）。光谱数据挖掘，其含义就是充分利用精细的光谱信息，发掘地物光谱曲线中微妙的变化并进行特征化、参量化，去粗取精，去伪存真，从而实现对地物的分类与识别。而光谱的含义，随着数据获取手段的发展，已经由地面测量光谱，延伸到图像光谱，时间序列谱，光谱指数谱等范畴，光谱数据挖掘的对象不再局限于地面测量光谱。光谱数据挖掘的

对象和方法随着高光谱遥感的应用而不断扩展外延。

## 1.2 研究重点和论文结构

本论文的研究重点是高光谱数据库的设计与构建、以及高光谱数据挖掘，主要研究了高光谱数据库中的光谱数据模型、影像数据模型，并对面向应用和数据挖掘的高光谱数据库设计进行了探讨，并利用数据挖掘技术对光谱数据及相关属性数据进行初步的研究探讨。

本论文论文结构如图 1.4 所示，全文共分七章。第一章引言主要描述了整个研究的技术背景发展，包括高光谱遥感技术的发展、数据库技术的应用与发展以及数据挖掘技术的发展，从而对高光谱数据库概念作了阐述，并对光谱数据挖掘的定义作了简单的解释。第二章主要论述国内外地面光谱数据库、遥感影像数据库的发展和研究现状、数据挖掘技术的发展和研究现状，并对空间数据挖掘和影像数据挖掘技术的作了探讨。第三章对高光谱数据库的设计作了全面深入的研究，在研究高光谱数据库的特点基础之上，提出了光谱数据模型和高光谱影像数据模型，利用面向对象的数据库设计方法，对高光谱数据库的应用框架、系统结构、应用接口、应用逻辑以及模型方法进行了详细的设计，并提出了基于高光谱数据库的通用模型和方法设计。第四章在数据库设计基础之上，对数据库系统建设进行了详细的论述，包括数据库系统物理结构、数据、典型方法、典型应用模型等等。第五章主要研究了在高光谱数据库系统基础之上，数据挖掘技术在地物光谱中的应用，包括岩矿光谱数据的模拟、光谱参量和理化参数的相关关系挖掘、光谱波段组合的自动优化等。第六章在高光谱数据库基础上，利用数据挖掘技术对高光谱影像数据进行了信息挖掘研究，包括利用最小描述长度模型来对高光谱影像数据进行波段选择、利用非负矩阵分解进行特征提取和数据压缩，利用支持向量机对高光谱影像进行目标提取等。



图表 1.4 论文结构图

## 1.3 本文使用的数据源描述

本文在高光谱数据库建设中，主要集成了如下数据：

- 863 波谱库岩石矿物数据（863 波谱库项目、中科院遥感所高光谱室）
- 小麦光谱曲线和对应生化参量（北京市自然基金项目、中科院遥感所高光谱室）
- 城市地物光谱数据（中科院遥感所高光谱室历史数据积累）
- 膨胀土光谱和对应理化参量（数据提供：燕守勋）
- MODIS 2004 年 23 波段 AVI 影像数据（数据提供：张霞）
- 太湖地区水质光谱和叶绿素 a 数据（数据提供：李俊生）

# 第二章 地物光谱数据库、遥感影像数据库及数据挖掘

在高光谱遥感应用中，图像分类、信息提取和目标识别是高光谱遥感的根本优势，也是所有高光谱遥感应用研究者在数据处理和分析中的聚焦点（童庆禧，2003）。在高光谱图像的分类与识别中，基于物性，即基于地物的光谱反射或发射曲线的分类识别方法最具特色（Jimenez L O， 1999）。这种方法就是利用光谱库中已知的光谱数据，采用匹配的算法来识别地物。因此，完整的光谱数据采集和配套参数收集，并通过数据库进行管理和发布成为了遥感应用的重要基础。而光谱数据库也在众多领域成为研究的重要组成，例如地质填图（Clark，1990），植被分析（Li et al.，1999），海洋水色研究（Barnard et al.，1999）等等。

在随着航空、航天遥感器的日益发展，遥感影像数据获取越加频繁，数据量也与日俱增。除了各个卫星地面站每日接收的原始数据之外，遥感影像数据随着图像处理的手段和阶段不同，也产生了各种级别的数据产品，对于这些海量的遥感影像数据的管理成为了遥感影像数据库的重要任务。而高光谱遥感影像更是因为其数据量巨大而对影像数据库的发展提出了新的挑战。

在海量的光谱数据和影像数据被数据库有效的管理起来之后，从这些 TB 级的数据中挖掘有用的信息，成为了高光谱遥感应用的主要研究方向之一。这些数据并不完全结构化，数据质量也不尽完好，因此数据挖掘领域的相关技术方法被应用到高光谱遥感领域，在数据的海洋之中萃取知识的精华。

本章将主要介绍国内外地面光谱数据库的发展，国内外遥感影像数据库的研究与进展，以及数据库技术和数据挖掘技术的研究现状和发展方向，并对空间数据挖掘技术和影像数据挖掘技术作详细的介绍。

## 2.1 地物光谱数据库

收集和积累各种地物的光谱数据信息历来是遥感基础研究和应用研究中不可缺少的重要环节，它对发展遥感信息处理的新方法、提高遥感分类识别水平起着非常重要的作用。随着高光谱遥感技术的发展和应用，光谱数据的剧增对地物光谱数据库的建设提出了更高的要求。建立地物光谱数据库、运用数据库技术来保存、管理和分析这些信息，是遥感定量化研究不断深入的重要表现。

地物光谱数据库，主要是指对地面光谱仪采集的地面物体目标光谱数据以及配套的相关参数数据进行存储、管理、显示和检索的数据库系统。这在高光谱定量遥感中扮演着十分重要的角色。由于地面光谱测量数据能够在电磁波紫外到近红外（300-2500nm）的太阳反射波谱段获取地物连续的光谱曲线，它对高光谱遥感数据分析的支持作用能够体现在以下几个方面：

- 地面光谱仪在遥感器过顶时间可以同步获取下行太阳辐射，以用于机上或星上传感器定标

- 在经验线性法反射率转换中，通过地面点光谱测量完成 DN 值图像到反射率图像的转换
- 通过地面光谱仪测量数据，研究某一地物被高光谱遥感探测的可能性
- 获取对特定地物最佳遥感效果的探测时间、最小空间分辨率、信噪比等信息
- 用于图像识别目的目标光谱数据获取和特征建立

除辅助遥感应用之外，地面光谱数据能够直接辅助行业应用。例如在地面地质矿物识别方面，利用矿物特殊的光谱吸收特点，通过对测量光谱分析，可以直接实现地面矿物和矿物集合的识别。这比从野外取回样本做实验室理化分析要更加有效率，并且更加经济实用。在环境监测领域中，也可以直接利用光谱数据反演叶绿素含量、悬浮颗粒物等等各种成分含量，能够迅速快捷的对水质进行监控，这比传统的水体取样后做化学分析要更加简捷，并且具有更高的时效性。

由于地物光谱数据库在遥感和各行业中具有不可替代的作用，因此，从地物光谱库诞生之时到现在，地物光谱数据库的建设一直在不断持续、不断发展。

### 2.2.1 国外地物光谱数据库研究进展

地物光谱数据库起源于地物光谱特性的研究。

地物波谱特性研究可以追溯到三四十年代，当时苏联对 370 种地物的可见光光谱进行测量，1947 年出版了国际上第一部地物光谱反射特性的专著，《自然物体的光谱反射特征》，书中包括植被、土壤、岩矿、水体 4 大地物的光谱反射特性，是研制各类专用胶片发展航空摄影遥感的主要参考书，之后前苏联科学家又进行了许多基础性测量研究。

六十年代，美国的密执安大学等开始进行大规模的地物波谱特性测量，从遥感器通道设置合理性和遥感器性能参数等方面对陆地卫星计划的可行性进行研究，地物波谱测量的波段扩展到了中红外甚至微波，并进行了一系列卫星遥感器航空样机的飞行试验，至 1971 年已在全国建立了 289 个试验场，进行遥感器应用评价和辐射校正，其中的白沙导弹靶场直到现在一直是美国和欧空局的重要辐射校正场。

美国NASA在60年代末到70年代初建立了地球资源信息系统（The NASA Earth Resources Spectral Information System， ERSIS），共包括植被、土壤、岩石矿物和水体等四大类地物的电磁波波谱特性数据。

八十年代后期，在美国 USGS 牵头十几个国家参与的国际地质比对计划中，从光谱仪、定标、测量规范、数据库结构与格式，到光谱特性与地质的关系分析，专门就地质光谱特性进行了比较全面的研究。USGS 对各种主要岩石类型和部分植被类型进行了比较系统的光谱测量，测量中除了采用实验室及野外地面光谱测量方法外，还采用了遥感光谱学（Remote Sensing Spectroscopy）的测量方法，即利用航空

高光谱成像的测量方法测量地物目标的光谱特征，并制成了光谱数据库。现在的光谱库版本是 splib04， 包含近 500 条特征矿物与典型植被光谱数据，覆盖波谱范围为 0。4-3。0μm。美国 USGS 正在进一步丰富其波谱库内容，增加更多的矿物、混合矿物、植被和人造材料波谱，并计划将波谱覆盖范围扩展到 150μm。由于波谱库建设耗费巨大人力、财力，USGS 下一个波谱库版本的推出日期还没有确定。

美国喷气推进实验室（JPL）对矿物的反射光谱进行了实验室测量研究，建立了 ASTER 光谱库，其中 135 种矿物提供三种不同粒径的反射谱，以区分粒子尺寸对光谱的影响。他们采用 Beckman5240 光谱仪测量。包括 160 种矿物岩石在 125—500 微米，45—125 微米，小于 45 微米三种微粒尺度下的光谱，以研究微粒尺度与光谱之间的关系。光谱波段宽度在 400—800nm 之间为 1—4nm，800—2200nm 之间小于 20nm，2200—2500nm 之间为 20—40nm。除光谱数据外，JPL 还规范了样品采集、样品纯度和组份分析方法。（Grove，1992）

九十年代，美国 Johns Hopkins 大学（JHU）建立了包括岩石（火成岩，变形岩，沉积岩）、矿物、地球土壤、月球土壤、人工材料、陨石、植被、水体、雪和冰、以及人工目标的波谱数据库。其中，矿物和陨石采用双向反射波谱测量，采用 Beckman 和 FTIR 光谱仪测量，波谱覆盖范围为 2.08-25 um，其它大都采取半球反射测量，波谱覆盖范围略有不同，但大致在 0.3-15 um 范围内。2-25μm 热红外及植被波谱库中，用户可查看、建立、重采样标准波谱库和自己的波谱库，从而使用户可进行物质成份、热红外分析和植被分析。（Korb，1996；Salisbury，1994）。

由美国 IGCP—264 项目于 1990 年收集建立 IGCP—264 光谱库，包括由 5 种光谱仪测量所得到的 5 个光谱库，它们是：

科罗拉多大学空间对地研究中心（SCES）采用改制的 Beckman5270 双光路反射光谱仪测量的光谱。光谱分辨率为 3.8nm，重采样成 1nm 分辨率。

SCES 采用 GER 公司 SIRIS 便携式野外光谱仪测量的光谱。SIRIS 是单光路三个光栅的光谱仪，第一个光栅波长范围为 350nm—1080nm；第二个为 1080nm—2500nm；第三个为 1800nm—2500nm。

SCES 采用 PIMAⅡ野外光谱仪在实验室条件下测得的光谱，光谱分辨率约为 2.5nm。

布朗大学采用 Relab 光谱仪测量的光谱。光谱分辨率为 2—13nm，在 400—2500nm 范围内重采样成 5nm。

USGS 丹佛光谱实验室采用计算机控制的 Beckman 光谱仪测量的光谱。光谱分辨率在可见光范围为 0.2nm，在近红外为 0.5nm。

此外，有些国家和单位还结合特定的应用研究开展了某些地物光谱特性测试与数据库的建立工作。美国环保局和空军部门针对大气污染和空气成分的诊断建立了 AEDC／EPA 光谱数据库；英国 90 年代初针对海水颜色研究建立了海水光谱数据库，以研究海水光谱分析模型；澳大利亚的 CSIRO 建立的高光谱分辨率地物光谱数据

库；美国的基于 HYDICE（Hyperspectral Digital Imagery Collection Experiment）超光谱传感器的森林高光谱数据库；NIST（National Institute of Standards and Technology）发展的有害气体污染物质的标准定量化光谱数据库等；日本的地质调查所主持的 IGCP-264 计划也期望弄清楚光谱与地质体及其蚀变、土壤、植被之间的对应关系，建立起标准化的目标光谱数据库，用于谋求建立新的分类和识别判据技术。其光谱范围遍及紫外到热红外和微波波段的反射谱，辐射、荧光也都是研究范围。

目前主要的遥感图像处理软件，都具有高光谱数据处理功能模块并挂接了国际上比较通用的光谱库。例如 ENVI 中拥有波谱库管理、编辑及分析模块。它包含了美国地质调查局的 USGS 光谱库、喷气推进实验室的 JPL 标准物质成份波谱库、John Hopking 大学的光谱库以及 1990 年作为 IGCP264 计划的一部分收集的光谱，从而使用户可进行物质成份、热红外分析和植被分析。在 PCI 软件的高光谱分析（Hyperspectral Data Analysis）模块中也提供了基于 USGS 光谱库发展的高光谱地物库，同时提供用户各种光谱分析能力，自动地物判识能力（根据光谱特点）。ERDAS 软件中提供了 USGS 的 500 种地物光谱和喷气推进实验室（JPL）的 160 种矿物波谱（0.4-2.5μm）。ERMaper 中也同样挂接了 USGS 光谱库。这些光谱库与相应软件模块在地质、水文、海洋、大气科学中都发挥了巨大作用。例如在地质科学中，用光谱库来定义各种矿物和材料类型，然后利用光谱角映射分类法对像素进行分类，即可探测特定矿物。

这些都充分显示出地物光谱数据库在遥感发展中不可或缺的基础性地位。

### 2.2.2 国内地物光谱数据库研究进展

我国地物波谱测量研究可以从七十年代末的腾冲航空遥感试验算起，在利用直升飞机进行遥感飞行试验的同时，也进行了地物波谱测量工作。

八十年代，中科院长春光机所和长春地理所等单位建立了长春净月坛试验场，场内地物反映了东北地区的自然地理特点，值得一提的有一个可以从多个角度测量样品对太阳模拟光反射情况的室内物理模拟测量装置。

水利部遥感中心在湖南建立了洞庭湖试验场，洞庭湖是我国第二大淡水湖，周围河网交错，这一带是我国重要的粮食基地，易受洪灾。为了发展我国资源卫星，在江苏安徽交界的宁芜地区建立了遥感试验场。

1982 年，由中科院空间科学技术中心主持，十多个研究所参加，制定了地物波谱测试规范，获得了该地区岩矿、水体、土壤、植被及农作物的 1 千多条光谱曲线，出版了《中国地球资源光谱信息资料汇编》一书。

“七五”期间，国家攻关项目“高空遥感实用系统”的子课题支持在全国范围建立了 13 个遥感基础实验场，全面规范了典型地物波谱的收集和分析方法，在中科院安徽光机所和中科院遥感所等众多单位的共同努力下，收集并建立了全面包括植被、

土壤、岩矿、水体和人工目标五大类地物、300 余种约 15000 条光谱组成的地物光谱数据库。数据库中除野外测量光谱外，还有选择地收集了一部分 380～2500nm 的室内光谱数据及 400～1100nm 的航空光谱数据，并存放了相应的环境参数、大气参数及理化参数。最后由中科院安徽光机所主持汇总建立了具有 15000 余条标准化数据的《全国地物波谱特性数据库》。

此外，中科院遥感所八十年代末出版了《中国典型地物波谱及其特征分析》一书，其中给出了 173 种植物、31 种土壤、66 种岩石、7 种水体，共计 277 种中国典型地物波谱特征。

进入九十年代，在国防科工委和国家计委的领导下，国家卫星气象中心牵头，资源卫星应用中心、中科院上海技术物理所、安徽光机所等十多家单位参与就中国遥感卫星辐射校正场的建立进行了一系列的研究，包括地物波谱测量。

1994 年进行了辐射校正场选场考察，对敦煌、格尔木和青海湖预选场进行了光谱测量和相关信息收集，为确定敦煌作为遥感卫星可见-近红外波段辐射校正场、青海湖作为热红外波段辐射校正场提供了依据。

1996 年、1999 年在敦煌、青海湖试验场进行了大量的光谱测量，为进行场地光学特性分析和评价积累了大量数据。

1999 年、2000 年在敦煌、青海湖试验场进行了 FY-1C、FY2B 气象卫星遥感器在轨辐射定标同步观测，获得了高光谱和遥感器通道场地测量数据（光谱覆盖可见-热红外）。

九五期间，作为国防科工委卫星应用重点项目的子项目“典型下垫面辐射光谱特性数据库”，中科院上海技术物理所在国家卫星气象中心、中科院安徽光机所和中科院遥感所等单位工作基础上，建立了基于 Windows 界面的地物波谱数据库，具有图像、曲线和数据的分析和交互功能，从数据源上看，特别增加了超光谱地物波谱仪获得的不同水体的光谱测量数据（并配有采样分析数据和 GPS 等数据）以及神舟飞船中分辨率成像光谱仪样机和 863 计划 128 波段航空成像光谱仪获得的云等的波谱数据，编写了《典型下垫面辐射光谱特性数据库》设计报告和用户使用手册。

此外，“九五”期间，在相关研究项目的支持下，中科院安徽光机所还完成了《目标/背景光谱特征数据库》。

中科院遥感所高光谱研究室在 1998 年建立了基于 Foxpro 的高光谱数据库，库中共有标准光谱数据 1152 条，其中植被 562 条，岩石 125 条，军事目标 146 条，城市地物 399 条，包涵了对象的光谱数据、测量的地学属性数据，例如：风速、云量、太阳高度角等、测量的仪器参数等等。该系统主要实现了数据库基本的查询检索、添加、删除，修改等功能，可以为对象的光谱画出曲线，并且可以实现去除包络线，以突现特征波段。尤其是这个高光谱数据库与高光谱图像处理系统建立了系统化的联系，使得光谱数据库在光谱图像处理中得到了实际应用。

2001 年 4 月，北京师范大学主持的国家 973 项目在北京顺义进行了星-机-地同

步大型遥感实验，首次获取了大量的冬小麦地面光谱测量数据、飞行图象和配套的结构参数、农学参数、农田小气候参数和气象参数等全面系统的实验数据，进一步丰富了其原有的 85 组地面 BRDF 测量数据库，和 395 组 POLDER 二向性反射率数据，为光谱结构知识库的建立提供了经验和保证。

此外通过各种遥感应用项目，在林业、农业、环境和城市等方面也进行了许多地物波谱测量。如在国防科工委的支持下，国家海洋局大连环保所联合中科院上海技术物理所等多家单位以水污染特别是海洋污染为重点获得了 3 百多组光谱测量数据，编写了《中国污染水体光谱特征》一书。

中科院上海技术物理所通过 863 信息获取与处理主题、863 航天领域和 921 工程的支持，以水污染为重点开展了地物波谱测量研究，相比于我国过去的地物波谱测量研究，他们采用了高光谱乃至超光谱地面和航空波谱测量仪器，强化了同步测量，并在优化通道选择、通道组合方法和应用模型的实用化定量化方面进行了有益的尝试。

浙江大学 1979 年在国内率先开展农业遥感研究，八五和九五期间通过多项自然科学基金和国家科技攻关项目的支持，以水稻为重点建立了农学光谱估产模式，近几年又对高光谱参数（红边位置及其变动）与水稻农学参数的关系进行了实验研究。

国土资源部航遥中心在 1998 年建成了主要针对岩石矿物的地物光谱数据库（GOSDBS）。

中科院广州地化所在 1988 年建立了遥感地物光谱数据库及其管理系统；

湖南省遥感中心在 1989 年建立了洞庭湖地区地物光谱数据库；

冶金部天津地质研究院于 1992 年建立了微机岩矿波谱数据处理及矿物识别软件系统；

北京农业信息技术研究中心在北京昌平小汤山拥有 2500 亩以农作物为主要目标的遥感试验场。自 1998 年以来，获取了小麦返青、起身、拔节、孕穗及灌浆期的生化组分与光谱同步数据 64000 多组。其中 2001 年 3～5 月以每星期两次（其中一次进行同步农学采样）的时相频率，取得光谱数据、农学采样及生理生化数据 41000 多条。

2005 年刚刚通过验收的国家典型地物波谱数据库项目收集了岩矿、农作物、水体等类型、覆盖全国范围的成套波谱数据三万多条。

### 2.2.3 小结

从国内外光谱数据库的建设与发展中，可以看出以下几点发展趋势：测量光谱的光谱分辨率逐渐提高、覆盖波长逐渐拓宽；光谱的配套数据逐渐完善；光谱数据库功能逐渐增多，由简单的数据库管理功能向光谱分析和模拟功能过渡；数据库逐渐走向专业化与应用化。

同时，也能看出，对于以高光谱图像立方体和像元波谱为核心的高光谱数据库的设计、建立与系统应用，目前还只是刚刚开了个头，还有待进一步进行系统的研究。

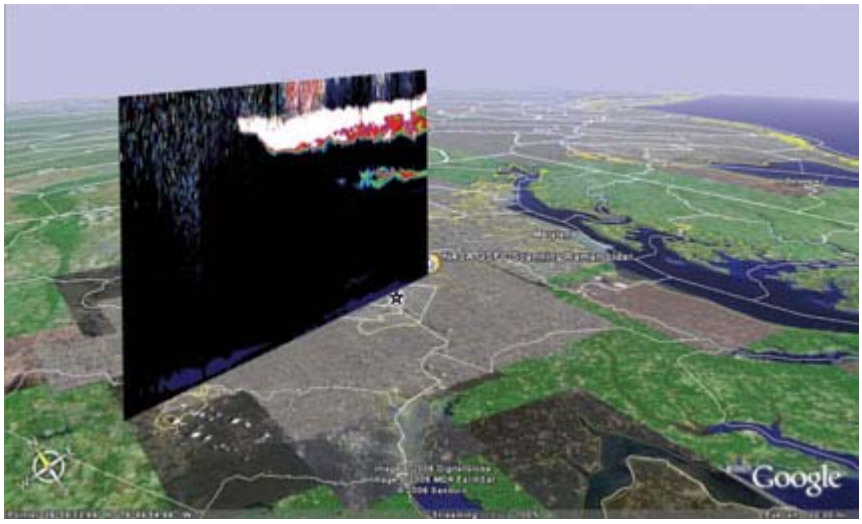
## 2.2 遥感影像数据库

### 2.2.1 国外遥感影像数据库现状

我们可以通过几个著名的研究机构或公司的产品中来探讨国外遥感影像数据库的现状。本节主要列举了 Google, Microsoft, NASA 等几个主要的遥感影像数据库作为案例进行阐述。

#### 2.2.1.1 GOOGLE EARTH

2006 年 2 月 16 日，Nature 杂志采用 Google Earth 的图片（如图 2.1）作为封面，并且发表了 Declan Butler 的文章 Virtual Globes: The web-wide world。Declan 认为 Google Earth 正在改变我们与空间数据的交互方式，（Declan B， 2006）。




图表 2.1Google Earth 地图影像浏览及大气轮廓数据切片

2004 年 10 月 27 日 GOOGLE 宣布收购了位于加州的 Keyhole 公司，Keyhole 是一家卫星图像公司，总部位于美国加州山景城（Mountain View），成立于 2001 年，从事数字地图测绘等业务，它提供的 Keyhole 软件允许网络用户浏览通过卫星及飞机拍摄的地理图像，这一技术依赖于数以 TB 计的海量卫星影像信息数据。2004 年 11 月 Google 推出了 Google Maps，紧接着于 2005 年 6 月 29 日推出了免费的桌面卫星地图服务工具——Google Earth，其最大特色就是结合本地搜索和卫星图片，可以让用户看到建筑物或地形的三维图像。2005 年 9 月 5 日， Google 宣布向中国用户推出 Google Local，这是 Google 继美国、加拿大、英国和日本以后，第五个开通本

地搜索服务的国家。2005 年 7 月 20 日，Google 推出 Google Moon（月球搜索），纪念人类登上月球 36 周年。2005 年 11 月 8 日，Google 宣布用户可通过部分型号的手机无线连接使用 Google 的卫星地图搜索服务，其使用方式与电脑用户完全一样。2006 年 1 月 5 日，Google 与摩托罗拉签署了一项全球移动搜索合作协议，摩托罗拉的手机用户以后将能轻松使用 Google 提供的以地图搜索为核心的新型搜索服务。2006 年 2 月 3 日，Google 宣布与德国大众美国分公司合作，将卫星地图软件 Google Earth 整合到大众汽车里，共同开发车载导航系统。2006 年 3 月 9 日 Google 与明基（BenQ）达成协议，将在一些型号的手机上提供一系列 Google 搜索工具，包括地图搜索。2006 年 3 月 13 日 ，Google 推出 Google Mars（火星搜索），使用户能够利用鼠标近距离地观察火星的地貌。2006 年 3 月 15 日，Bentley 将 MicroStation 同 Google Earth 服务进行结合，用户首次能够在 Google Earth 环境中进行基建工程二维/三维模式的观看与导航。（江寒，陈露，2006）Google Earth 一经推出，发展迅速，相应的商业应用也随即拓展。

Google Earth 提供了三个版本（不含企业服务器版）：个人免费版、Plus 版、Pro 版。三个版本的对比如表 2.1 所示：

表格 2.1Google Earth 软件三个版本对比

| | Google Earth | Google Earth Plus | Google Earth Pro |
|---|---|---|---|
| 价格 | 免费 | 20 美金/年* | 400 美金/年* |
| 影像数据库 | 主数据库 | 主数据库 | 主数据库 |
| 各区域浏览 | √ | √ | √ |
| 查询学校、公园、餐馆和旅馆 | √ | √ | √ |
| 查询行车线路示意图 | √ | √ | √ |
| 以 3D 方式倾斜与转动 | √ | √ | √ |
| 打印尺寸 | 1000 像素 | 1400 像素 | 2400 像素 |
| 绘制草图工具 | × | √ | √ |
| GPS 设备数据导入（只读）** | × | √ | √ |
| 电子表格数据批量导入 | × | 100 点 | 2500 点 |
| 客户服务支持 | 仅网上 | 网上、Email | 网上、Email、电话 |
| 测量区域距离 | √ | √ | √ |
| 视频电影生成模块*** | | | 200 美金/年 |
| 高精度打印模块*** | | | 200 美金/年 |
| GIS 数据导入模块*** | | | 200 美金/年 |
| GDT 交通计量数据导入模块*** | | | 200 美金/年 |
| NRB 商务信息数据模块*** | | | 200 美金/年 |

随后，Google Earth 的第四个版本 Enterprise Solution 也面世。Google Earth Enterprise 的融合模块整合了 Google Earth imagery，可以支持和显示光栅图、GIS、

地形图和点数据;其客户软件可以从 Google Earth Server 那里获得数据;Google Earth EC （Enterprise Client）还支持浏览、打印和修改数据。

Google Earth 的影像数据包括卫星影像与航片。其卫星影像部分来自于美国 DigitalGlobe 公司 QuickBird 商业卫星与 EarthSat 公司（影像来源于陆地卫星 LANDSAT-7 卫星居多），航拍部分的来源有英国 BlueSky 公司，美国 Sanborn 公司等。

#### 2.2.1.2 MICROSOFT VIRTUAL EARTH

微软公司于 1998 年联合美国地质调查局、美国航天局以及俄罗斯空间署建立了面向城市地区的 TERRASERVER，把美国和俄罗斯数十年集成的高精度全色卫星影像通过互联网络向全世界展示。该影像数据库数据总量达到 5TB，以 JEPG 格式存储，通过四级金字塔提供缩放和浏览。在此基础上的 MSN Virtual Earth 是在 2005 年 5 月举行的一次会议上，由微软董事长兼首席软件工程师比尔·盖茨宣布推出，并计划免费提供。2005 年 7 月 26 日微软正式推出其 MSN Virtual Earth Beta 服务，成为继 Google 后第二个推出全面地图搜索服务的网络服务商。目前，MSN Virtual Earth 仅针对于美国地区开放，用户可以从地图上找到最近的中餐厅或者 ATM 机地点。与 MSN Virtual Earth 配套的客户端软件 Microsoft Location Finder 是一个独立的地图搜索软件，由 Microsoft Research 开发。它能够与 MSN Virtual Earth 无缝连接， 与 WiFi 网络配合可以确定用户的所在地。

MSN Virtual Earth 的基本服务是搜索某地区的地理信息，例如搜索该地区商店的信息，然后将搜索结果显示在卫星图片上。在“Where 2.0”会议的演示中，在美国西雅图市的卫星图片上搜索并显示了该地区餐厅、咖啡馆及电影院的信息。用户可以通过“便签本（Scratch Pad）”功能把搜索结果保存起来，也可将保存的图片和搜索结果通过电子邮件发送给其他用户。此外，还可与保存用户“博客”及个人信息的服务“MSN Spaces”配合使用。

MSN Virtual Earth 与 Google Earth 的最大不同是，Google Earth 在操作中需要专用的客户端软件，而 MSN Virtual Earth 则不需要插件。MSN Virtual Earth 只需要微软的“网络资源管理器”与开放源码 Web 浏览器“Firefox”即可进行操作。据微软称，MSN Virtual Earth 利用“AJAX”（Asynchronous JavaScript and XML）技术管理图像。

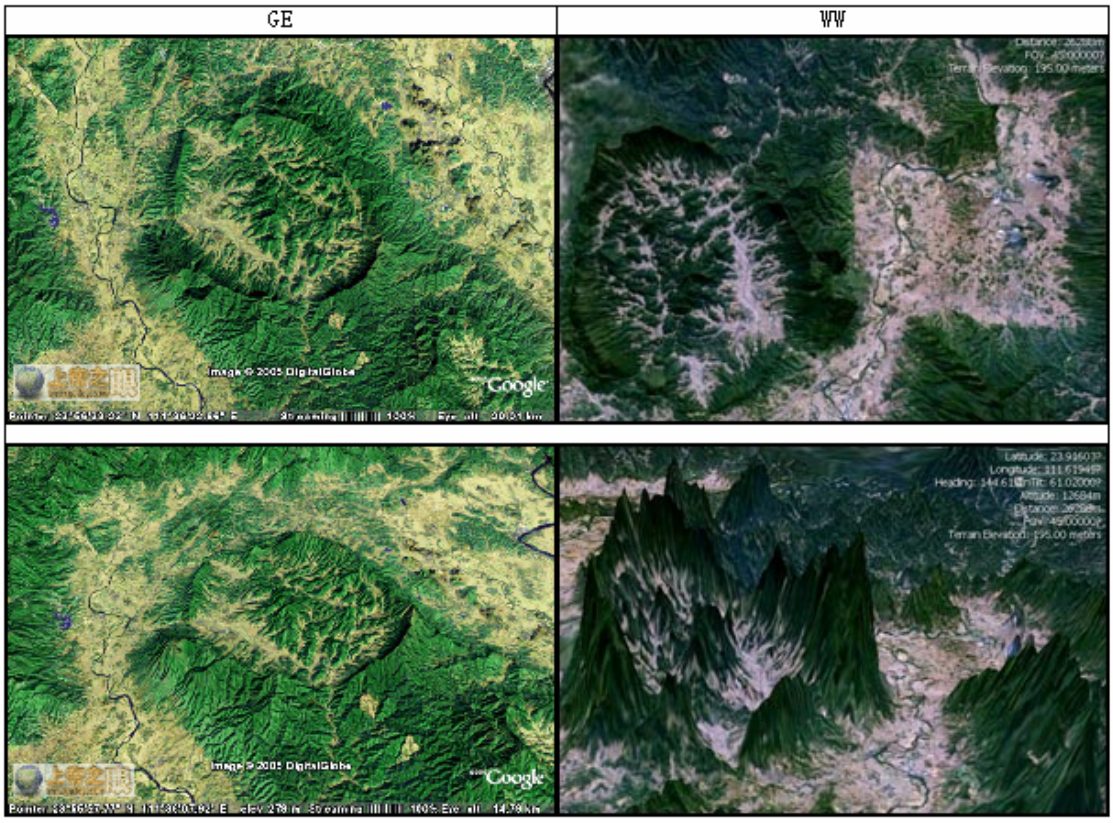
MSN Virtual Earth 的影像数据主要由 ORBIMAGE 公司提供。

#### 2.2.1.3 NASA WORLD WIND

World Wind 是由美国的 NASA（美国航天宇航局）机构开发出来的，整个系统

的开发初衷是面向于专业领域，利用 NASA 卫星的优势，给使用者营造一个有关卫星图片的查看平台。随着版本的不断升级，World Wind 在最新的版本中加入了各种插件，例如：比如以前的网络卫星图片，只能够通过指定的服务器下载，利用相关插件，让用户可以得到更多网络卫星图片服务器中的资源。通过这套完全免费的软件系统的 3D 引擎，可以从外太空看见地球上的任一角落。World Wind 的作用还不仅仅是浏览真彩色高分辨率遥感影像，它的数据源包括：Blue Marble 提供分辨率达到 15 公尺的真彩色遥感影像，MODIS 提供可见光以外的波段影像，Terra，Landsat 7 查看以前的历史遥感影像， SRTM（Shuttle Radar Topography Mission，与 landsat 整合可以从各个角度查看地球表面，模拟近地面飞行），NASA SVS（查询龙卷风动画等）， GLOBE（查询当日即当日以前的全球温度及温度变化）等等。World Wind 的虚拟旅游功能比 Google Earth 更逼真，透过及时动画形成的模组可以体验飓风席卷佛罗里达或者体验气候变化情况等等。

World Wind 提供了上百个插件，用以提升其功能和性能，而且其所有的代码都是开源的，可以提供给开发人员和研究人员作二次开发。此外，如图 2.2 所示：World Wind 的图像质量和 3D 渲染效果相当出色，不过，在速度和资源占用方面，World Wind 比 Google Earth 尚显不足。




图表 2.2 Google Earth（左）和 World Wind（右）卫星影像浏览和 3D 展示

2.2.1.4 Digital-NGP

Digital-NGP （Digital Northern Great Plains Project） 是一个在线的遥感图像存储及发布系统，是由美国的 North Dakota 大学 Upper Midwest Aerospace Consortium 开发的一套系统。用它可以查询，下载美国地区各种尺度的各种卫星遥感影像。

提供的影像类型包括：MODIS， ETM， TM， QuickBird， MSS， ASTER， ETM+， SRTM 等，并能查看指定日期的各地区的各种类型的影像。并对于多光谱影像可以选择特定波段，生成 RGB 图像。然后还可以选择生成特定的图层的内容，如 Boundaries， Transportation， Water， Geology， Raster Layer， Agriculture 等分类下的特定图层。同时也可进行 NDVI 等指数的分析。生成的影像均为 png 格式，可以自由打印和下载。

系统数据库用的是 ORACLE，底层影像的处理主要是依靠开源的 Mapserver 来完成。影像数据通过 GeoTIFF 存储，然后通过 Mapserver 根据用户的请求生成 png 图像。（数据库介绍：http://digital-ngp。aero。und。edu/ ，数据库应用界面（3。7 版）：http://digital-ngp。aero。und。edu/newdngp370/index。php ）

2.2.1.5 NOAA SERVER

NOAA 影像数据库主要包括来自：NOAA 公司，国家气候数据中心（NCDC），国家环境卫星数据和信息中心（NESDIS），国家地理数据中心（NGDC），国家海洋渔业中心（NMFS），国家海洋服务中心（NOS），国家冰雪数据中心（NSIDC），国家气象服务中心（NWS），海洋气象办公室（OAR），国外数据库（FDL），日本科技公司（JST）等等，影像内容涉及农业、气象、空气污染、太阳辐射、酸雨等等。

### 2.2.2 国内遥感影像数据库现状

北京大学遥感与地理信息系统研究室研制了遥感影像 WEB 发布系统，该系统根据客户区和用户选择区的大小在服务期端对影像进行抽点拼合，将拼合后的数据返回给客户端，用户可以对影像进行放大缩小，并以当前分辨率漫游，还可以将所看到的影像保存下来，并具备矢量图叠加功能。但是该系统只是针对某一地区的示例，没有影像数据库管理和原数据库查询的功能（王宇飞，2001）

中科院卫星地面站（http://cs.rsgs.ac.cn/cs_cn/query/query_map.asp）通过在数据库中存储影像的相关属性数据和缩略图片，提供对卫星影像的检索。他们通过 WEBGIS 对经纬度坐标的查询，显示查询区域坐标范围内符合相关属性描述（云量、数据获取时间、卫星类型）的影像缩略图。

武汉大学研究了基于 C/S 的 GeoImageDB 影像数据库系统，能够实现对影像数据的导入、查询、浏览，并支持将数据导出为 BMP 格式的标准图像

王宇飞以 JPEG2000 静止影像压缩标准为基础，开发出了基于网络的海量遥感影像浏览系统，将遥感影像数据与矢量数据、属性数据、三维 DEM 数据进行叠加、套合与集成，形成具有多元数据性质的网络地理信息系统。（王宇飞，2001）

张雄飞等通过分析高光谱数据的特点，结合数据库开发实践，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（张雄飞，张兵，张霞等，2004）

杜云艳等基于海洋数据的多样性和时空复杂性的特点，采用 oracle9i 的 ArcSDE 和 ArcGIS 桌面系统，完成了海岸带及近海科学数据平台实体的装载和组织工作，并在 Arc Object 组件的基础上开发了基于 C / S 的数据平台的前端管理系统，实现面向海洋用户的遥感数据无缝拼接、多源数据多种快速查询、提取、影像数据的装载以及影像元数据的管理和修改等多种功能。（杜云艳、苏奋振、杨晓梅等，2004）

李军等在大型遥感图像处理项目应用的基础之上，进行了集成数据库的设计与应用，提出了虚拟数据库的概念，并结合实例说明了集成数据库以元数据为链条的使用机制与方法。（李军、刘高焕、迟耀斌等，2001）

李军等在北京市综合遥感影像数据库的前期建设中，对遥感影像数据库进行了先期研究，对于常见的三种数据库建设方案（基于 ESRI 技术、基于吉奥技术、基于 ermapper 技术的解决方案）作了技术分析与比较，认为，目前的技术已经能够满足建设大型的综合遥感影像数据库系统的要求。在此基础上进一步分析系统需要实现的功能、面临的服务对象和硬件网络环境，分析遥感影像数据库的综合应用和与其他数据库的结合是系统实际建设方案选型的重要依据。（李军、李琦、毛东军等，2003）

吴信才等利用 COM 技术与 SQL 数据库技术实现了海量影像数据库（IMIDB）的逻辑结构设计，通过遥感影像数据的金字塔结构来存储不同尺度的数据，并通过分块策略进行管理实现影像数据的 web 发布。（吴信才、郭玲玲、李军等，2002）

冷秀华等设计并开发了高光谱遥感数据管理系统原型，能够实现对高光谱影像数据进行检索、查询，空间截取等功能。（冷秀华、张杰、马毅等，2003）

陈华等针对遥感影像讨论了利用 ASP 技术动态访问图像数据库的原理，通过“虚拟目录”来存储图像，保证数据的安全性。系统采用 B/S 架构，利用 ADO 访问 SQL SERVER，实现了图像的动态访问。（陈华、曹锦云、郑学峰等）

刘鹏等通过对不同遥感数据的研究，提出了遥感影像原数据标准草案，构建了影像数据的元数据库；利用大型关系数据库对遥感影像数据及其原数据进行组织与管理；通过基于 XML 的影像数据发布，实现用户通过网络对遥感影像数据的检索、查询和访问。他们认为，基于元数据的遥感影像管理方案能够有效的解决还亮遥感影像数据的组织、存储和管理问题。（刘鹏，毕建涛，曹彦荣等，2005）

李航、岳丽华在研究如何高校存取和管理还亮遥感图像数据，并深入分析基于COM 组建的 GIS 二次开发技术和 ArcSDE 中间件技术的基础上，提出了利用基于COM 的 GIS 组建开发客户端应用程序，利用 ArcSDE 开发后台遥感图像数据库的解决方案，并给出了在 VC++6。0 和 ORACLE9i 环境下的一个开发实例。他们认为，基于 COM 的 GIS 开发可以降低开发难度，提高开发效率，增强系统的灵活性和开放性。ArcSDE 空间数据引擎利用客户／服务器计算模式和关系数据库管理的先进特点，为遥感图像数据的存储和管理提供了新的解决思路。（李航、岳丽华，2005）

王密、龚健雅等分析了我国数字正摄影像产品的空间参考特性和建立大型无缝影像数据库所带来的问题，在此基础上提出了海量影像数据管理中分带存储模式和跨带漫游算法。他们采用这种方式成功地管理了广东省全省 l：l 万的 37 度、38 度、39 度三个高斯投影带的影像，并且整个数据库可以进行快速的无逢漫游。（王密、龚健雅、李德仁，2001）

数据压缩是解决海量数据远程传输的关键技术，编解码速度和重建图像质量是评价压缩系统性能的重要指标。李飞鹏、秦前清等通过影像分块、小波变换、最优量化、MQ 算术编码、容错编码、光谱转换等方法，设计和实现的一个海量遥感影像压缩系统。该系统采用较新的编码技术，能在半分钟内完成 100M 以上图像的高保真压缩，并支持单波段和三波段影像压缩。（李飞鹏、秦前清，李德仁，2003）

李宗华、彭明军以 ORACLE GEORASTER 和 ARCSDE 为例，研究了基于扩展的面向对象关系数据库进行遥感影像建库和基于中间件技术进行遥感影像建库的数据模型、数据存储方式、空间索引、数据管理与维护等问题。他们对遥感影像建库实施的数据分块、影像金字塔的构建、数据压缩处理等具体问题进行了实验研究，得出在 32 位 windows 2000 server 操作系统中，采用 oracle 9i 建立影像数据库，数据库的数据块大小设置为 16k，遥感影像数据块大小设置为 128*128 时，系统运行效率最高。（李宗华，彭明军，2005）

陶冶宇等通过对 oracle spatial 和 oracle intermedia 的研究，探讨了基于对象-关系模型的多分辨率遥感影像数据的组织和管理，设计和建立了多分辨率遥感影像数据库的原型，采用了地形瓦片和影像金字塔的快速索引结构。他们的实验表明这种设计能够平稳快速的实现多分辨率遥感影像的快速调度显示，并且显示速度不受影像分辨率以及影像显示范围所影响。（陶冶宇，马东洋，徐青等，2005）

赵艳玲等对基于 web 海洋卫星遥感产品查询系统的架构、数据库的建立和动态网页的编程作了研究，采用了三层体系的组织方式，利用 VC++开发数据层中间件，用 ASP 技术构建业务层与表示层，实现了对遥感产品的网上实时发布。（赵艳玲，何贤强，王迪峰，2005）

张芬，高炎将 MRSID 影像压缩方法应用到上海市测绘院研制的影像数据库系统中，通过较少的数据量存储 MRSID 压缩格式的影像数据，并利用 MRSID 的解压接口实现数据自定义范围的提取功能，满足规划管理和资源调查等对影像数据的特

殊需求。（张芬，高炎，2005）

方涛、龚健雅等研究了建立大型遥感影像数据库中需要加以研究和解决的主要问题和若干关键技术，包括影像数据库的体系结构，扩充的关系数据库，无缝拼接，正射影像图的制作，数据压缩等。（方涛，龚健雅，李德仁，1997）

### 2.2.3 小结

国外遥感影像数据库已经进入大规模的网络化服务时代，普通用户可以通过网络浏览、查询、下载。数据流管理和图像快速显示成为这些遥感影像数据库的主要技术核心；图像之间的镶嵌、无缝集成是其主要特点；影像数据的浏览是当前遥感影像数据库的主要功能。遥感影像数据库的现状是向横向发展，侧重于数据库应用的广度，而对于遥感影像数据库应用的深度挖掘还显不够，尤其是在基于数据库技术基础之上的数据挖掘应用研究并不太多。

## 2.3 数据挖掘技术

### 2.3.1 数据挖掘技术

数据挖掘是人工智能、数据库、统计理论相结合的技术，具有广泛的应用前景，它是指从大型数据库或数据仓库中提取隐含的、未知的、非平凡的及有潜在应用价值的信息，这些信息可表示为概念（Concepts）、规则（Rules）、规律（Regularities）、模式（Patterns）等形式。数据挖掘是一个多学科交叉的新兴学科，涉及数据库技术、人工智能、机器学习、神经网络、统计学、模式识别、知识获取、信息检索、图像与信号处理、空间分析、高性能计算和数据可视化等学科领域。数据挖掘借鉴和学习这些学科的理论、方法，形成了适合自身特点的一系列的数据挖掘算法和技术，实现数据中隐含知识的提取。

自 20 世纪 60 年代以来，数据库和信息技术已经从原始的文件处理演化成复杂的、功能强大的数据库系统，这大大加快了世界信息化的进程，为数据的大量积累提供了基础。数据库技术的发展过程经历了三次革命：第一次革命发生于 60~70 年代，出现了功能强大的数据库系统；第二次发生于 70~80 年代，代表技术是数据的组织和使用、数据库中的信息检索和事务处理，关系数据库管理系统逐渐成熟并被广泛使用；第三次发生于 80~90 年代，先进的数据模型被采用，如扩充关系模型、面向对象模型、对象——关系模型和演绎模型，产生了分布式数据库、扩展关系数据库、面向对象数据库、演绎数据库和异质数据库管理系统，结合特定行业的需求，产生了面向应用的数据库系统，如空间数据库、时态数据库、多媒体数据库、科学数据库、知识库、全球信息库等。

数据挖掘首次出现于 1989 年 8 月在美国底特律召开的第十一届国际人工智能

联合会议（INTCAI）上。随后在 1991 年、1993 年和 1994 年度举行了 KDD（Knowledge Discovery from Databases） 的专题讨论会，汇集了来自各个领域的研究人员和应用开发者，以及他们在各自领域的研究成果。由美国人工智能协会主办的 KDD 国际会议从 1995 年开始，成为一年一度的大型国际学术会议，规模由二三十人发展到七八百人，研究重点也逐渐从发现方法转向系统应用。此外，数据库、人工智能、信息处理、知识工程等领域的国际学术刊物也纷纷开辟了 KDD 专题或专刊。

数据挖掘与知识发现是目前国际上数据库和信息决策领域的最前沿研究方向之一，引起了学术界和工业界的广泛关注。一些高级别的工业研究实验室，如 IBM Almaden 和 GTE，以及众多的学术单位，如美国的 UC Berkeley 和加拿大的 Simon Fraser 大学，都在这个领域开展了各种各样的研究计划，研究的主要目标是发展有关的方法、理论和工具，以支持从大量数据中提取有用的、让人感兴趣的知识和模式。

经过十多年的努力，数据挖掘技术的研究已经取得了丰硕的成果，国际上许多数据库和数据仓库供应商、统计分析软件开发商以及专门的 DMKD 开发商、研究所等，相继研制开发出了数据挖掘软件产品。例如，IBM 公司的 QUEST 和 Intelligent Miner；加拿大 Simon Fraser 大学的 DBMinner；SGI 公司的 Mine Set 等。此外，一些关系数据库产品也加入了相关的数据挖掘功能和支持数据挖掘的接口，用户可以通过构造数据仓库来实现数据挖掘功能，或者进行二次开发， 来建立专门的数据挖掘系统， 如 Microsoft SQL Server2000、Oracle9i、Informix 等数据库产品。随着国外知识发现的兴起，我国也很快跟上了国际步伐。《计算机世界》报技术专题版于 1995 年 3 月发表了由李德毅教授组织的 KDD 专题；于 1995 年 4 月发表了由中国科学院组织的“机器学习、神经网络”专题；于 1995 年 12 月发表了由国防科技大学陈文伟教授组织的“机器发现和机器学习”专题，这些都推动了我国数据挖掘和知识发现技术的发展。

数据挖掘应用取得了很大的成就。在多学科相互交融和相互促进的信息时代，数据挖掘为大型数据库的利用提供了有效工具，是决策支持系统的一个重要组成部分。在科研、工业、商业、经济、金融、管理等领域，数据挖掘已引起极大关注，吸引了众多的研究人员和商业机构。CAT（sky image cataloging and analysis tool），使用 SKICAT 对天体数据进行分析，一方面是通过机器学习将知识提取过程由学习算法完成，从而实现对大批量数据的分析；另一方面是辩识那些亮度很低、人工难以判读的天体图像，以进行后续分析。Simon 等人研制的 BACON 系统成功地重新发现了理想气体定律、库伦定律、开普勒第三定律、欧姆定律和伽利略定律。

基于数据挖掘技术，“啤酒__尿布”规则指导美国加州某连锁店的销售，ISPA 系统分析钢铁产品的性能规律，Fidelity Stock Selector 系统选择投资，LBS Capital Management 系统管理有价证券，AC Nielson 和 Infrates Burk 等国际市场研究公司预

测市场，BBC 广播公司预测电视收视率，信用卡公司 American Express 吸引客户，ＡＴ＆Ｔ公司侦探国际电话欺诈行为，公安部门侦破案件等，都为决策者提供了极有价值的知识和效益。由 AcknoSoft 公司用 KATE 发现工具开发的 CAS SIOPEE 系统，已用于诊断可预测在制造波音飞机过程中可能出现的问题。在通信网络管理方面，芬兰 Helsinki 大学与一家远程通信设备制造厂合作的 TASA 系统，可用于网络故障的定位检测和严重故障的预测等任务中。我国的宝钢和零售企业等也应用了数据挖掘技术。中国科学院计算机技术研究所与国家税务部门合作，开发了计算机选案系统，用于稽查和追缴偷、漏、欠税款，查处和纠正纳税人的违法行为。

当前，数据挖掘领域的研究与发展趋势可概括为：

（1）应用领域的探索和扩张。在注重理论、技术研究的同时，强调实际应用研究，如一般化的、通用的及针对特定领域的数据挖掘系统的开发。

（2）算法的效率和可伸缩性。数据挖掘直接面对的是海量数据，且这些数据之间已含着各种繁杂关系，这就导致挖掘过程中搜索空间和搜索维数的激增，且其间的许多不确定因素和干扰因素也就随之增加，但许多成熟的算法是基于内存的，这就对算法的效率提出了严峻的挑战。另一方面，由于数据量是随时间激增的，因此，针对单独、集成的数据，挖掘功能的可伸缩性就显得非常必要。

（3）数据挖掘系统的交互性。数据挖掘中适当的用户参与是必不可少的，是基于以下几个方面的原因：①友好的、完善的交互界面是用户准确表达其要求和挖掘策略的保证；②用户的背景知识和指导作用可以提高挖掘效率，并保证发现知识的有效性。目前一个重要的研究方向就是所谓基于约束的挖掘（Constraint-bad）它致力于在增加用户交互的同时如何改进挖掘处理的总体效率；③通过交互界面，系统可以直接、有效地获取用户的感兴趣模式，从而提高挖掘的有效性。

（4）复杂数据源和数据类型的处理。数据挖掘系统的理想体系结构是与数据库和数据仓库系统的紧耦合方式，把事务管理、查询处理、联机分析处理和联机分析挖掘集成在一个统一的框架中，然而，正如前所述，实际应用中的数据挖掘对象是各种类型的数据库，甚至是没有完整数据结构的数据集，因此，如何把这些特殊数据类型的专用分析方法与现在成熟的基于关系数据库和数据仓库的数据挖掘方法集成起来，实现这些复杂数据源和数据类型的处理，是一个重要的发展方向。

（5）隐私保护与信息安全。在发展数据挖掘的同时，需要进一步开发有关方法以便在适当的信息访问和挖掘过程中保护隐私和信息安全。

数据挖掘面临的主要挑战是：数据输入形式的多样性；数据挖掘算法的有效性与可测性；用户参与和领域知识；证实技术的局限；知识的表达和解释机制；知识的维护和更新；私有性和安全性支持的局限、与其它系统的集成。

所以，当前数据挖掘的研究是以知识发现的任务描述、知识评价与知识表示为主线，有效的知识发现算法为中心，面向具体应用，开发原型系统与实用系统，研究与开发基于数据挖掘的通用工具。

### 2.3.2 空间数据挖掘技术

1994 年在加拿大渥太华举行的 GIS 国际会议上，李德仁院士首次提出了从 GIS 数据库中发现知识———KDG（Knowledge Discovery from GIS）的概念。他系统分析了空间知识发现的特点和方法，认为从 GIS 数据库中可以发现包括几何特征、空间关系和面向对象的多种知识，KDG 能够把 GIS 有限的数据变成无限的知识，可以精练和更新 GIS 数据，使 GIS 成为智能化的信息系统，并第一次从 GIS 空间数据中发现了用于指导 GIS 空间分析的知识。（Li D.R. 1994）

空间数据挖掘是在空间数据库的基础上，综合利用统计学方法、模式识别技术、人工智能方法、神经网络技术、粗集、模糊数学、机器学习、专家系统和相关信息技术等，从大量的空间生产数据、管理数据、经营数据或遥感数据中析取人们可信的、新颖的、感兴趣的、隐藏的、事先未知的、潜在有用的和最终可理解的知识，从而揭示出蕴含在数据背后的客观世界的本质规律、内在联系和发展趋势，实现知识的自动获取，提供技术决策与经营决策的依据（李德仁，2001）。

Murray 和 Estivill.Castro（1998）回顾了探测性空间数据分析的聚类发现技术，分析了基于统计学、数据挖掘和地理信息系统的空间模式识别和知识发现方法。Koperski.Adhikary 和 Han（1996）总结了空间数据挖掘的发展，认为巨量的空间数据来自从遥感到 GIS、计算机制图、环境评价和规划等各种领域，空间数据的累积已经远远超出人们的分析能力，数据挖掘已经从关系数据库和交易数据库扩展到空间数据库。他们就空间数据生成、空间数据聚类和挖掘空间数据关联规则等方面总结了空间数据挖掘的最近发展。

Lenarcik 和 Piasta 把概率论和粗集相结合，利用条件属性推理决策知识，开发了 Prob Rough 系统（Probabilistic rough classifier generation）。利用基于决策树的概率图模型， Frasconi，Gori 和 Soda 对带有图形属性的数据库进行挖掘，得到了用于指导机器学习的知识。Cressie 利用地理统计数据、栅格数据和点数据三种空间数据描述现实世界，并据此提出了一个通用模型。由于大部分空间数据挖掘的研究偏重于提高静态数据查询的效率，所以 Wang、Yang 和 Muntz 基于统计信息，研究了一种由用户定义的主动空间数据挖掘的方法。应用空间统计学的克吕格方法，由一组已分类的观测点直接估计未观测点位的属于各类别的验后概率，求得类别变量在任一位置上所观测到的各类别的概率知识，就可以从影像上获取模糊分类信息。冯建生也利用空间统计学揭示了影响冲击韧性的因素知识。

Koperski 和 Han（1995）提出了一种在地理信息数据库中挖掘强空间关联规则（空间数据库中使用频率较高的模式或关系）的算法，并给出了两步式的空间优化技术。程继华和施鹏飞（1998）提出了多层次关联规则的挖掘算法，利用集合“或”、“与”运算求解频繁模式，提高了挖掘的效率。许龙飞和杨晓昀（1998）分析了广义关联规则模型的挖掘方法、挖掘策略和规则挖掘语言。Eklund， Kirkby 和 Salim

（1998）在土壤盐度分析中把决策支持系统和 GIS 数据相结合，发现了用于环境规划和二级土壤盐碱化监测的关联规则。Aspinall 和 Pearson（2000）把风景生态学、环境模型和 GIS 结合在一起，通过综合地理评估，研究了美国黄石国家公园的汇水处环境条件，发现了用于环境保护的关联规则。涂星原（1998）研究了基于数值属性的关联规则的挖掘。Clementini，Felice 和 Koperski（2000）在宽边界的空间实体中挖掘出了多层次的空间关联规则。左万利（1999）研究了在含有类别属性的数据库中提取关联规则的类型转换技术。丁祥武（1999）在关联规则模型中增加了描述关联规则时效性的时态信息。丁祥武（1999）根据数据记录之间的时间间隔和相邻记录中项目的类别合并同类记录，肖利等（1998）用时间窗刻画时间约束。程继华（1998）等提出了基于概念的关联规则的挖掘算法。肖利（1997）等提出了一个基于关系操作的挖掘广义关联规则算法，在多概念层上交互挖掘关联规则。

杨学兵等（1999）的实时数据挖掘算法能在实时过程控制中自动挖掘，并根据挖掘的知识预测趋势。Levene 和 Vincent（2000）发现了关系数据库的功能独立和包含独立的规则，信息处理使用了基于知识规则挖掘的分类方法。Ester， Kriegel 和 Xu（1995）使用聚类技术研究了在大型空间数据库中挖掘类别判读知识的技术。Knorr 和 Ng（1996）分析了空间数据挖掘中的聚类和特征关系，提出了发现聚合亲近关系和公共特征的算法。Edwin 等（1997）通过构造地理信息系统中的聚类器，发现了空间物体的边界形状匹配关系的部分规律。Lin，Zhou 和 Liu（1999）根据类别和特征，研究了空间数据库中的临近关系匹配算法。Tung，Hou 和 Han（2001）提出了一种在空间数据挖掘中实行空间聚类时，处理河流、高速公路等阻隔的算法。Reinartz（1999）给出了他关于现实世界的数据挖掘方案及其实验结果。

周成虎和张健挺（1999）从信息熵的基本概念出发，认为地学空间数据子集划分产生的互信息或熵减源于子集划分，使得各个子集的不确定性或模糊性降低，并且子集之间的差异性增大，因此具有最大熵减的子集划分方案代表一定的地学模式和地学规律。并以此为基础分别探讨了地学数据属性要素的子集划分产生多维属性关联规则，以及通过空间和时间的子集分割来进行聚类的方法。Ester 等（1995）以空间的点为基本单位，研究了多空间物体的相邻关系的处理技术，集成了空间数据挖掘算法和空间数据库管理系统，同时利用相邻图形和路径以及小型的初始数据库操作挖掘空间模式，使用相邻索引来提高初始数据库的处理效率。Mouzon， Dubois 和 Prade（2001）在空间可能因果关系的属性异常诊断索引中，使用一致和诱导的算法挖掘了属性不确定性对异常诊断影响的知识。

布和敖斯尔（1999）提 出了基于知识发现和决策规则的盐碱地 GIS 和遥感分类的方法，把盐碱地分类的地学专家思想和区域专家的思想应用到 GIS 数据挖掘中，并把从 GIS 数据库中发现的知识，按一定的规则应用到华北平原地区的盐碱地分类的决策中，能够简化数据运算过程，减少或避免分类过程中人为误差的产生。陈春香（1999）应用机器学习中的数据驱动发现学习方法处理广东云浮__阳春地区的地

球化学数据的实践证明，可以挖掘出隐含在数据间的各参数间理解，为地球化学找矿提供更合理的决策信息。

### 2.3.3 影像数据挖掘技术

Simon 大学开发的 MultiediaMiner，可以进行图像集的相关规则的挖掘；Mihai Datcu 和 Klaus Seidel 开发了一个智能卫星挖掘系统；Michael C．Burl et a1．用图像挖掘的技术对 NASA 的目标进行分析，获取信息。Mihai Datcu et a1．用贝叶斯方法进行信息聚合和图像数据挖掘 ；Michael C．Burl et a1．讨论了图像数据挖掘的分布式结构；讨论了用图像挖掘技术进行语义特征的抽取，进行图像检索的新方法。我们可以看到，人们对图像挖掘研究的问题主要在于挖掘系统的建立和挖掘算法的发现。（薄华、马缚龙、焦李成，2004）

王晋年、张兵等（1999）认为高光谱信息挖掘技术是高光谱数据应用延拓与深入的重要环节，其核心在于光谱信息的挖掘。他们基于高光谱遥感信息的特点，探讨分析了以地物识别与分类为目标的高光谱数据挖掘技术，包括基于模式识别的高光谱信息挖掘技术、基于光谱波形特征的挖掘技术以及亚像元光谱信息挖掘。

马建文和马超飞（1999）分析了地面物质和结构光谱与卫星遥感信息之间的关系，建立了空间角度模型，通过对ＴＭ卫星数据的挖掘说明了基于空间角度算法在处理多波段遥感数据时的数学能力。

戴晓军、淦文燕、李德毅（2004）提出了基于数据场的图像数据挖掘，采用数据场和势的概念，提出了一种把非结构化数据转化为结构化数据场的思想，通过提取数据场不同层次的局部极大值点，实现概念粒度的跃升，达到图像数据的降维。他们通过比较图像经过不同的非线性变换后对局部极大值大小和位置分布的影响，找到了合适的变换函数，并证明该方法能够突出局部特征，对于特征提取提供了一种新思路。

叶静、蔡之华（2003）描述了遥感图像上的数据挖掘技术，尤其是遥感图像的关联规则挖掘，并详细介绍了 Apriori 算法在遥感图像中的应用，并且对其它两种数据挖掘方法：决策树分类以及神经网络方法作了探讨。

刘钊、蒋良孝（2003）从介绍一个图像数据挖掘可以采用的系统原型出发，对图像数据挖掘的常用方法和基本过程进行了详细的阐述。他们认为图像数据挖掘技术可以广泛的应用于医学影像诊断分析、遥感影像分析、地下矿藏预测等等各种领域，但是现阶段研究工作尚不成熟，很有必要进行相应的技术研究。

邸凯昌和李德仁（2000）采用数据发掘技术从数据库和遥感图像中发现知识，用于改善遥感图像分类。提出了两种实施空间数据归纳学习的途径：在空间对象粒度上学习和直接在像元粒度上学习。分析了两种粒度学习的特点和适用范围，同时提出了一种归纳学习与传统图像分类法的结合方式。用北京地区多光谱图像和数据

库进行土地利用分类的试验证明，归纳学习能较好地解决同谱异物、同物异谱等问题，显著提高分类精度，并且能够根据发现的知识进一步细分类，扩展了遥感图像分类的能力。

帅艳民、朱启疆等（2003）进行了地物波谱数据仓库系统设计的研究，按照参数配套原则，提出用树状结构的事实表和分维表实现数据管理，并利用北京顺义野外实验 SE590 波谱数据为实例，提出了基于相关作物模型和数据挖掘采掘观测间隙的光谱数据的方法。

马超飞，刘建强（2003）初步研究了卫星遥感数据的关联规则挖掘及其在土壤侵蚀和退耕还林上的应用，根据多维空间数据的特点，将遥感数据的数性质划分为不同的块，同时利用现有的关联规则挖掘的算法，将划分好的数据转变为事务数据库形式，最后利用 Apriori 算法提取了土壤侵蚀强度与坡度、植被覆盖度以及坡耕地之间的有意义的关联，为退耕还林还草决策提供了有力支持。

宫辉力、赵文吉、李京（2005）对多源遥感数据挖掘系统流程进行了研究，并设计了系统框架，提出了多源遥感影像挖掘系统原型。

### 2.3.4 小结

随着信息时代的到来，数据量的猛增，数据挖掘技术已经成为数据库发展的必然趋势之一。随着李德仁院士 1994 年提出 KDG，地理信息系统领域对于数据挖掘技术的应用作了深入的研究，并得到了一些较好的成果，数据挖掘算法日益改进更新，空间数据挖掘的应用也日益扩展。而对于遥感影像数据挖掘，是当今影像数据研究的趋势之一。目前研究主要集中在利用关联规则来实现对像元 DN 值和属性的关联分析上，而对于高光谱遥感数据方面，基于高光谱数据库的数据挖掘系统化研究还为数不多。

## 2.4 本章小结

本章主要介绍了地物光谱数据库、遥感影像数据库、数据挖掘技术在国内外的研究现状和主要的研究重点。通过对高光谱数据库背景和现状的研究，我们可以发现，地面测量光谱数据库建设和数据采集在国外已经较为成熟，在国内已经通过各项研究逐步完成各种数据的搜集工作，而且也有了比较好的数据平台，数据库建设正在逐步完善中；影像数据库研究在国外侧重于数据的无缝集成（镶嵌）、压缩、快速网际传输等，而且主要集中在高空间分辨率的全色或真彩色卫星影像方面。国内也有相关研究，而且有一些影像数据库集成了对图像属性的查询；数据挖掘技术在空间数据库领域有较为深入的研究，在遥感影像方面，则更多的是侧重于影像的关联分析，对于其他的数据挖掘技术与方法的应用，尚不多见。

# 第三章 数据挖掘和应用导向的高光谱数据库系统设计

当人类进入空间时代并跨入信息时代的门槛之时，各种运行与空间、翱翔于空中的遥感平台连续不断的在多尺度上对我们的地球进行着观测，各种先进的对地观测系统源源不断地向地面提供着各种信息源。作为地球空间信息科学与技术（曾澜，2001）发展的一个重要组成部分，高光谱遥感近年来在国际对地观测领域扮演起了越来越重要的角色。高光谱数据的应用日益广泛，已经涵盖了地质、农业、林业、城市土地利用、环境保护等等诸多领域。光谱数据库也如雨后春笋般地建立起来。在面向各个领域的数据库设计与建设之时，综合各方面应用需求，挖掘高光谱数据应用的本质，进行面向应用的核心高光谱数据库设计，是高光谱遥感应用对高光谱数据库提出的迫切需求。

尽管我国在十五期间为加强对地观测信息源建设和提高现有遥感信息资源利用效率，建成了国家地理空间信息交换中心，形成了国家空间信息基础设施的主干网和一定数据服务能力。但是，无可争辩的事实是人们对遥感信息的认识和利用程度要远远落后于通过空间和航空系统收集信息源的速度。据估计，人们可能用到的遥感信息仅在全部获取信息的 5%左右，而深层次的信息开发则更少。这就是所谓信息“过剩”现象。其重要的原因在于技术系统，或对信息的处理、加工、分析技术的发展落后于信息获取技术的发展（陈述彭、童庆禧等，1998）。高光谱遥感数据相对一般遥感数据而言，更加具有海量数据的特点，它在时间序列上的积累和空间覆盖上的拓展，对相应的数据处理与分析能力提出了更高的要求。早期的光谱数据库主要应用在于提供标准的端元数据辅助影像分类，或者与通过光谱匹配实现地物识别。随着计算机技术的不断发展和遥感应用的不断深入，信息提取与挖掘的需求也不断深化。由于高光谱数据具有光谱维和空间维的连续性特点，这使得面向数据挖掘技术的高光谱数据库设计成为发展的方向之一。

本章将针对高光谱数据库的设计需求，提出高光谱数据库的光谱数据模型和影像数据模型，并根据数据挖掘和应用导向对高光谱数据库的设计进行了阐述。

## 3.1 高光谱数据库需求分析

到目前为止，据不完全统计，在世界各国已投入试验或运行的航空高光谱系统就有五十多台/套，已有近十颗卫星或航天器带有高光谱系统在空间运行或曾经运行过。EO-1/Hyperion，MODIS，ASTER，以及欧空局小型高光谱小卫星系统 CHRIS 和 ENVISAT 系统中的 MERIS 等航天高光谱遥感器的成功发射、试验和运行为航天高光谱应用起到了巨大的推动作用（童庆禧，2003）。随着高光谱遥感应用的不断深入，高光谱遥感数据获取日益频繁，随着高光谱遥感卫星的研制和应用，高光谱遥感数据量越来越大。高光谱数据的存储、管理、发布对高光谱数据库提出了迫切的需求。

数据库对于当今任何科学研究来说，都是至关重要的。数据库技术用于高效管理大量的已收集到的数据，并使的数据能够安全的长期保存，实现快速检索。这对于科学研究而言，显得尤为重要。考虑到高光谱遥感应用与众不同的特点，在进行高光谱数据库设计时，需要充分考虑到以下几个方面：

- 海量数据的存储结构

高光谱遥感最典型的特征之一，便是其细分的光谱波段导致图像数据量的剧增。高光谱遥感数据一个对象对应几十波段乃至上千波段不等的数据，每个波段的数据都可能需要单独提出以用作分析、计算。在实际应用中，对于高光谱影像数据的检索和读取往往有更加详细的要求，比如说对图像进行空间截取和光谱波段截取。在影像数据存储结构的设计时，一方面要考虑到大容量数据存储的高效性，另一方面也要考虑到进行操作时的灵活性。

- 高光谱数据的完整性

数据的完整性一般通过对数据表和表中的列添加约束来实现。广义的高光谱数据包含了影像数据、配套地面测量光谱数据、配套测量参数还有针对应用的属性参数等等。由于针对不同的应用目的，测量参数和属性参数在命名、格式、值域、数量上都不尽相同。对应的数据表的结构难以提炼成统一的实体关系模型。这种半结构化的一套数据为数据约束增加了难度。因此，在维护这些数据的过程中，需要通过对数据模型的设计，来保证数据的完整性。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。

- 图谱合一的特点

高光谱数据中图像和光谱是紧密联系的一个整体。从平面图像空间提取光谱曲线，直观的提供地物目标光谱信息，是高光谱遥感的重要特点，也是高光谱数据库应该保存的重要特点。高光谱影像数据结构，即图像立方体，本身就是一个图谱合一的数据模型。在数据库存储中，如何保存图像立方体的特点，将变化性很强的图像光谱维数据与图像空间维数据很好的结合在一起，需要对现有的数据类型进行改进。

- 光谱匹配算法的融合

在高光谱遥感应用中，光谱维信息的应用是高光谱遥感的重要特点，它是地物识别的物理基础。高光谱数据库的一个重要作用便是辅助地物光谱识别。基于高光谱遥感在众多窄波段获取数据的特点，可以由已知地物类型的反射光谱，通过波形或特征匹配比较来达到直接识别地物类型的目的。人们对地球上的各种物质已经做了长期的研究，逐步认识了电磁波与地物的相互作用机理；长期的高光谱试验也收集了大量的实验室标准数据，建立了许多地物标准光谱数据库；在高光谱应用研究中，人们也已经解决了图像数据的光谱重建的难题。在这些研究工作的基础上，我们已经具备了从图像直接识别对象的条件（白继伟，2002）。而光谱匹配则是实现地

物识别的关键技术之一。所谓的光谱匹配是指通过研究两个光谱的相似度来判断地物的归属类别。在高光谱数据库中融合这些算法，是强化高光谱数据库的重要手段之一。

- 光谱特征参量化

高光谱遥感数据大量的光谱波段为我们了解地物提供了极其丰富的遥感信息，这必然有助于我们完成更加细致的遥感地物分类和目标识别，然而波段的增多也必然导致信息的冗余和数据处理复杂性增加。当光谱维数增加时，其特征组合更是成指数方式增加。因此，光谱特征空间的减少和优化显得十分重要。在高光谱数据库中，辅助地物识别、定标、农林业等行业的特定应用都需要对影像或者地物光谱进行参量化，实现特征提取与检索。

- 数据的快速检索与动态浏览

高光谱数据的辅助数据和属性复杂，种类繁多，而且随着不同的应用领域显示出不同的特点。在进行数据检索过程中，除了针对影像的属性信息和图像内容检索，还要考虑对纷繁复杂的应用中共同信息内容的精炼与萃取，并实现对数据的综合信息检索。另外，目前的影像数据浏览多为静态图片，对于高光谱遥感影像而言，静态的缩略图浏览已经不能满足高光谱多波段的特点和应用的需求，动态的影像数据提取与浏览也应该成为高光谱数据库设计的考虑因素。

- 数据的数据挖掘

高光谱遥感技术的不断发展和应用，使得积累的遥感数据越来越多。但数据的增多，随之而来的问题就是如何从海量的数据中抽取出具有决策意义的信息，更好的对科学研究和行业应用带来促进和发展，这就需要数据挖掘。从我们对数据管理需求的角度看，可以划分两大类：一是联机事务处理（OLTP）应用，即对光谱和影像数据检索、查询、上传、下载；二是联机分析处理（OLAP）与辅助决策，既通过数据库对联机数据的强大管理功能，对遥感影像数据及其配套数据、相关的参量信息进行分析、模式发现、异常提取等等，从而为决策提供支持。尽管目前众多的数据库中包含了大量的有价值的历史数据，但如果我们直接去看这些数据是没有任何实际意义的，必须有方便有效的工具能够很方便的对其中的数据进行分析处理。（张大昕等，2003）。从遥感影像数据库和光谱数据库的发展现状可以看出，目前大部分遥感数据库主要的应用，还在于对遥感数据的检索和浏览，遥感影像和光谱数据的联机分析和数据挖掘还比较少。然而，不同行业的应用对于数据挖掘的要求又不尽相同，因此，在数据库设计中，要考虑对数据挖掘进行设计。

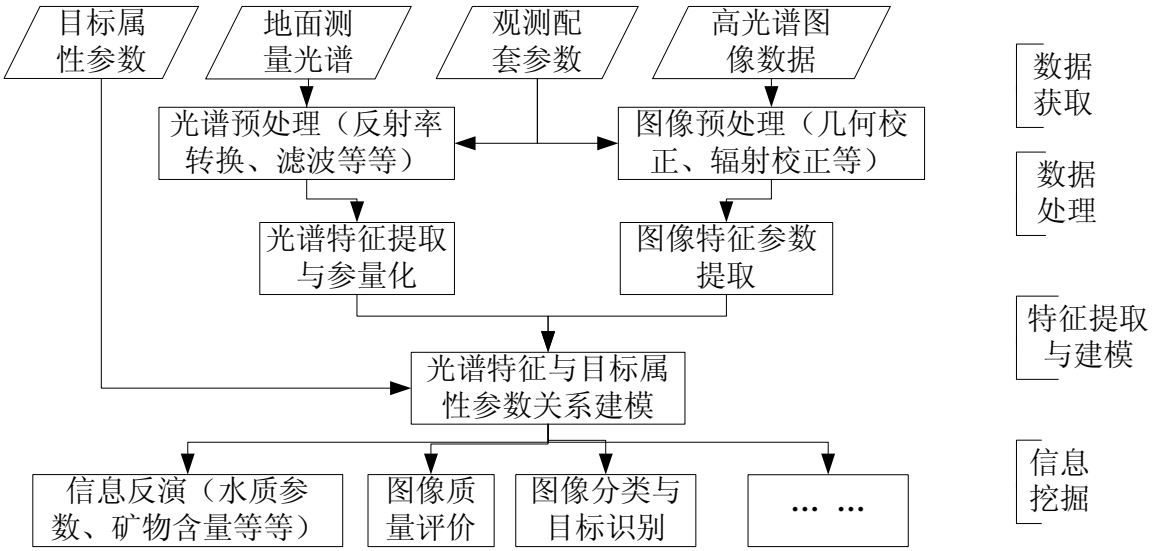
- 网络共享

从第二章光谱数据库和遥感影像数据库的发展趋势可以看出，数据库的网络化的横向发展和面向数据挖掘的纵向发展成为趋势。新一代的网格技术则是把整个网络整合成一个虚拟的巨大的超级计算环境，实现计算资源、存储资源、数据资源、信息资源、知识资源和专家资源的全面共享。目的是解决多机构虚拟组织中的资源

共享和协同工作问题。在网格环境中,不论用户工作在何种“客户端”上,系统均能根据用户的实际需求,利用开发工具和调度服务机制，向用户提供优化整合后的协同计算资源,并按用户的个性提供及时的服务。按照应用层次的不同可以把网格分为 3 种：计算网格，提供高性能计算机系统的共享存取；数据网格，提供数据库和文件系统的共享存取；信息服务网格则支持应用软件和信息资源的共享存取（Jagatheesan A，2003）。在高光谱数据库的设计中，首先应该能够实现数据库的共享存取。

### 3.2 高光谱数据库数据流程分析

高光谱数据库的数据流程一定程度上决定了高光谱数据库的应用框架体系结构设计原则。高光谱数据库的存在和发展与高光谱遥感应用是离不开的。传统的光谱数据库和遥感影像数据库的数据流程主要是基于数据管理的目的，其核心在于数据存储与发布，着重点在于数据库，而不是高光谱。在分析高光谱遥感应用的数据流程之后，我们能够将重心转移到高光谱遥感之上，通过高光谱遥感数据流程来重组高光谱数据库数据流程。在利用数据库对高光谱遥感进行辅助应用时，一般的数据流程，如图 3.1 所示：




图表 3.1 高光谱遥感应用数据流程

这是一个相对完整的高光谱遥感数据应用流程，并不是所有的这些数据流都必须发生，才能到达高光谱遥感应用目标。从传统的数据库观点出发，数据库的作用仅仅停留在数据获取与数据处理之间,这是局限于原有的关系数据库技术基础之上。随着数据库技术的不断发展，基于对象的数据模型能够将特定的数据与操作整合在一起，重组数据库流程。高光谱遥感应用数据流程对于高光谱数据库的设计原则的重大启示有以下几点：

- 发展数据模型

由于关系数据库技术的局限，数据库对于遥感影像及其配套参数等大型非结构化数据的应用往往捉襟见肘。数据库的应用潜力往往取决于数据结构模型，在对象

关系数据库技术发展至今，面向对象的数据结构模型应用到数据库中，使得数据库的能力大大提升。在高光谱数据库中，可以将光谱数据和影像数据存储方式从 BLOB 中解脱出来，通过对数据模型的优化将数据的配套参数融合到存储结构中。

- 扩展应用接口

数据存储和管理一直是数据库技术的优势，基于关系代数的 SQL 语言是数据库查询与检索应用的核心，它能够灵活有效的处理和检索关系型数据。但是这种非过程化的语言限制了在数据库端的应用。PL/SQL 等的过程化数据库查询语言能够将过程性结构和数据库查询语言无缝集成在一起，便于在数据库端进行编程，为数据库的应用提供更加灵活的接口。在高光谱数据库中，则可以把基本的非应用性目的的数据操作作为数据模型的方法集成到数据模型中，为专业性的应用开发提供程序接口。

- 将表观数据发布变为深层数据挖掘

由于关系数据库给于人们的印象太深刻，以至于人们对于数据库的作用的理解，除了数据存储与管理之外，大部分还停留在辅助数据发布之上。以在目前已经商业化运营的航空高光谱遥感器为例，如 AVIRIS（http://aviris.jpl.nasa.gov/cgi/flights.cgi）和 HYMAP（http://www.aigllc.com/data_acq/intro.htm），在提供影像数据库服务时，他们仅仅是把获取的高光谱图像数据和配套的参数经过数据处理之后，存放在数据库中。提供给用户检索浏览方式，仍然采取是静态的，即通过高光谱遥感数据做成真彩色的 jpg 格式缩略图片。USGS 光谱数据库从 1995 年开始将其测量的岩石矿物光谱数据在网上公开提供下载，至今已经是第五个版本（R. N. Clark，2003）。USGS 光谱库提供了光谱曲线、光谱图片以及对应的样品描述。而 USGS 的 SPECLIB 发布，还没有用到数据库技术，只是通过 ftp 和 web 的方式提供原始数据的下载。Google Earth 和 MSN Virtual Earth 的应用核心也是基于 GIS，对于遥感影像的利用也是停留在数据发布与浏览层面。NASA World Wind 则将 MODIS，SRTM 等数据集成到一起，在数据发布层面更进了一步。然而，遥感数据，尤其是高光谱遥感影像数据的价值，远远不应局限于通过数据发布和浏览来体现。对于光谱数据和遥感影像数据的深层挖掘，并在此之后的信息发布是数据库设计时要考虑的重要环节。

## 3.3 应用框架与体系结构设计

### 3.3.1 高光谱数据库应用框架设计

高光谱数据库的应用框架设计原则，在于体现高光谱数据库的扩展性能，网络性能，和跨平台性能。网络环境中的数据应用框架，经历了从文件服务器到两层客户端/服务器再到多层客户端/服务器的转变（杨勤，2001）。文件服务器模型用以解

决个人 PC 和工作站的数据和外部设备共享，但是无法提供多用户应用的数据并发性，并且容易导致网络堵塞。两层客户机/服务器模型则是一种分布式计算模式，主要包括数据库服务器、客户应用程序和网络。它将应用的处理需求分开并共同实现，在这种模式中，所有的业务逻辑和形式逻辑驻留在客户端，而服务器端则称为数据库服务器，负责各种数据的处理和维护。随着应用范围扩大，这种模式显得缺乏灵活性，同时可靠性降低、维护费用增高，因此在两层结构中增加了一层 web 服务器用来承担浏览服务器和应用服务器。这种三层客户端/服务器模式将应用功能分为表示层、功能层和数据层。通过对应用服务器的扩展能够实现系统的扩充需要，可以用较少的资源建立伸缩性比较强的系统。和以前的结构相比，三层 C/S 结构具有更灵活的硬件系统构成,对于各个层可以选择与其处理负荷和处理特性相适应的硬件。合理地分割三层结构并使其独立，可以使系统的结构变得简单清晰，这样就提高了程序的可维护性。三层 C/S 结构中，应用的各层可以并行开发，各层也可以选择各自最适合的开发语言，有利于变更和维护应用技术规范。按层分割功能使各个程序的处理逻辑变得十分简单（刘卫忠，徐重阳等，2000）。一般而言，三层结构与过去的两层结构相比有如下优点：

（1）进程管理：通过对服务进程的管理，使得在正常情况下，能用尽量少的服务进程处理尽量多的请求，减少进程的启动/终止次数。在峰值情况下，控制服务进程的总数，使得服务器在设定的负载下工作，不被压跨。总之，通过中间件对服务进程的有效管理，可以使系统在额定的功率下稳定工作，当请求服务的数量超过了服务器的处理速度时，中间件会把请求排队进行缓冲。

（2）保持和复用数据库连接：服务进程访问数据库都要和数据库建立连接，如打开和关闭数据库等。中间件通过采用长驻服务进程的手段，使得与数据库的连接被保持和复用，从而大大减少与数据库连接的次数和时间。

（3）优化了系统结构：将系统分为三层（或多层），业务逻辑放在应用服务层，软件的维护集中在应用服务层，客户端的维护就相对简单多了，有利于软件维护及系统管理。

（4）提高了应用系统的安全性：将客户端与数据库隔离起来，客户端无权限直接访问数据库，有利于安全管理，可有效防止恶意攻击。还可以利用中间件的安全管理特性进一步加强权限控制管理。

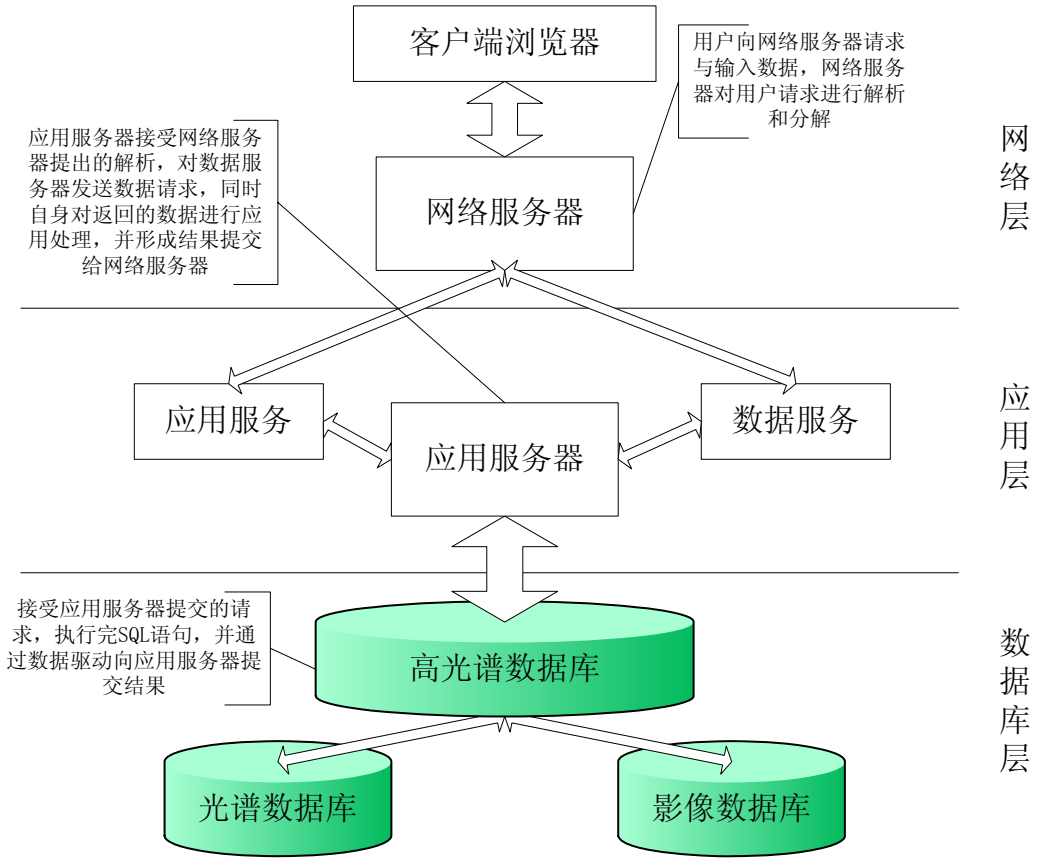
（5）卓越的扩展能力：若要提高系统性能、处理速度，可增加应用服务器，分担一部分应用服务工作即可，而原来的应用服务器几乎可以不动。

（6）减少网络数据流量和提高数据库响应速度：两层应用体系结构中客户机直接（或通过存储过程）访问数据库，会造成数据库访问瓶颈及网络瓶颈，从而降低了整个系统的性能。三层应用体系结构中，应用服务层的引入有效地解决了网络瓶颈和数据库连接数过多引起数据库性能下降的问题。应用服务层往往有多台服务器，可有效地解决客户机访问服务层瓶颈。应用服务器与数据库服务器（物理距离

很近）可方便地采用宽带网连接，不会产生与数据库服务层网络瓶颈。

（7）提高系统性能：三层应用体系结构能更好地调整应用体系，还可利用中间件的特点来选择路由、平衡负载，提高整个系统的性能。

总的来说，三层应用体系结构使系统的性能、安全性、扩展性有了很大的提高，也方便了系统的维护和管理。因此，在高光谱数据库系统的应用框架设计上，我们采取了如图 3.2 所示的三层客户端/服务器框架体系：




图表 3.2 高光谱数据库应用体系框架

- 网络层：

网络层最主要的是实现对多用户和网络用户的数据和应用展示。在 WEB 技术的发展促进下，客户端的应用一般都能够通过通用浏览器进行实现。用户通过 WEB 界面向网络服务器提出对数据的请求和操作，网络服务器负责解析这些请求，并将指令传递到应用服务器。

网络服务器按是否能同时处理多个客户请求，可分为一次只能处理一个客户请求的循环服务器及一次能处理多个客户请求的并发服务器。循环服务器采用单进程结构，一次只能处理一个客户连接，结构简单，易于实现。对访问服务的客户数量不大，服务处理需时短的场合，采用该类型结构服务器非常合适。并发服务器通过创建多个子进程，分别与多个客户机请求对应，为其提供同步处理。响应速度快，但需更多的系统资源开销，适用于客户请求密集、服务处理时间较长的情况（郑庆良，张翔等，2004）。在高光谱数据库系统中，网络服务器为不同规模的客户群提供

服务，一般而言，在大规模应用之前，数据的并发处理并不是高光谱数据库系统所重点关注的。对高光谱数据库而言，首要的设计要求是效率，即服务器的响应速度满足客户要求。第二，服务器在向客户机提供服务的同时，尽可能少地占用系统资源，尤其是长时间占用，以提高系统资源的利用率。实际上，这两者往往相互牵制，实现时应追求两者的平衡、统一。

- 应用层：

应用服务器已经是网络应用中关键的一种中间件技术。在网络应用开发的早期，大家都使用 Web 服务器提供的服务器扩展接口，使用 C 或者 Perl 等语言进行开发，例如 CGI、API 等。这种方式可以让开发者自由地处理各种不同的 Web 请求，动态地产生响应页面，实现各种复杂的 Web 系统要求。但是，这种开发方式的主要问题是对开发者的素质要求很高，往往需要懂得底层的编程方法，了解 HTTP 协议，此外，这种系统的调试也相当困难。在第二阶段，开始使用一些服务器端的脚本语言进行开发，主要包括 ASP、PHP、Livewire 等。其实现方法实质上是在 Web 服务器端放入一个通用的脚本语言解释器，负责解释各种不同的脚本语言文件。这种方法的首要优点是简化了开发流程，使 Web 系统的开发不再是计算机专业人员的工作。此外，由于这些语言普遍采用在 HTML 中嵌入脚本的方式，方便实际开发中的美工和编程人员的分段配合。对于某些语言，由于提供了多种平台下的解释器，所以应用系统具有了一定意义上的跨平台性。但是，这种开发方式的主要问题是系统的可扩展性不够好，系统一旦比较繁忙，就缺乏有效的手段进行扩充。这种方式不利于各种提高性能的算法的实施，不能提供高可用性的效果，集成效果也会比较差。（张志，2005）

为了解决这些问题，应用服务器便得到发展。应用服务器是独立的进程，对业务进行处理，并进行事务管理，将其中的所有数据操作转给第三层，也就是数据处理层的数据库服务器。应用服务器体系结构的核心在一般的网络服务器和数据库服务器之间，增加专门的应用服务器来完成业务处理。

应用服务器层的增加，使得高光谱数据库系统的扩展性得到增强。在将来高光谱数据库系统需要进行规模扩大时，可以仅仅增加几台新的服务器，安装应用服务器软件，进行恰当的配置，通过应用服务器的负载均衡能力，将用户发来的请求，恰当地分配给各个应用服务器，就可以使各个服务器分别负担系统的负载。这样可以方便地进行扩充，不需要进行应用的重新开发和调整等高风险性的操作。

对于高光谱数据库系统而言，数据库处理是最关键的步骤，各种数据库操作的步骤中，数据库的连接和释放往往又特别耗时。集成在应用服务器中的数据库连接池（Connection Pool）的技术，即在系统初起，或者初次使用时，完成数据库的连接，而后不再释放此连接，而是在处理后面的请求时，反复使用这些已经建立的连接。这种方式可以大大减少数据库的处理时间，有利于提高系统的整体性能。

另外，应用层还承担了对各个领域应用的业务组件开发，用于实现业务逻辑。

- 数据层

将数据库和应用逻辑隔离，可以通过对象关系模型对数据库进行架构设计，同时，也可以完全用面向对象的思想去设计实体，而不用首先去考虑表的结构和关系。这样大大降低了应用层对数据库的耦合度，方便了数据库的移植。在高光谱数据库中，数据层的设计尤为重要。高光谱数据的特殊性决定了其数据访问方式和检索方式需要重新审视，这不仅仅要考虑海量数据在数据库中的存储，还要考虑对于大对象的读取效率，更要考虑对于将来存储空间的扩展和分布式数据库的拓展。另外，为了提高对数据的利用效率，数据层还应该能够相应用户对数据的一些简单操作，在数据库端通过存储过程完成数据的简单处理。在数据层设计存储过程完成数据的操作，可以改善系统性能，降低网络流量，方便系统维护和功能扩展，对于数据库系统的安全性而言，也是非常有益处的。

### 3.3.2 系统结构设计

高光谱数据库系统结构设计应该遵循数据库设计的一般原则，即：

- 完整性原则（Integrity）

数据库中的数据值应满足指定的约束, 且对数据库进行更新后仍然保持这种性质，称为数据库具有完整性。数据库完整性是指数据库数据的正确性和一致性。完整性考虑的参数有：实体完整性、域完整性、参照完整性；强制和有效执行约束——触发条件、测试出违反完整性约束时采取的措施（动作）等。

- 一致性原则（Consistency）

如果多个用户同时、以同样方式、对同一数据查询，数据库的回答结果是一样的，那么称数据库对多用户具有一致性。一致性考虑的参数有：修改数据的方法：执行强制修改前等待时间，执行强制修改时其它活动用户的等待时间，修改的响应范围，数据修改的算法等；同时要求发生修改同一数据的用户数目（并发度）。

- 可靠性原则（Reliability）

数据库的可靠性包括：故障发生的可恢复性（ Recoverable ）、故障恢复所需要的时间和故障发生的频率。当数据库发生故障时，具有恢复数据库完整性的能力，称数据库具有可恢复性。可恢复性的设计过程包括建立一个检查系统，防止事务和数据的丢失；当故障发生时，在合理时间内把事务和数据的状态恢复到故障发生前的情况。一般数据库管理系统都具有这一功能，如数据库备份、 镜像、日志等。

- 安全性原则（Security）

数据库安全性是指对数据库有意或无意的泄漏、修改或丢失的保护能力。设计数据库安全性的主要目的是以最小的代价防止对数据的非法的使用。实现方法是控制对数据库数据的访问，DBMS 提供控制访问的功能，如创建子模式或存储过程、授权/收权、授予角色、用户确认、审计等功能。

- 效率原则（efficiency）

主要是指计算机资源的利用和系统的响应时间。效率原则的目标包括三个部分：模式设计的合理性，使每个应用的执行时间最小，联机响应时间不超过设计的要求值；查询优化，最小的数据传送量和通信次数、最优处理顺序、最少 I/O 次数和操作量、最优访问路径；合理的数据冗余，处理时间/存储与执行空间的最优选择。例如在模式中保留一定的冗余（属性的冗余）可以减少检索的响应时间，但它增加了新的开销和一致性问题，这根据数据库应用的具体需求而定。（张雄飞，2003）

在确定数据库结构设计原则之后，需要根据需求分析和数据流程分析对数据库进行结构划分。高光谱数据库由两大数据库（高光谱影像数据库和地物光谱数据库）、一个模型库和一个方法库组成。

高光谱数据库的核心部分是高光谱图像数据库系统；为了适应不同的应用需求，有必要将一些主要的模型和方法内迁到数据库中。这对于扩展数据库的功能，强化数据库在存储、搜索等各方面的应用，有着很大的帮助。

⑴ 高光谱影像数据库系统：这是高光谱数据库系统的核心所在，它主要负责高光谱影像光谱数据以及其对应的属性数据的存储、查询、浏览、添加、修改、删除等基本操作。其中存储的可以是地面成像光谱辐射计或者航空、航天高光谱成像仪采集的图像光谱数据，其兼容性保证各种光谱波段设置的高光谱成像仪采集的样本均可以存储。各种级别的遥感影像都可以存储在高光谱影像数据库中，针对其不同的处理过程，可以生成各种不同级别的影像。一般而言，针对应用的数据是经过纠正、处理的标准图像光谱数据，同时其相应的属性中不但可以包括地学、测量方面的属性数据，还可以面向需求扩展到经纬度信息乃至分类信息。

⑵ 地物光谱数据库系统：高光谱数据库系统是来源于已有的光谱数据库系统之上的，而且现有的很多高光谱航空、航天成像仪还是需要地面调查，取得地面光谱的配合。

⑶ 高光谱数据模型库：为了满足科研的需要，作为一个成熟的数据库应用系统，光有数据库基本的管理功能是不够的。尤其是高光谱数据中富含了很多信息，相应的处理模型、方法也已经作了很多的研究，将一些常见的模型内置到数据库中，为用户提供常见的数据模型应用，能够扩大高光谱数据库系统的应用范围，给高光谱研究应用带来极大的便利。

⑷ 高光谱方法库：高光谱方法库是将各种数据操作中常用的操作方法提炼出来，在对高光谱数据库进行针对行业应用级别的二次开发或者应用时，可以复用这些方法，用于产生新的模型。

一景 1024×1024 分辨率的高光谱图像数据就将含有 1M 条光谱曲线，如果每条光谱曲线有 100 个波段，就将是 100M 的数据量，明显可以看出高光谱数据是典型的海量数据。同时，高光谱图像光谱数据具有很强空间几何信息，相互之间具有复杂的关系，多景同类图像之间更会有复杂的关系，这就给利用数据仓库整合数据、

然后进行数据挖掘提供了良好的基础。数据挖掘正是当前数据库发展的前沿和热点，其目标就在于数据积累的基础上，深入探索数据间的关系与规律，从而得到进一步有价值得信息和结论。由高光谱影像数据库、地物光谱数据库、高光谱模型库以及高光谱方法库构成了一个面向高光谱数据应用的、具备数据挖掘能力的完整的高光谱数据库系统。

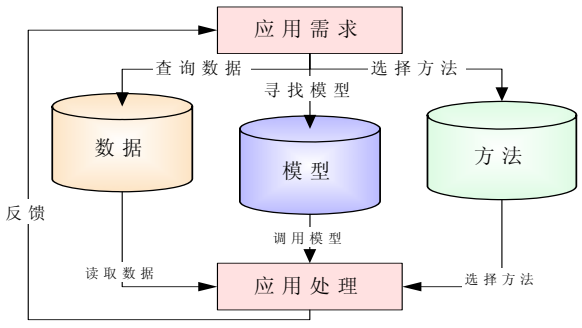
### 3.3.3 数据挖掘与数据库的耦合设计

根据数据挖掘与数据库的耦合程度可以分为零耦合、松散耦合、半紧密耦合和紧密耦合四种结构。零耦合是指数据挖掘和数据库没有任何关系，输入数据是从文件中读取的，存放结构也是存到文件中，这种结构一般较少使用。松散耦合是利用数据库作为数据挖掘的数据源，其结果写入文件或者数据库，但是不使用数据库提供的数据结构和查询优化方法。半紧密耦合是指部分数据挖掘原语出现在数据库中，例如 sorting, indexing, aggregation , histogram analysis, multiway join, 以及一些统计方法，如 sum, max, min, standard deviation 等等。机密耦合是将数据挖掘集成到数据库中，作为其中的一个组成部分，并利用数据库的数据结构、索引模式查询处理过程对挖掘查询进行优化。目前的发展趋势是紧密耦合的系统结构。（邵峰晶，2003）

从高光谱数据库的的数据挖掘需求出发，紧密耦合的数据挖掘能够高效的利用数据结构和数据库查询机制，能够节省数据挖掘在其它耦合状态下对于海量数据的读取检索所带来的时间损耗。而高光谱数据库的应用在不同领域的特点使得紧密耦合又存在扩展上的局限，因此，在面向数据挖掘设计时，需要对挖掘算法进行筛选，将应用领域无关性的算法集成在数据库中，将相关性较强的挖掘算法在应用服务器端实现。

### 3.3.4 应用逻辑模块设计

高光谱数据库的应用逻辑，虽然随着不同的应用领域体现出不同的特色，但是，在经过抽象化之后，能够得出如图 3.3 所示的宏观应用逻辑示意图。




图表 3.3 高光谱数据库应用逻辑图

将高光谱遥感应用和高光谱数据库的应用逻辑进行高度抽象之后，得到三个基本的逻辑模块：数据、模型和方法。

高光谱数据模块，主要包括两个数据模型，光谱数据模型和影像数据模型。该模块响应应用的数据需求，在数据库中搜索、截取、整合之后，将数据提供给模型或者方法模块，进行处理分析。数据模块应对的是各种各样的数据需求，有图像、光谱、属性参数、环境参数等等。

高光谱模型模块负责存储高光谱应用模型，这是构建和管理模型的核心部分，是高光谱数据库系统中最复杂与最难实现的部分。模型模块中主要存储的是能让各种应用问题共享或专门用于某特定问题的模型基本模块或单元模型，以及模型之间的关系。对应于那些结构性比较好的问题，其处理算法是明确规定了的，表现在模型上，其参数值是已知的。对于非结构化的问题，有些参数值并不知道，需要使用数理统计等方法估计这些参数的值。

高光谱方法模块是以程序方式管理和维护各种决策常用的方法和算法，是存储、管理、调用及维护决策各部件要用到的通用算法、标准函数等方法的部件，方法一般用数据库中的存储过程或者函数实现，包括基本数学方法、统计方法、预测方法、计划方法、优化方法，同时也包括具有高光谱特色的算法，比如包络线去除等等。

## 3.4 高光谱数据库数据模型设计

数据库技术从诞生到现在,在不到半个世纪的时间里，形成了坚实的理论基础、成熟的商业产品和广泛的应用领域，吸引了越来越多的研究者加入，使得数据库成为一个研究者众多且被广泛关注的研究领域。随着信息管理内容的不断扩展和新技术的层出不穷，数据库技术面临着前所未有的挑战。面对新的数据形式,人们提出了丰富多样的数据模型，同时也提出了众多新的数据库技术（孟小峰，2004）。数据模型决定了数据库中数据的组织、描述与存储。高光谱遥感的应用需求和特点对数据库技术提出了新的挑战，因此，对于数据模型的应用与改进成为高光谱数据库重要的研究内容。

### 3.4.1 地面光谱数据模型

#### 3.4.1.1 地面光谱数据特点

地面光谱数据一般理解为光谱反射率数据，其表现形式为一组两列的二维数组。一列数据为波段值，另一列为对应的反射率值。然而，广义的地面光谱数据还包括各种光谱参数，例如光谱仪的各项性能参数、参考板的定标数据、测量方式参数、测量环境数据等等。在应用层面上的地面光谱数据还包括测量的相关理化参量、

组分含量，样品颜色、特征等等数据。随着不同的应用领域，或者同一领域不同的应用项目，这些参数的数量、内容和数据格式都是变化的。

对于地面光谱数据的应用处理，主要集中体现在参量化、光谱截取和重采样等等。在对光谱数据进行具体的操作上面，新的模型算法层出不穷。例如：二值编码（Goetz，1985），导数光谱模型（Johnson,1996），光谱曲线的包络线去除（白继伟，2003），光谱重排（耿修瑞，2004）等等。

因此，对于地面光谱数据对象建模，需要充分考虑光谱数据本身的特点，也要考虑其在操作方法上面的成熟性和灵活性。

3.4.1.2 数据库中地面光谱的传统存储方式

在数据库中，表是存储的基本单位，在早期对于光谱数据的存储考虑更多的是对于数据表空间利用的设计。由于早期磁盘存储技术的限制，对数据的冗余有着比较苛刻的要求。因此，在数据库设计中，剔除数据的冗余是主要方向。白继伟等在FOXPRO 基础上建立的光谱数据库在六个部分数据中建立了 47 个表。（白继伟，2002）在设计中充分考虑了数据库的扩展性和数据库对象分离，但是在大量的数据检索中，这种结构会引致频繁的多表链接，从而降低检索效率。

张雄飞等在经过长期研究后，提出在关系数据库中，合理而优化的高光谱数据基本的存储方式为：光谱数据表组＋属性数据表组。这一方法代表了数据库在地面光谱的传统存储方式。

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。这种传统存储方式有以下几个突出的优点：（张雄飞，2003）

- 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。
- 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。
- 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。
- 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。

傅莺莺等在国家典型地物波谱数据库的建设过程中，提出了波谱数据的文件存储格式。波谱数据总共分为四类文本文件：A 文件用于存储光谱文件，其形式为每行两列制表符间隔的光谱数据；B 文件用于存储光谱属性文件，用【 】间隔开元数

据和属性值；C 文件和 D 文件是对于植被光谱 BRDF 数据的扩展，用于共享除了时间和观测角度之外的其它完全相同的属性数据（Fu Yingying, 2004）。

由此可以看出，对于地面光谱的传统存储方式，主要是将光谱数据和属性数据完全分离，通过关键字进行链接。根据这一模型运用到比较常见的 ORACLE 数据平台，对光谱数据存储有以下三种模式:（张雄飞，2004）

- 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

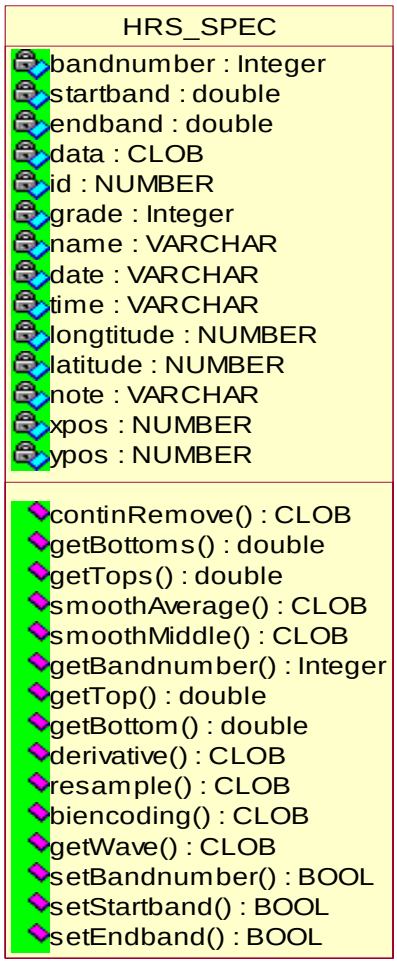
这是最容易理解的一种方式。由于高光谱遥感数据量大，类型繁多，所以一种类型的数据对应一个表是最简单最容易理解的方式。既可以以某种高光谱仪器的波段为单位来建表，也可以以对象为单位建表等。它们的共同点是会有一些比较固定的数据维，或是波段数固定，或是光谱数固定，这样就可以根据该维来设定该表的结构。这样实施，数据存储量不会冗余和浪费空间，查询操作的效率高，开发容易，但是缺点是扩展性差。数据库开发不应该允许用户随着数据量的增大经常进行建表工作。所以这种方式仅限于一些特殊要求或者数据量比较固定的情况下使用。这种方式易实现开发。

然而，这一传统的数据模型仅仅考虑到对于光谱数据的高效存储和快速读取。随着应用需求的拓展，光谱数据的基本操作也成为高光谱数据库的基本功能，同时，通过面向对象的方法，使光谱数据和属性数据在模型内统一，来实现对数据的完整性维护也成为可能。

#### 3.4.1.3 地面光谱数据模型设计

在高光谱数据库中，地面光谱数据是以 HRSSPEC 对象存储的。HRSSPEC 对象

除了对光谱数据进行存储，还包括光谱数据的基本属性以及对应的操作方法，如图3.4。




图表 3.4 地面光谱数据概念模型

#### 3.4.1.3.1 HRSSPEC 属性

如表格3.1所示，HRSSPEC属性主要包括了对光谱数据本身的描述，包含在数据库中的唯一标识、光谱数据的等级、观测目标名称、光谱获取时间、关侧地点经纬度等等。如果该光谱是作为像元光谱或者是辅助图像定标等应用，图像记录外键不为空值，则有对应的图像中的坐标属性。

表格 3.1HRSSPEC 模型属性

| 属性名称 | 属性数据类型 | 属性说明 |
|---|---|---|
| BANDNUMBER | INTEGER | 波段数 |
| STARTBAND | DOUBLE | 光谱起始波长 |
| ENDBAND | DOUBLE | 光谱终止波长 |

| | | |
|---|---|---|
| ID | N10 | 光谱数据的唯一标识，主键 |
| GRADE | N1 | 光谱数据等级 |
| NAME | N10 | 光谱名称 |
| DATA | CLOB | 光谱数据 |
| DATE | VA8 | 光谱获取日期 |
| TIME | VA8 | 光谱获取时间 |
| LONGITUDE | N7,4 | 经度 |
| LATITUDE | N6,4 | 纬度 |
| NOTE | VA64 | 说明 |
| XPOS | N10 | 在图像中 X 座标 |
| YPOS | N10 | 在图像中 Y 座标 |

#### 3.4.1.3.2 HRSSPEC 方法

如表格 3.2 所示，HRSSPEC 模型包含了对于光谱对象的基本操作。这些操作方法在高光谱数据库中经常被应用来实现对光谱的匹配、辅助端元提取等等。

表格 3.2 HRSSPEC 模型方法

| 方法名称 | 返回类型 | 方法功能说明 |
|---|---|---|
| smoothAverage（） | CLOB | 均值滤波 |
| smoothMiddle（） | CLOB | 中值滤波 |
| getBottoms（） | VARRY | 获取吸收位置 |
| getTops（） | VARRY | 获取反射峰 |
| getBandnumber（） | NUMBER | 获取光谱波段数 |
| getWave（） | CLOB | 获取光谱波长参数 |
| getTop | NUMBER | 获取指定区间波峰 |
| getBottom（） | NUMBER | 获取指定区间波谷 |
| continRemove（） | CLOB | 对光谱曲线进行包络线去除 |
| derivative（） | CLOB | 对光谱曲线进行求导 |
| resample（） | CLOB | 对光谱曲线给定参数的重采样 |
| biencoding（） | CLOB | 对光谱曲线进行单门限二值编码 |
| setBandnumber（） | BOOL | 设置波段数 |
| setStartband（） | BOOL | 设置起始波段波长 |
| setEndband（） | BOOL | 设置终止波段波长 |

### 3.4.2 高光谱影像数据模型

3.4.2.1 高光谱影像数据特点

如何存储和管理，特别是图像信息的检索和相关辅助数据（试验数据、特征描述等）相互对应的关系，是目前图像数据库研究中需要解决的问题，也就是说，如何用计算机技术、数据库技术、信息处理技术对图像数据库模型进行研究，并建立相应地图像数据库来满足实际应用的需要（纪钢，2003）。建立高光谱影像数据模型的难点在于高光谱影像数据本身比较复杂，包括影像数据、影像特征和辅助信息等等。

高光谱影像数据通常以三种格式排列：BIP（波段按像元交叉）、BIL（波段按行交叉）和 BSQ（按波段顺序）格式。设 P、L、B 分别表示像元维、扫描行和波段维，以三维数组表示图像 D 有：BIP 格式为 D（P, L, B），其中波段维 B 为最低维，扫描行 L 为最高维；BIL 格式为 D（P, B, L）；BSQ 格式为 D（P, L, B）。通常成像光谱仪获得的数据流是 BIP 格式的，也有些面阵成像光谱仪以 BIL 格式获得数据。（白继伟，2002）

因为高光谱图像数据波段很多数据量很大，数据存储的格式对数据处理的效率影响较大。我们知道，在数据读写过程中，数据在存储空间上的方法很大程度上影响着访问效率。例如，要分析图像像元的光谱特性，提取光谱曲线，需要访问图像某像元所有波段的数值，这时以 BIP 格式存储图像比较合适。如果这时图像以 BSQ 格式储存，则需要从磁盘多个不同的地方读取数据，对于波段数很多而图像数据又很大的成像光谱数据来说，将花费大量的时间。如要对图像进行空间分析，如作空间滤波，只能对图像同一波段内的数据进行处理，此时以 BSQ 格式存储图像比较合适。BIL 格式一般是遥感器获取数据的初始格式，介于 BIP 和 BSQ 格式之间，它适合于在光谱空间对像元进行操作，但是对于具体应用来说，即不利于空间分析也不利于光谱分析。由于高光谱数据库侧重于对高光谱影像的光谱维信息挖掘，因此，采用 BIP 格式作为高光谱原始图像数据的存储方式。

在确定数据存储方式之后，需要根据影像特征等对具体的数据模型进行设计分析。影像特征主要是描述图像实体的形状、颜色、纹理和空间关系等。对于高光谱影像而言，影像特征除了上述空间信息特征之外，还包括光谱为的信息特征，例如波长范围、各波段反射率、吸收特征等等各种光谱信息特征。因此，高光谱影像特征集成了空间维和光谱维的信息特征，这对于高光谱影像数据库的存储与检索也是非常有意义的。

另外，高光谱影像数据的辅助信息对于高光谱影像的应用也有非常重要的作用。高光谱遥感影像的应用离不开辅助信息，辅助信息主要包括： 遥感器参数、测

量参数、环境参数和相应的地面辅助数据；遥感器参数主要包括遥感器型号、CCD行列数等等；测量参数包括飞行姿态姿态参数等等；环境参数包括太阳高度角、能见度、风速等等。

从这几类数据不难看出，图像数据内容难以准确描述，属非结构化的，而且其中的空间对象及其关系自身不包含语义信息，如果直接将语义信息同空间对象及其关系相联系将严重限制图像信息的使用。因为同一对象可以有多种解释，在不同时期使用方式可能不同，其解释是近似的，它随着图像识别技术的发展而发展。同时，图像数据库还必须支持图像查询和空间推理，由于解释本身的不确定性，系统必须提供特定的领域知识，以便用户逐步精炼其要求，这就要求模型是可进化的。这些使得传统的3级模式数据库系统结构无法管理图像数据，必须扩充到5级模式，既：用户视图、语义特征视图、图像特征视图、特征表示和特征组织，其中用户视图完成空间推理；语义特征视图从用户观点出发，描述图像特征在某一领域的视图；图像特征视图包括图像实体的形状、颜色、空间关系等特征；特征表示层支持同一图像特征的多重表示，如“圆”可以用轮廓表示，也可以用圆心和半径表示；特征组织层是特征表示层的底层描述，用来定义特征的存储方式、物理结构和索引组织等。

目前图像数据库模型应继承关系／对象模型的诸多优点，如视图定义、代数优化等，同时解决关系／对象模型所不能解决的问题，如非结构化图像数据。模型的主要思想应将对象的模式分解为由稳定属性组成的主表和若干个可变属性组织的副表，查询推理时对主表和副表临时组织供进一步使用，如图像信息数据的存取等，将它们构成不可分割的整体，对外部来说，副表是不可见的，并且支持基于对象标识符的对象引用，从而形成嵌套结构。

在一个完善的图像数据库中，图像信息检索、图像解释及图像信息的识别与处理就必须得到支持，而数据模型要具体体现数据查询的要求，因此，图像特征及其表示是建立图像数据库原型的基础，在这之上，确定相应的检索策略、索引策略等。对于图像数据模型应是一种极其灵活的模型，它可以方便地对图像的语义特征等图像数据库结构的5个模式统一建模。模型可以在线进化，极大地支持图像系统的空间推理，能够自动地适应用户对特定领域知识的不断精炼，并灵活地支持非结构化图像数据建模。

#### 3.4.2.2 数据库中遥感影像的传统存储方式

在数据库中，影像，包括遥感影像的存储方式一般是以BLOB，BFILE两种方式。

BLOB方式：BLOB是数据库提供的一种存储大对象数据的数据类型，它以二进制的形式存储数据，所以常被用于存储图片等数据。使用BLOB字段存储数据主要特点是使得图片数据与整个数据库的数据成为一个整体，这样在库内容的迁移、

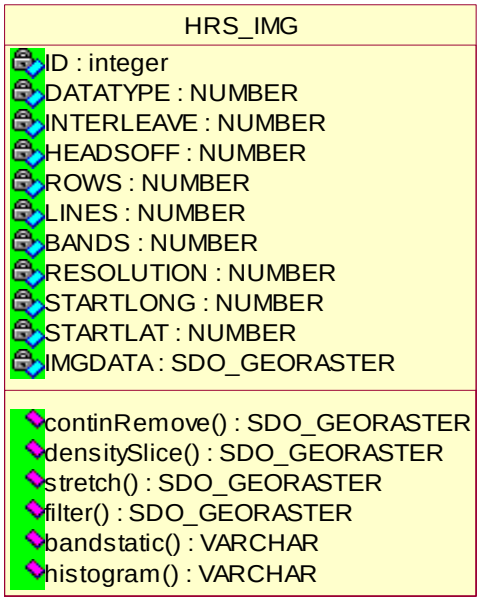
升级、恢复等方面与其他数据都将统一为一个整体，十分有利于数据库的管理和操作。其缺点是读取时速度略慢，不如在库外读取时效率更高。

BFILE 方式：BFile 字段类型实际上是 ORACLE 数据库中一种文件指针，该字段类型仅仅将图像数据在系统中的存放位置存储在数据库中，所以该图像文件整体还是游离于数据库之外的。这对于数据库的完整性的维护增加了困难。

两种方式解决了遥感影像这类二进制对象在数据库中的存储方式，除了其本身的特点之外，都存在着一些局限性。两种方式都把影像数据对应的特征信息和辅助信息分开存储，忽略了影像的信息完整性；在影像数据庞大时，大对象或者大文件的读取是制约网络应用的瓶颈。

### 3.4.2.3 高光谱影像数据模型设计

在高光谱数据库系统中，高光谱影像数据是以 HRSIMG 对象类型存储在数据库中的。在高光谱影像数据模型的设计中，继承了高光谱数据的数据特点和应用，把高光谱影像数据的主要应用参数集成到模型中，使得原始影像数据与影像参数集成在一起，大大提高数据应用的效率。同时，把基本的影像处理方法，也作为对象本身的扩展方法，这为数据库前台的展示和数据的跨平台转移提供了很好的基础。



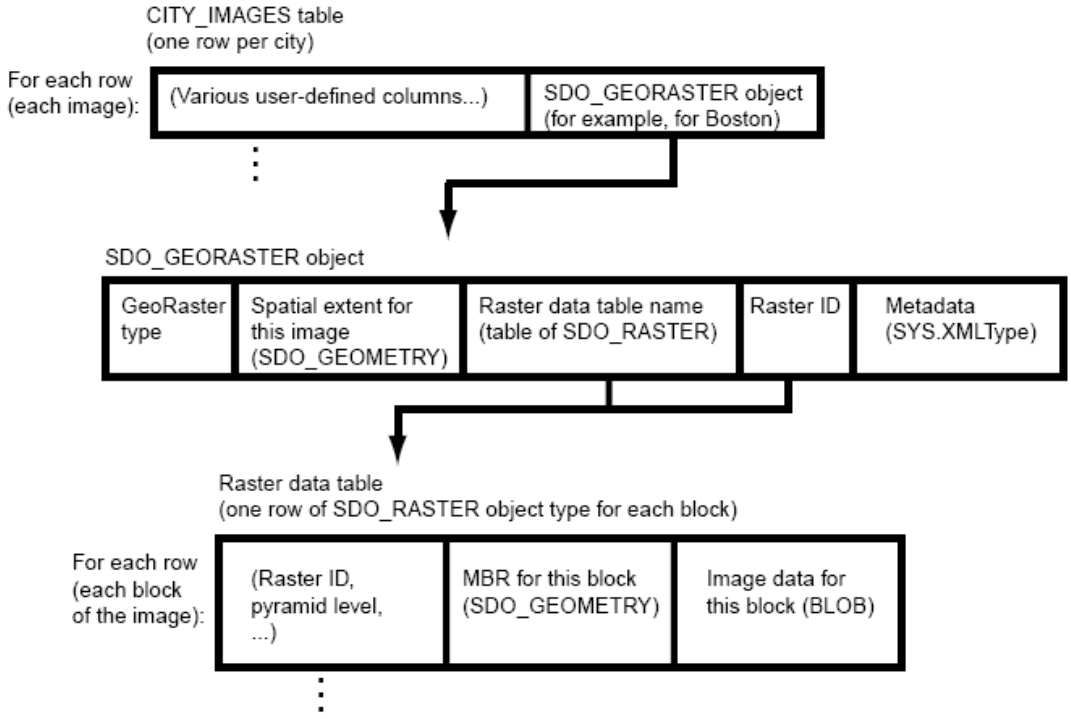

图表 3.5 高光谱影像数据概念模型

#### 3.4.2.3.1 HRSIMG 属性

HRSIMG 模型的核心，即影像数据 imgdata 的存储类型是 GeoRaster。GeoRaster 是 Oracle 数据库 10g 中一个全新的空间数据类型，包含对象关系模式、全面的元数据、操作栅格数据。它允许存储、索引、查询、分析和传送影像及其元数据。

GeoRaster 体系结构提供了支持在 Oracle 数据库 10g 中使用影像或基于网格的栅格数据所需的核心功能。

在 GeoRaster 模型中，核心数据为栅格影像数据，元数据则是除了栅格影像数据之外的以 XML 格式存储的所有其它数据，包括影像结构信息、空间坐标参照信息、图像创建时间信息、波段信息等等。在具体的数据存储中，如图 3.6 所示，在建立的数据表中，影像定义为表中的一个字段，字段类型为 GeoRaster，每一行对应着一个影像数据，而 GeoRaster 对象类型通过初始化栅格序号和最初定义的栅格数据表来对影像数据进行存储。在指定的栅格数据表中，根据 GeoRaster 对象中的存储参数的设定对影像数据进行分块存储。




图表 3.6 GeoRaster 模型的物理结构

### 3.4.2.3.2 HRSIMG 方法

ContinRemove 方法是对光谱维进行包络线去除，从而将数据进行标准化。

DensitySlice 方法是对影像的某一波段根据给定的数值值域进行分割。

Stretch 方法是对图像进行拉伸，采用的是线性拉伸方法。

Filter 方法是对光谱维进行滤波，包括指定步长的中值滤波和均值滤波方法。

BandStatic 方法是对影像某一波段的极值、均值、方差等进行统计

Historgram 方法是对影像的某一波段进行直方图分析。

对于 HRSIMG 模型，除了该模型本身扩展的方法之外，HRSIMG 中影像存储模型 GeoRaster 提供了一系列对于影像数据的基本方法（Jim Farley，2003）：

用于 GeoRaster 数据库管理的主要方法如表 3.3 所示：

表格 3.3 GeoRaster 数据管理

| | |
|---|---|
| init | 初始化一个空的 GeoRaster 对象 |
| createBlank | 创建一个空的 GeoRaster 对象。 |
| copy | 复制现有的 GeoRaster 对象。 |
| importFrom | 将文件中或 BLOB 对象中的图像导入到 GeoRaster 对象中。 |
| exportTo | 将 GeoRaster 对象或子集导出到文件存储在文件系统或 BLOB 对象中。 |
| validateGeoraster | 验证 GeoRaster 对象。 |
| schemaValidate | 根据 GeoRasterXML 模式验证 GeoRaster 对象的元数据。 |

用于 GeoRaster 数据操作的主要方法如表 3.4 所示：

表格 3.4 GeoRaster 数据操作

| | |
|---|---|
| changeFormat | 更改现有 GeoRaster 对象的存储格式，包括更改分块等 |
| changeFormatCopy | 使用不同的存储格式来复制现有的 GeoRaster 对象。 |
| generateSpatialExtent | 生成包含 GeoRaster 对象空间范围的空间几何图形。 |
| georeference | 使用指定的变换系数，对 GeoRaster 对象进行地理参照。 |
| generatePyramid | 生成 GeoRaster 对象的金字塔数据。 |
| deletePyramid | 删除 GeoRaster 对象的金字塔数据。 |
| subset | 执行截取操作：（a）空间挖子区、剪切或剪辑，或（b）层或段子集。 |
| scale | 伸缩（扩大或减小）GeoRaster 对象。 |
| scaleCopy | 伸缩（扩大或减小）GeoRaster 对象，并将结果插入到新对象中。 |
| mosaic | 将无缝的 GeoRaster 对象嵌入到一个 GeoRaster 对象中。 |

用于 GeoRaster 单元数据及元数据的更新与查询的主要方法如表 3.5 所示：

表格 3.5 GeoRaster 对象更新与查询方法

| | |
|---|---|
| getID | 返回与 GeoRaster 对象相关的用户定义标识符的值。 |
| setID | 将用户定义的标识符设置成与 GeoRaster 对象相关，而如果指定一个空的 id 参数，则删除现有的值。 |
| getVersion | 返回 GeoRaster 对象的用户指定版本。 |
| setVersion | 设置 GeoRaster 对象的用户指定版本。 |
| getInterleavingType | 返回 GeoRaster 对象的单元数据交替类型。 |
| getSpatialDimNumber | 返回 GeoRaster 对象的空间维数。 |
| getSpatialDimSizes | 返回 GeoRaster 对象每个空间维的大小/单元数。 |
| getTotalLayerNumber | 返回 GeoRaster 对象中的总层数。 |
| getBlockSize | 将 GeoRaster 对象每个块中每维的单元数以阵列形式返回，显示行维、列维和（如果相关）波段维的单元数。 |
| isSpatialReferenced | 如果 GeoRaster 对象在空间上作了参考则返回 TRUE，而如果 GeoRaster 对象没有作空间参照则返回 FALSE。 |

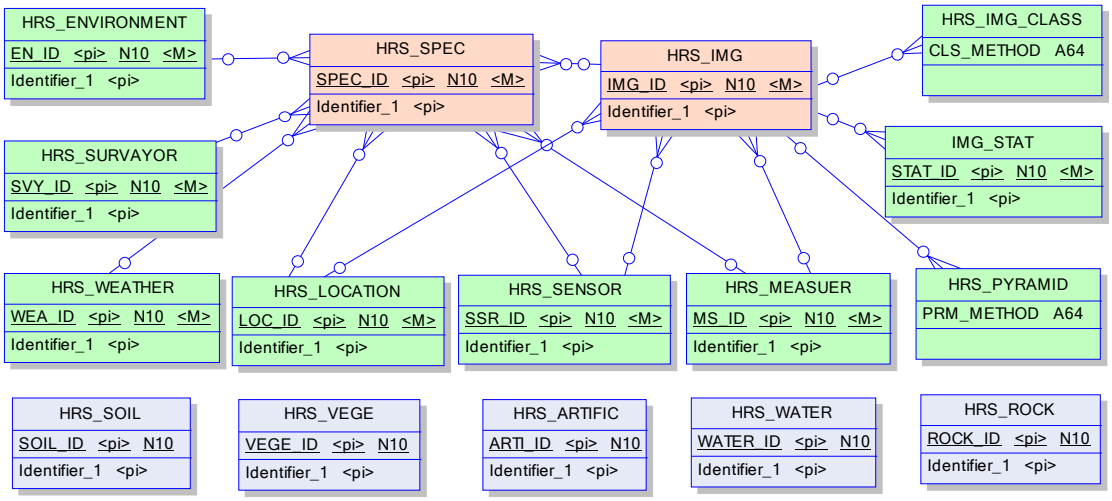
| | |
|---|---|
| setSpatialReferenced | 指定 GeoRaster 对象是否作了空间参照 |
| getSRS | 返回与 GeoRaster 对象的空间参照相关的信息。 |
| setSRS | 设置 GeoRaster 对象的空间参照信息，而如果指定一个空的 srs 参数，则删除现有的信息。 |
| getModelSRID | 返回与 GeoRaster 对象的模型（地面）坐标系统相对应的 SDO_SRID 值。 |
| setModelSRID | 设置 GeoRaster 对象的模型（地面）空间的坐标系统（SDO_SRID 值），而如果指定一个空的 srid 参数，则删除现有的值。 |
| getBeginDateTime | 返回 GeoRaster 对像元数据中栅格数据收集的开始日期和时间。 |
| setBeginDateTime | 设置 GeoRaster 对像元数据中栅格数据收集的开始日期和时间，而如果指定一个空的 beginTime 参数，则删除现有的值。 |
| hasPseudoColor | 检查层中是否包含伪彩色信息（色彩映射表）。 |
| getColorMap | 返回表示某层的伪彩色显示的色彩映射表。 |
| setColorMap | 设置 GeoRaster 对象中某层的色彩映射表，如果指定一个空的 colorMap 参数，则删除现有的值。 |
| getVAT | 返回与某层相关的值属性表（VAT）的名称。 |
| setVAT | 设置与 GeoRaster 对象的某层相关的值属性表（VAT）的名称，而如果指定一个空的 vatName 参数，则删除现有的值。 |
| getPyramidMaxLevel | 返回 GeoRaster 对象顶级金字塔的级数。 |
| getModelCoordinate | 返回与指定单元（栅格）坐标点相关的模型（地面） |

由以上设计可以看出，高光谱影像数据模型是影像数据在数据库中存储的解决方案，它不仅包含了对高光谱影像数据的高效存取，还包含了对影像的基本操作，而且由于这些方法和数据一起内嵌在 ORACLE 数据库中，因此，可以实现数据和方法的跨硬件平台、跨操作系统的迁移，大大提高了高光谱数据库的可应用性。

### 3.4.3 高光谱数据库概念结构

在确定光谱数据模型与影像数据模型之后，需要对数据表中其它数据进行进一层的抽象化、对象化。在高光谱数据库设计时，对应用属性和光谱/影像对象内在属性进行了分离，光谱/影像的内在参数作为映射属性，在关联表中进行存储，而对针对不同领域的应用属性存储在自由表中，如图 3.7。在实际操作中，将通过建立视图对应用属性和测量属性等进行联接，在外模式中展现完整的应用对象。高光谱数据库中，总共包含十六个实体，分别代表光谱数据、传感器参数、测量参数、地区参数、天气参数、环境参数、测量人员信息、影像数据、影像金字塔、影像统计信息、影像分类信息、土壤属性参数、植被属性参数、人工地物属性参数、水体属性参数、岩矿属性参数。每个实体都对应了在数据获取和应用时的具体对象，对于单个实体的扩展不需要对数据库结构作大的变动，而且，在很大的程度上减小了数据的冗余

度。该模式也将光谱数据和影像数据做了有机结合，为像元波谱应用提供了结构基础。光谱数据和影像数据可以分别成为单独的一套存储体系，如果光谱数据是影像的像元波谱或者对应的地面观测数据，可以通过外键和影像数据关联，辅助具体应用。




图表 3.7 高光谱数据库内模式

## 3.5 高光谱数据库方法设计

高光谱方法也是数据模型和应用模型中的基础。考虑高光谱数据库系统的扩展性，在进行高光谱数据模型设计和应用模型设计时，需要把高光谱数据库中能够应用到的方法提炼出来。在高光谱数据库的应用过程中，紧耦合的数据挖掘设计、将数据和基本的数据操作整合在一起的数据模型设计、针对实际应用的模型设计，提高了数据的应用能力和效率,但是在一定程度上也束缚了数据库的扩展性和灵活性。因此，在具体实施数据和模型的时候，需要把方法提炼出来，通过对方法的引用来实现数据操作和模型应用。在高光谱数据库中能够应用到的典型高光谱方法主要包括：反射率转换方法、光谱维滤波方法、光谱匹配方法、包络线去除方法。

### 3.5.1 反射率转换方法

人们为了研究地表物质的光谱响应行为，长期以来测量了大量的地面光谱数据，建立起了地物光谱数据库，使我们有可能通过光谱匹配技术从图像直接识别地物覆盖类型，而我们所获得的高光谱影像数据大多为 DN 值图像，为此，有必要将遥感器获得的辐射亮度 DN 值转换为反射率值。太阳光线通过大气传播到地面，与其发生作用，又通过大气传播到遥感器，受到大气的影响。因此图像反射率的转换实际上就是通过大气校正来实现定标,定标是定量遥感的基础。一般图像反射率转换包括三种方式：利用辐射传输模型进行反射率转换、利用图像本身做反射率转换、利用地面光谱作反射率转换。

3.5.1.1 利用辐射传输方程进行反射率转换

大气削弱和散射的乘性和加性效应以及太阳光谱形状等的影响可以利用辐射传输模型来确定。针对不同的成像系统以及条件发展出多种大气校正模型，利用LOWTRAN 7 进行反射率反演的模型为：

R= （L-L0）/LX 公式 3.1

其中，R：像元某一波段的反射率；L：相应的 DN 值；L0，LX：由 LOWTRAN 7 所计算的 0 反照度表面的程辐射及 100%反射的朗伯体表面的反射辐亮度。

这种方法需要知道经纬度坐标、数据获取时间、光学厚度、大气水分含量等参数。

3.5.1.2 利用图像本身进行反射率反演

这类方法仅从图像数据本身出发进行反射率反演，不需要其它辅助数据。典型的方法有：

（1） 内部平均法

内部平均法是假定一幅图像内部的地物充分混杂，整幅图像的平均光谱基本代表了大气影响下的太阳光谱信息。因而，把图像 $DN$ 值与整幅图像的平均辐射光谱值的比值确定为相对反射率光谱。

$$\rho_\lambda = R_\lambda / F_\lambda \quad \text{公式 3.2}$$

式中，$\rho_\lambda$ 表示相对反射率，$R_\lambda$ 是像元辐射值，$F_\lambda$ 为全图像的平均辐射光谱。

（2） 平场域法

平场域法是选择图像中一块面积大且亮度高而光谱响应曲线变化平缓区域（Flat Field），利用其平均光谱辐射值来模拟飞行时的大气条件的太阳光谱。通过将每个像元的 DN 值除以该平均光谱辐射值的比值作为地表反射率，以此来消除大气的影响。

$$\rho_\lambda = R_\lambda / F_\lambda \quad \text{公式 3.3}$$

式中，$\rho_\lambda$ 表示相对反射率，$R_\lambda$ 是像元辐射值，$F_\lambda$ 为定标点（平场域）的平均辐射光谱。

使用平场域法消除大气影响并建立反射率光谱图像有两个重要的假设条件：① 平场域自身的平均光谱没有明显的吸收特征；② 平场域辐射光谱主要反映的是当时

大气条件下的太阳光谱。

（3） 对数残差法

对数残差法的意义是为了消除光照及地形因子的影响。假设有：

$$DN_{ij} = T_i \times R_{ij} \times I_j \quad \text{公式 3.4}$$

式中，$DN_{ij}$ 是波段 j 中像元 i 的灰度值；$T_i$ 是像元 i 处表征表面变化的地貌因子，对确定的像元所有波段都相同；$R_{ij}$ 为波段 j 中像元 i 的反射率；$I_j$ 为波段 j 的光照因子。

如果假设 $DN_{i\cdot}$ 表示像元 i 的所有波段几何均值，$DN_{\cdot j}$ 表示波段 j 对所有像元的几何均值，$DN_{\cdot\cdot}$ 表示所有像元在所有波段的数据的几何均值。则 $DN_{\cdot j} / DN_{\cdot\cdot}$ 表示 $DN_{ij} / DN_{i\cdot}$ 对一个波段中所有像元的几何平均值：

$$(DN_{ij} / DN_{i\cdot}) \big/ (DN_{\cdot j} / DN_{\cdot\cdot}) = Y_{ij} \quad \text{公式 3.5}$$

$Y_{ij}$ 消除了地形因子与光照因子的影响。

3.5.1.3 借助地面特殊地物的已知光谱反射率进行反射率转换

（1） 混合光谱法（Smith,1987,1990; Gillespie, 1990）

混合光谱法一般用来进行混合像元的分类。采用线性混合光谱模型以及一些地物的已知反射率响应，可用来进行反射率反演。设图像中有 N 个端元，M 个波段，图像上任意一个像元的在波段 b 上的 DN 值为 Lb，第 i 个端元在波段 b 上的图像 DN 值为 Li，b,在像元中的百分比为 Fi，则有:

$$L_b = \sum_{i=1}^{N} F_i L_{i,b} + E_b, \text{ 其中，} b = 1,2,...,M, \sum_{i=1}^{N} F_i = 1 \quad \text{公式 3.6}$$

Eb 为残余误差。对图像的每个像元解方程组，可以得到每个像元的 Fi，如果杰出的 Fi 在政府图像都比较合理，误差水平也比较低，则说明端元的选择是合理的，否则，需要重新设置端元，重解方程组。通过这个过程迭代直到找到满意的解。端元确定以后，利用光谱数据库或者是地光谱测量将端元的图像 DN 值和光谱反射率联系起来，即：在图像的 DN 值和反射率之间找到一系列的增益系数和偏置量，把图像的 DN 值转换为反射率。

（2） 经验线性法

经验线性法需要两个以上光谱均一、有一定面积大小的目标——分别为暗目标和亮目标。假定图像 DN 值与反射率 R 间存在线性关系：

$$DN = kR + b \qquad \text{公式 3.7}$$

实测两个定标点的地面反射光谱值，计算图像上对应像元点的平均辐射光谱。然后，利用线性回归建立起反射光谱与辐射光谱间的相关关系（Roberts et al. 1985，Elvidge 1993，Kruse et al. 1990）。求出系数 k、b 后就得到了 DN 值与反射率 R 之间的关系式，可以进行像元灰度的反射率反演。在使用经验线性法的过程中，对定标点有如下要求：定标点要选择尽可能各向同性的均一地物；定标点地物在光谱上要跨越尽可能宽的地球反射光谱段；定标点要尽可能与研究区域保持同一海拔高度。

以上各种方法各有优劣。辐射传输方法理论较为成熟，但具体使用起来需要知道大气参数及其它有关数据，比较复杂；对航空遥感系统来说，姿态不稳定，没有实时大气参数测试记录，难以使用。利用图像本身做反射率转换的方法，如内部平均法，平场域法，对数残差法等，尽管简单，不需要其他数据，然而所得到的只是反射率的相对值，与绝对意义的反射率在概念上是不同的。混合光谱法对最终光谱单元的选择要求严格，其精度依赖于最终单元的质量，需要通过选择-检验误差-重选的迭代过程以确定最终光谱单元。而经验线性法只需要在获取高光谱遥感图像的同时，进行地面同步典型地物光谱测量，并不需要额外的大气参数及精确的大气模型，它在实现图像光谱重建的过程中，获得的是绝对反射率而非相对值，既简单易行，又更有利于与标准数据库相结合，定量分析实现地物识别。研究表明（刘建贵，），经验线性法反演出的反射率值与实际的反射率值相当吻合，误差也比较小，跟内部平均法得出的光谱曲线相比较，能更贴切地反映地物的光谱响应行为。

在高光谱数据库中进行高光谱影像反射率转换时，由于高光谱数据库可以将各种辅助参数集成在一个环境中，因此，可以把各种方法均集成到数据库中，通过完备的环境参数和辅助信息，来实现对高光谱影像反射率转换的高效和高精度。就目前应用而言，经验线形法是开展研究的主要应用方法。

### 3.5.2 光谱维滤波方法

由于地物光谱曲线在响应滤波对于光谱曲线的特征提取和吸收参数的计算具有非常重要的作用。一般来说，图像的能量主要集中在其低频部分，噪声所在的频段主要在高频段。高光谱影像和地物光谱的光谱维滤波过程，其实质就是一维向量的去噪、平滑，获取光谱曲线主要信息的过程。噪声并不限于人眼所能看的见的失真和变形，有些噪声只有在进行图像或者曲线处理时才可以发现。常见噪声主要有加性噪声、乘性噪声和量化噪声等。实际上，在信号获取的过程中，噪声往往和信号交织在一起，尤其是乘性噪声，如果平滑不当，就会使图像本身的细节如边界轮

廓、线条等变的模糊不清。

滤波算法可分为线性与非线性方法。线性方法提出较早，且有较完备的理论基础。在对付零均值的高斯噪声时，均值滤波是非常有效的方法。但线性方法在滤除噪声的同时也破坏图像中的快变信号，如边缘及细节等，从而使图像变得模糊。同时线性方法无法滤除冲激噪声。1974 年 Tukey 首先将非线性的滤波算法中值滤波应用于图像处理，由于这种方法在保护图像细节的同时能有效地滤除冲激噪声，因此在图像处理领域得到广泛的应用。中值滤波算法，如多级中值滤波（MSM）（Coyle E J，1988），中心加权中值滤波（CWM）（ Ko S J，1991），广义中值滤波（WM）等，这些方法在保护图像边缘方面较普通中值滤波算法有了进一步的改善。由于中值滤波算法与均值滤波算法在滤除冲激噪声与高斯噪声方面各有所长，因此 Lee 和 Kassam 将这两种方法结合起来，提出了改进的均值滤波算法（MTM）（Lee Y H，1985）。

3.5.2.1 均值滤波

均值滤波是一种典型的低通滤波器，传统的均值滤波是是用一个有奇数点的滑动窗口在光谱曲线上滑动，将窗口中心点对应的光谱曲线波段的灰度值用窗口内的各个点的灰度值的平均值代替，如果滑动窗口规定了在取均值过程中窗口各个像素点所占的权重，也就是各个像素点的系数，这时候就称为加权均值滤波；加权均值滤波的表达式如下：

$$\widehat{x}_{k,l}=\frac{\sum_{x_{i,j}\in M(k,l),i\neq k,j\neq l} w_{i,j}\times x_{i,j}}{\sum_{i\neq k,j\neq l} w_{i,j}} \qquad \text{公式 3.8}$$

其中 $Xi_{,}j$ 是中心点（k，1）邻域内像素的灰度值，$\widehat{x}_{k,l}$ 为中心像素点的滤波后的灰度估计值，$Wi_{,}j$ 为滤波窗口中 $i_{,}j$ 对应的权重。模糊加权均值滤波就是通过一定的模糊规则或隶属度函数得到滤波窗口中各点权重的加权均值滤波。

3.5.2.2 中值滤波

中值滤波也是一种典型的低通滤波器，它的目的是保护曲线边缘的同时去除噪声。所谓中值滤波，是指把以某点 x 为中心的小窗口内的所有象素的灰度按从大到小的顺序排列，将中间值作为 x 处的灰度值（若窗口中有偶数个象素，则取两个中间值的平均）。它是一种能够在去除脉冲噪声、椒盐噪声的同时又能保留图像边缘细节的平滑方法。并且由于中值滤波不会明显的模糊边缘，因此可以迭代使用。

经过实际运行证实，中值滤波能有效去除曲线中的噪声点，特别是在一片连续变化缓和的区域中，几乎100％去除突变点（可以认为是噪声点），也因为如此，中值滤波不适合用在一些细节多，如细节点，细节线多的光谱曲线中，因为细节点有可能被当成噪声点去除。中值滤波同时伴随着模糊化效果。

### 3.5.3 光谱匹配方法

基于成像光谱仪在众多窄波段获取数据的特点，可以由已知地物类型的反射光谱，通过波形或特征匹配比较来达到直接识别地物类型的目的。人们对地球上的各种物质已经做了长期的研究，逐步认识了电磁波与地物的相互作用机理；长期的高光谱试验也收集了大量的实验室标准数据，建立了许多地物标准光谱数据库；在高光谱应用研究中，人们也已经解决了图像数据的光谱重建的难题。在这些研究工作的基础上，我们已经具备了从图像直接识别对象的条件。从概念上出发，光谱匹配主要有以下三种运作模式：

（1）从图像的反射光谱出发，将像元光谱数据与光谱数据库中的标准光谱响应曲线进行比较搜索，并将像元归于与其最相似的标准光谱响应所对应的类别，这是一个查找过程。

（2）利用光谱数据库中，将具有某种特征的地物标准光谱响应曲线当作模板与遥感图像像元进行比较，找出最相似的像元并赋予该类标记，这是一个匹配过程。

（3）根据像元之间的光谱响应曲线本身的相似度，将最相似的像元归并为一类，这是一种聚类过程。

在前两种运作模式中，解决问题的关键一是图像辐射亮度值到地物表面反射率的精确反演，使用经验线性法已经可以较好的对此加以解决；二是光谱匹配算法的研究。

在常规情况下，通过地面调查获得地物分布的先验数据，然后通过选取训练样本集对图像进行分类来达到识别的目的，这就是监督分类技术。但地面样本的收集往往费时费力，而借助于标准光谱数据，则可以直接对图像进行识别。地物覆盖由于化学成份差异形成可诊断的典型光谱吸收特征，这成为地物光谱识别的理论基础。以高光谱数据库中的典型地物标准光谱曲线为依据，针对光谱吸收特征的光谱识别有以下几种。

#### 3.5.3.1 二值编码匹配

对光谱库的查找和匹配过程必须是有效的，而且，对成象光谱数据这种大程度的光谱数据冗余度来说，为实施匹配，全部光谱数据的原始形式可能并不必要，所以提出了一系列对光谱进行二进制编码的建议（A. F. H. Goetz，1990）。使得光谱可

用简单的 0～1 来表述。最简单的编码方法是：

$$\begin{cases} h(n)=0, & \text{如果} x(n \le T) \\ h(n)=1, & \text{如果} x(n) > T \end{cases} \qquad \text{公式 3.9}$$

其中 $x(n)$ 是像元第 n 通道的亮度值，$h(n)$ 是其编码，$T$ 是选定的门限值，一般选为光谱的平均亮度，这样每个像元灰度值变为 1bit。但是有时这种编码不能提供合理的光谱可分性，也不能保证测量光谱与数据库里的光谱库相匹配，所以需要更复杂的编码方式。

（1）分段编码

对编码方式的一个简单变形是将光谱通道分成几段进行二值编码，对每一段来说，编码方式同上所示。这种方法要求每段的边界在所有像元矢量都相同。为使编码更加有效，段的选择可以根据光谱特征进行，例如在找到所有的吸收区域以后，边界可以根据吸收区域来选择。

（2）多门限编码

采用多个门限进行编码可以加强编码光谱的描述性能。例如采用两个门限 Ta，Tb 可以将灰度划分为三个域：

$$h(n)=\begin{cases} 00 & \text{如果} x(n) \le T_a \\ 01 & \text{如果} T_a < x(n) \le T_b \quad n=1,2,...N \\ 11 & \text{如果} x(n) > T_a \end{cases} \qquad \text{公式 3.10}$$

这样像元每个通道值编码为 2 位二进制数，像元的编码长度为通道数的两倍。事实上，两位码可以表达 4 个灰度范围，所以采用三个门限进行编码更加有效。

（3）仅在一定波段进行编码

这个方法仅在最能区分不同地物覆盖类型的光谱区编码。如果不同波段的光谱行为是由不同的物理特征所主宰，那么我们可以仅选择这些波段进行编码，这样既能达到良好的分类目的，又能提高编码和匹配识别效率。

一旦完成编码，则可利用基于最小汉明距离的算法来进行匹配识别（X. Jia and J. A. Richards 1993）。

3.5.3.2 光谱角度匹配

当模式类的分布呈扇状分布时，定义两矢量之间的广义夹角余弦为相似函数，这即为较为广泛应用的广义夹角匹配模型。将像元 N 个波段的光谱响应作为 N 维空间的矢量，则可通过计算它与最终光谱单元的光谱之间广义夹角来表征其匹配程度：夹角越小，说明越相似（F. A. Kruse et al, 1993）。两矢量广义夹角余弦为：

$$\theta=\cos^{-1}\frac{T\cdot R}{\|T\|R|}$$ 即：

$$\theta=\cos^{-1}\frac{\sum_{i=1}^{n}t_i\bullet r_i}{\sqrt{\sum_{i=1}^{n}t_i^2}\sqrt{\sum_{i=1}^{n}r_i^2}},\quad \theta\in\left[0,\frac{\pi}{2}\right]$$ 公式 3.11

如果以图像中已知区为参考光谱，则将区域中的光谱的几何平均向量为类中心。设已知某类中有$M$ 个点$R_1,R_2$，.....,$R_M$，则类中心为$\overline{R}=\frac{1}{M}\sum_{i=1}^{M}R_i$。

从两个方面可以改进光谱角度填图方法。修改类中心的计算方法；修改分类准则，利用样区的统计参数。修改后的算法分为以下三个步骤。

⑴ 求类中心

首先将类中的各向量投影到单位半径的超球面上，即对各向量进行归一化，归一化向量为$R'_i=(r_{i1},r_{i2},...,r_{in})$：

$$R'_i=\frac{R}{\|R\|},\text{其中：}\ r'_{ij}=\frac{r_{ij}}{\sqrt{\sum_{j=1}^{N}r_{ij}}},$$ 公式 3.12

归一化后的向量的几何中心也在单位超球面上，以该向量为类中心。即类中心为：

$$\overline{R}'=\frac{1}{M}\sum_{i=1}^{M}R'_i$$ 公式 3.13

⑵ 计算各类的统计特征

以$\overline{R}'$为类中心，根据公式（3.11）求类中各向量$Ri$与类中心的广义夹角$\theta_i$，$i=1,2,\ldots,M$。假设$\theta$是以均值零方差为$\sigma$的正态分布，其概率密度函数为：

$$P(\theta)=\frac{1}{\sqrt{2\pi}\sigma}\bullet e^{\frac{1}{2}\left(\frac{\theta}{\delta}\right)^2}$$ 公式 3.14

当光谱向量$X$与类$i$的中心$Ri'$的广义夹角为$\theta_i$时，$X$属于类$i$的条件概率为$P_i(\theta_i)$。

根据最大似然参数估计，类方差为：

$$\sigma^2=\frac{1}{M}\sum_{i=1}^{M}\theta_i^2$$ 公式 3.15

（3） 分类准则

按照贝叶斯决策规则，当光谱向量 $X$ 属于类 $i$ 的条件概率 $P_i(\theta_i)$ 取最大值时，将 $X$ 划入类 $i$。为了简化计算，取 Pi（θ）的自然对数，得：

$$P_i^{'} = \ln P_i(\theta_i) = -\frac{1}{2}(\frac{\theta_i}{\sigma_i})^2$$

公式 3.16

定义分类准则，当 $P_i^{'} = \max(P_i^{'})$，则将 $X$ 归入类 $i$。

### 3.5.4 包络线去除方法

光谱曲线的包络线从直观上来看，相当于光谱曲线的“外壳”，因为实际的光谱曲线由离散的样点组成，所以我们用连续的折线段来近似光谱曲线的包络线。包络线去除（Continuum Remove）方法是利用包络线来对图像的光谱维进行归一化。具体的算法如下：

求光谱曲线包络线的算法描述如下：

设有反射率曲线样点数组：$r(i), i = 0,1,\cdots,k-1$；

波长数组：$w(i), i = 0,1,\cdots,k-1$；

⑴ $i := 0$，将 $r(i)$，$w(i)$，加入到包络线节点表中；

⑵ 求新的包络节点。如 $i = k-1$ 则结束，否则 $j := i+1$；

⑶ 连接 $i, j$；检查 $(i, j)$ 直线与反射率曲线的交点，如果 $j = k-1$，则结束，将 $w(j)$，$r(j)$ 加入到包络线节点表中，否则：

①. $m := j+1$；

②. 若 $m = k_m = k-1$ 则完成检查，$j$ 是包络线上的点，将 $w(j)$，$r(j)$ 加入到包络线节点表中，$i = j$，转到⑵)；

③. 否则，求 $i, j$ 与 $w(m)$ 的交点 $r1(m)$。

④. 如果 $r(m) < r_1(m)$，则 $j$ 不是包络线上的点，$j := i+1$，转到⑶；如果

$r(m) \le r_1(m)$，则$i, j$与光谱曲线最多有一交点，$m := m+1$，转到②。

⑷ 得到包络线节点表后，将相邻的节点用直线段依次相连，求出$w(i), i = 0,1,\cdots,k-1$所对应的折线段上的点的函数值$h(i), i = 0,1,\cdots,k-1$；从而得到该光谱曲线的包络线。显然有：

$$h(i) \ge r(i) \qquad \text{公式 3.17}$$

⑸ 求出包络线后对光谱曲线进行包络线消除：

$$r'(i) = r(i)/h(i), i = 0,1,\cdots,k-1 \qquad \text{公式 3.18}$$

图 3.8 为几条光谱曲线在包络线消除前后一组反射率图像光谱曲线的对比。

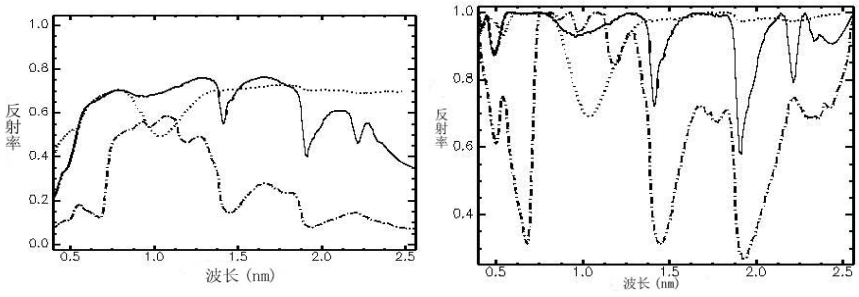



图 3.8 包络线消除前后一组反射率光谱曲线的对比

### 3.5.5 高光谱数据库方法通用设计

从以上四种高光谱数据库中的典型方法可以看出，高光谱数据库需要涵盖的方法，全是以光谱数据（包括地面测量光谱和高光谱影像数据的像元光谱）为输入参数，而输出参数主要是两种：光谱数据和某一数值型运算结果。

在高光谱遥感影像处理软件或者其他通用遥感影像处理软件对这些方法的设计，主要是以数组或者数值型变量来作为接口，而方法实体中的算法则是通过具体的语言逻辑来实现。由于各种开发语言，如 VC++，MATLAB，IDL 等提供了对数组处理的各种函数，这使得在研究中这些方法能够很方便的实现。然而，在数据库中，过程化的查询语言并没有对数组的操作提供良好的支持。比如在 ORACLE 中，严格意义上的数组类型并不存在，我们只能通过自定义类型来实现对数组的模拟，来实现对系列数值的读取，但是，对于数组的各种数学处理，数据库并没有提供直接的函数支持。因此，在对某些光谱数据方法设计和开发时，我们需要通过光谱数据进行转换，从而最大程度利用数据库提供的基本操作方法，从而优化高光谱数据库方法的效率。

数据库中的方法都是以表的字段和记录作为对象的。不管是数据库的结构化查

询语言（SQL），还是 ORACLE 的过程化查询语言（PL/SQL），数据库对关系表的运算和操作做了大量的优化，同时也把一些基本的数学方法应用到了数据库查询语言中。我们要把一些基本方法内嵌在数据库中，实现对数据和方法的集成与融合，就需要用这些查询语言来实现我们的数学运算。根据高光谱数据库方法的特点，我们在对其进行设计开发时，需要把光谱数据转换为关系表对象。这样，通过查询语言，我们可以轻松的对各条光谱/各个像元，即光谱数据表中的记录，进行高效而灵活的读取；我们可以通过强大的查询语句对其联合的属性数据进行有条件的选择与过滤，并对这些完全差异化的子集进行分析处理。同时，借助于 SQL 语言中的数学方法，可以对关系表中的数据集的各个波段进行统计分析处理，也就是说，我们可以对地面测量光谱或者像元光谱进行相关条件的过滤之后，在对全波段或者相关波段进行分析处理。而具体的算法实现，则可以通过 PLSQL。方法的输出方式若为光谱数据，则仍然可以采用光谱数据或者图像数据的存储方式；如果是数值型变量，则可以作为输出参数直接返回该变量。

在这种设计思想下，对于光谱曲线或者像元光谱的处理，转换为对关系数据表的处理。这克服了数据库 SQL 语言的弱点，而对于关系表的处理正是数据库的强项。而且，这种设计思想并不需要弱化我们对数据存储的高效性，而只需要通过开发分别针对光谱曲线和高光谱影像的转换方法即可。

## 3.6 高光谱数据库应用模型设计

高光谱遥感应用遍布各个研究领域，从农业、林业、环境、城市、军事等等各个方面。基于这些不同领域的应用，国内外研究者都建立了很多具有针对性的应用模型。本节主要对其中的一小部分应用模型算法作一些阐述，借此来讨论在高光谱数据库中应用模型的设计。

### 3.6.1 典型矿物波谱识别模型

Clark 等（1990, 1991） 提出了该技术方法, 称之为波段拟合算法。首先参考光谱和试验（像元）光谱包络线去除.选择三个点,分别是吸收中心和吸收中心两侧的点,为了均化包络线上的噪音,可以选择吸收中心两侧的若干个波段。由下列除式进行包络线去除:

$$L_c = \frac{L(\lambda)}{C_l(\lambda)} 和 O_c(\lambda) = \frac{o(\lambda)}{C_o(\lambda)} \qquad 公式\ 3.19$$

$L(\lambda)$ 是实验室光谱作为波长 $\lambda$ 的函数, O 是像元光谱, Cl 是实验室光谱,Co 是像元光谱。

用一个附加常数 K 来增加参考光谱的对比度:

$$L_c^{'} = \frac{L_c + K}{1.0 + K}$$ 公式 3.20

Lc'是调整的包络线去除光谱,与观测光谱最佳拟合.上式可另写为:

$$Lc' = a + bLc$$ 公式 3.21

其中，a=k/（1.0+K） b=1.0/（1.0+K）

在公式 3.21 中,试图解 a 和 b, 以给出对观测光谱 Oc 的最佳拟合.

用标准的 least squares 来解 a 和 b:

$$a = (\sum o_c - b\sum L_c)/n$$

$$b = \frac{\sum o_c L_c - \sum L_c / n}{\sum L_c^2 - (\sum L_c)^2 / n}$$

$$b^{'} = \frac{\sum o_c L_c - \sum o_c \sum L_c / n}{\sum o_c^2 - (\sum o_c)^2 / n}$$

拟合度为 $$F = \frac{n\sum o_c L_c - \sum o_c \sum L_c}{\sqrt{[n\sum o_c^2 - (\sum o_c)^2][n\sum L_c^2 - (\sum L_c)^2]}}$$

特定矿物都具有特殊的光谱吸收特征，矿物实验室或数据库波谱可以与野外及像元光谱匹配。

### 3.6.2 典型岩石矿物组分分析模型

典型岩矿矿物组分分析模型是利用参考光谱（光谱库中的光谱）对未知光谱（图像光谱或者地面光谱）进行匹配识别，该模型不仅能够识别纯的矿物类型，而且能够解算出混合矿物的主要成分类型及其含量。模型主要根据线性光谱模型进行光谱解混，利用参考光谱（光谱库中的光谱）对未知光谱（图像光谱或者地面光谱）进行匹配识别，从而得出岩矿中的矿物类型和相应的组成含量。

根据光谱线性模型，图象中的任意像元 P 都可以有一些纯粹像元（端元）$R_i(i = 1,\cdots,n)$ 线性混合而成：

$$P = \sum_{i=1}^{n} c_i R_i + \mathrm{E}, \qquad \sum_{i=1}^{n} c_i = 1, \qquad 0 \le c_i \le 1$$ 公式 3.22

其中 n 为图象中包含的端元数目，$c_i$ 表示像元 P 中端元 $R_i$ 所占的比例,E 为误差项。

在高光谱图像的波段数据组成的特征空间里，高光谱图象中的每个像元都是其

N 维波段空间中的一个点（N 为图象的波段数），其中有一些称之为端元的点构成了高光谱图象的基本元素，图象中的所有像元都可以由这些端元线性组合而成（省略去误差项 E），正如公式所示。而满足上述两式的所有点的集合正好构成一个 n-1 维空间的凸集，这些端元则坐落于这个凸面单形体的顶点。凸面几何学模型正是以高光谱数据在波段空间的这一特殊的几何特性为基本依据，我们下面将要给出的解混算法也是以此为重要依据。我们根据高光谱数据在高维波段空间中散点的分布为凸面单形体这一特点，提出了一种新的几何模型，并将模型成功地应用于高光谱图象的线性解混。（耿修瑞，2004）

以两个波段、三个端元为例来阐明算法的原理，如图 3.9，像元 P 是以端元 A，B，C 为顶点的三角行内部的一个点，则此像元中端元 A，B，C 对应的地物的含量分别为：

$$c_1 = \frac{S_{PBC}}{S_{ABC}}, \qquad c_2 = \frac{S_{PAC}}{S_{ABC}}, \qquad c_3 = \frac{S_{PAB}}{S_{ABC}}$$

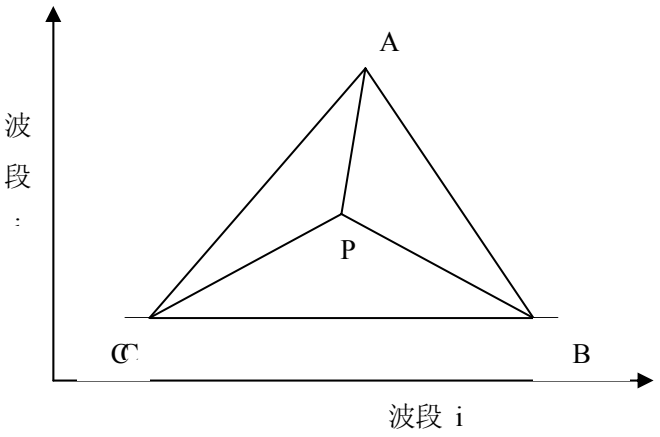



图 3.9 二维情况下混合像元中各端元比例的几何示意图

其中 $S_{ABC}$ 为三角行 ABC 的面积，$S_{PBC}$，$S_{PAC}$ 和 $S_{PAB}$ 也分别是相应三角形的面积。我们把上述结论推广到了高维空间，并且严格证明了在高维空间中对于凸面单形体仍有上述规律成立。根据高维空间凸面单形体的体积公式（蔡聪明）：

$$G = (P_1, P_2, \cdots, P_n) \qquad \text{公式 3.23}$$

$$V(O, P_1, P_2, \cdots, P_n) = \frac{1}{(n-1)!}\sqrt{|G^T G|} \qquad \text{公式 3.24}$$

其中，$P_i(i = 1,2,\cdots,n)$ 为图象中的任意像元. $O$ 为图象波段空间的原点。$V(O, P_1, P_2, \cdots, P_n)$ 为以 $P_i(i = 1,2,\cdots,n)$ 以及原点 $O$ 为顶点的凸面单形体的体积。这里，把原点 $O$ 当作图象的一个已知端元，为了描述方便起见，下面均把

$V(O,P_1,P_2,\cdots,P_n)$ 记为 $V(P_1,P_2,\cdots,P_n)$。

在给出完整的解混算法之前我们先证明如下引理：

引理 1：矩阵 G 的初等行（列）变换不改变行列式 $|G^TG|$ 的值。

证：另 $F=F(i,j(k))$ 表示把矩阵的 j 行的 k 倍加到 i 行的初等矩阵。

则显然有 $|F|=1$

另 $B=GF$

于是有 $|G^TG|=|F^T|\cdot|G^TG|\cdot|F|=|F^TG^TGF|=|B^TB|$。

下面的定理就是我们的光谱解混算法的理论依据：

定理 1：记 $V_0=V(R_1,R_2,\cdots,R_n)$，$V_i=V(R_1,\cdots,R_{i-1},P,R_{i+1},\cdots,R_n)$，则有：

$$c_i=\frac{V_i}{V_0}$$

证：令 $D=(\sum_{i=1}^{n}c_iR_i,R_1,\cdots,R_{i-1},R_{i+1},\cdots,R_n)$

由公式 3.23 有：

$$\begin{aligned}
&V_i=V(R_i,...,R_{i-1},P,R_{i+1},...,R)\\
&=V(P,R,...,R_{i-1},R_{i+1},...R_n)\\
&=V(\sum_{i=1}^{n}c_iR_i,R_1,...,R_{i-1},R_{i+1},...R_n)\\
&=V(D)\\
&=V(DF(1,2(-c_1))F(1,3(-c_2))...F(1,i(-c_{i-1}))F(1,(i+1)(-c_i))...F(1,n(-c_{n-1}))\\
&=V(c_iR_i,R_1,...,R_{i-1},R_{i+1},...R_n)\\
&=c_iV(R_i,R_1,...,R_{i-1},R_{i+1},...R_n)\\
&=c_iV_0
\end{aligned}$$

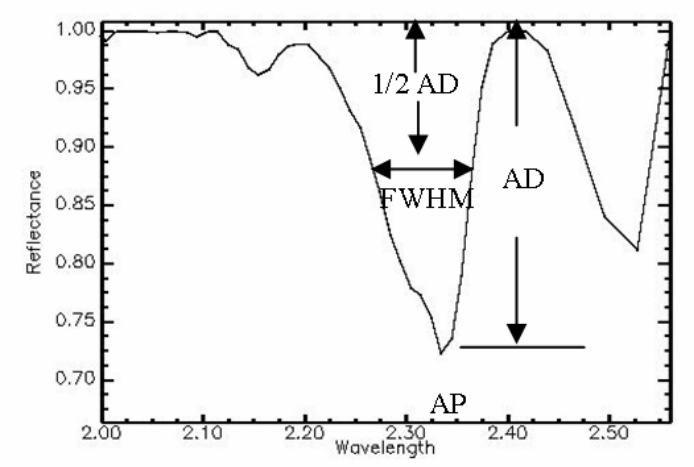
从而有 $c_i=\frac{V_i}{V_0}$。

定理 1 的意义在于，它把高光谱图象的混合像元中各个端元所占的比例归结为一个简单的体积比：凸面单形体内任意一点 P 与凸面单形体的顶点集 $(O,R_1,\cdots,R_{i-1},R_{i+1},\cdots,R_n)$ 所围成的体积 $V_i$ 与凸面单形体的所有顶点

$(O,R_1,R_2,\cdots,R_n)$所围成的体积$V_0$之比即为端元$R_i$在混合像元 P 中所占的比例。

与最小二乘法相比，由于比例系数$c_i(i=1,2,\cdots,n)$的求解表达式只用到了矩阵乘积及行列式的运算，所以时间复杂度会相应降低，同时不会出现$c_i(i=1,2,\cdots,n)$为负数的情况；与凸面几何学模型相比，虽然都运用了高光谱图象在波段空间中呈现凸面单形体这一几何特性，但是本模型又引入了一种含量与体积比的关系，并且本模型所提出的算法不需要对原始数据降维，从而不会导致‘忽视’小目标的现象；本模型算法显然更加有效了利用了高光谱数据本身所特有在波段空间中的几何分布特点以及它们之间内在的含量与体积比的关系，因而能够取得更好的解混效果。

### 3.6.3 岩矿光谱吸收参数提取模型




图表 3.10 光谱吸收特征

岩矿光谱光谱吸收特征的量化往往建立在包络线去除和归一化的光谱曲线上，通过对岩矿光谱曲线进行去包络线处理，能够将数据标准化，同时也突出吸收特征。图 3.10 显示了方解石（Calcite）的吸收光谱。量化的光谱吸收特征包括：

⑴ 吸收位置（Absorption Position，AP）：

在光谱吸收谷中,反射率最低处的波长, 即$AP=\lambda$，当$\rho_\lambda=Min(\rho)$。

⑵ 吸收深度（Absorption Depth，AD）：

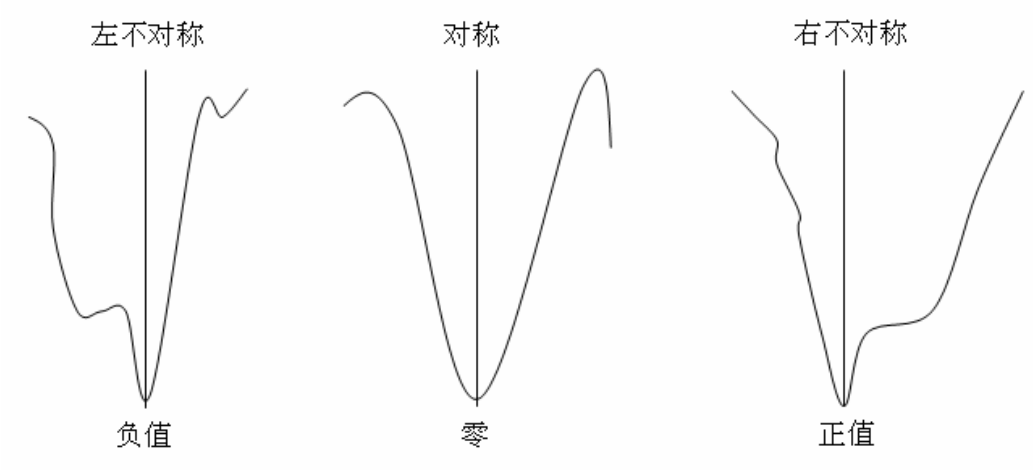
在某一波段吸收范围内，反射率最低点到归一化包络线的距离。

$AD=1-\rho_0$，$\rho_0$为吸收谷点的反射率值。

⑶ 吸收宽度（Absorption Width，AW）：最大吸收深度一半处的光谱带宽 FWHM（Full Width at Half the Maximum Depth）。

⑷ 对称性（Absorption Asymmetry，AA）：

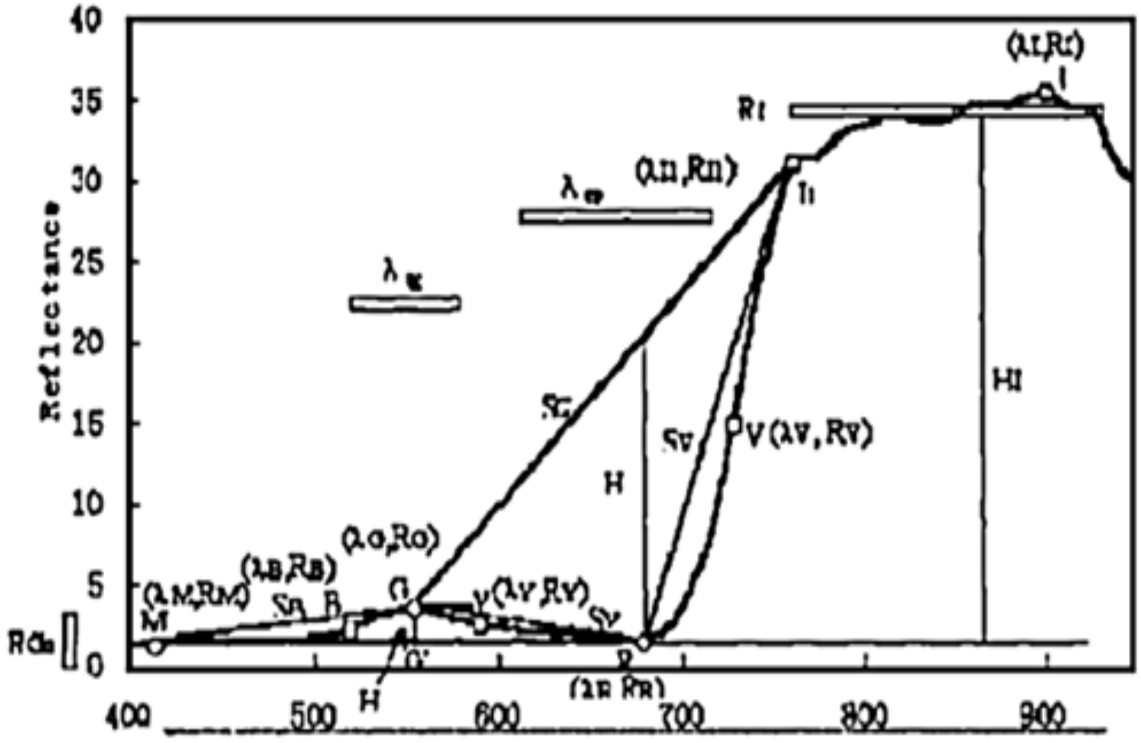
如图 3.11 所示，光谱吸收对称性定义为，以过吸收位置的垂线为界线，右边区域面积与左边区域面积比值的以 10 为底的对数（张兵，2002）。




图表 3.11 光谱吸收对称性分析

### 3.6.4 植被光谱特征参数提取模型

由于高光谱遥感在植被应用领域具有明显优势，针对植被的光谱特征参数的研究工作在国内外研究较为广泛。谭倩通过对植被光谱特征参数的研究，在以八个特征为之为基础，提出了 13 个特征参数的定义和算法，如图 3.12（谭倩，2001）：




图表 3.12 植被光谱维特征

（1）蓝边斜率 SB：为直线 MG 的斜率，近似反映了曲线 MYG 的特征，算法是： SB=（RG-RM）/（ λG -λM） 。

（2）黄边斜率 SY：为直线 RG 的斜率，近似反映了曲线 GYR 的特征，算法是：SY=（RG-RR）/（ λG –λR） 。

（3）VIR 边斜率 SV：为直线 RI1 的斜率,近似反映了曲线 RVI1 的特征，算法是：SV1=（RV1-RR）/（ λV1 –λR） 。

（4）包络线斜率 SC:为直线 GI1 的斜率,近似反映了曲线红波段吸收峰 GYRVI1 的背景,算法是:SC=（RG-RI1）/（ λG-λI1） 。

（5）绿峰净高度 HG：为 G 到直线 MR 的 y 向距离,相当于绿峰去背景后的净反射率,反映了反射峰 MBGYR 的特征,算法是:HG=RG-（（RR-RM）/（ λR-λM）×（λG-λR）+RR） 。

（6）红谷净深度 HR：为 R 到直线 GI1 的 y 向距离,相当于红吸收峰去背景后的深度,反映了反射峰 GYRVI1 的特征,算法是:HR=（RG-RI1）/（λG-λI1）×（λR-λG）+RG-RR。

（7）红外平台净高度 HI:为红外平台的平均高度与 R 点的反射率之差,近似地可以用直线 I1I 的中点与 R 点的反射率之差来代替,相当于红外反射率平台的净高度,反映了反射平台 I1I 的特征,算法是：HI=（Σ（（Ri+Ri+1）×（λi+1-λi）/2））/Δλ-RR≈（RI1+RI）/2-RR,其中 λi∈λI1-930nm,Δλ 为红外波段宽度=930-λI1。

（8）绿峰半高宽 λwG:反映了绿峰的宽度,为绿峰在净高的一半（HG/2）处两光谱位置的差,可用 BY 的水平距近似表示,算法是:λwG≈λY-λB。

（9）红吸收峰半高宽 λwR:反映了红谷的宽度,为去包络后红吸收谷的半高宽,显然可用 YV 的水平距来近似表示,算法是:λwR≈λV-λY。

（10）红外平台平均高度 RIa:为红外波段（I1-I 间）反射率的平均值,反映了红外吸收平台的特征,可以用 I1 和 I 处的平均值来代替,算法是：RIa=（Σ（(Ri+Ri+1）×（λi+1-λi）/2））/Δλ

≈（RI1+RI）/2=HI+RR,其中 λi∈λI1-930nm,Δλ 为红外波段宽度=930-λI1。

（11）绿峰面积 AG：为曲线 MBGYR 下的面积,反映了绿峰的强度,可以用折线 MBGYR 下的面积近似地代替,算法是: AG=（Σ((RI+RI+1）×（λI+1-λI）/2)) ≈[（Σ（Rp+Rp+1）×（λp+1-λp）/2)),p=M,B,G,Y]， 其中 λI∈λM-λR。

（12）绿峰净面积 AG′：为曲线 MBGYR 和直线 MR 所围的面积,反映了绿峰的净强度,显然可以用多边形 MBGYRM 的面积来近似,算法是:AG′=AG-((RM+RR）×（λR-λM）/2）≈[（Σ（Rp+Rp+1）×（λp+1-λp）/2)),p=M,B,G,Y]-((RM+RR）×（λR-λM）/2）,其中 λi∈λM-λR。

（13)）红吸收峰净面积 AR：为曲线 GYRVI1 与直线 GI1 所围的面积,反映了红谷的净强度,它可以用多边形 GYRVI1G 的面积来近似,算法是:AR=S 多边形 GYRVI1G。

通常在研究植被光谱时所用的参数 NDVI、红边斜率、红边位置等,完全可以由上述参数得到,例如:NDVI=（RI1-RR）/（RI1+RR）,红边位置=λV,红边斜率≈SV。

### 3.6.5 高光谱数据库应用模型通用设计

从以上模型的算法分析，可以看出，在基于高光谱数据库的应用中，模型的处理对象都是光谱。这一点，和高光谱数据库方法一样。因此，在将必要的高光谱遥感应用模型整合到高光谱数据库中时，我们同样可以采用高光谱数据库方法的通用设计思想，来对高光谱数据库应用模型进行设计与开发。这样，对于光谱及波段的处理，更具灵活性，只需要作一次性数据读取和转换，便可以利用数据库对关系表操作的高效性实现模型的应用目的。在此，高光谱数据库模型与方法的设计理念得以统一，而所不同的是模型具有针对性、专一性和应用性，而方法则是具有广泛性和普适性。

对于部分模型的设计，有一点需要补充考虑。模型往往要通过一些常数或者变量来实现对应用结果的控制。比如说回归模型的置信度和方差，监督分类模型的类别数量等等。因此，对于模型的设计，还需要建立模型表，通过在模型表中输入控制因子的名称和变换其数值大小来实现对模型的应用弹性。

## 3.7 本章小结

本章主要针对高光谱遥感的应用和数据挖掘的需求对高光谱数据库设计理念进

行了阐述。高光谱数据库要满足对数据发布和共享的需求，因此三层的网络体系结构是数据库首选的架构。为了满足应用层面的跨平台，对于数据挖掘模块（模型/方法）与数据库采取了紧耦合设计。针对高光谱遥感在光谱维信息研究中的侧重，在数据库的数据模型设计中，着重把光谱维空间的相关方法和必要属性集成到地面光谱数据模型和高光谱影像数据中。

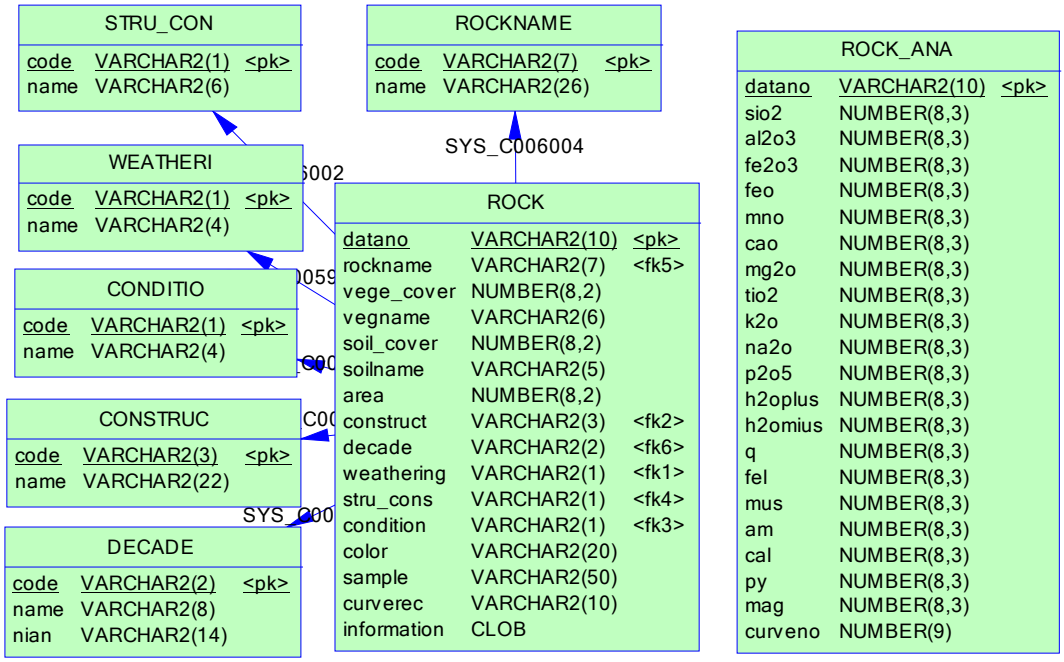
# 第四章 高光谱数据库系统建设

由于高光谱遥感应用的多元化，高光谱数据库内存储的数据也涵盖了各种应用领域。根据第三章的数据库概念设计理念，本章主要对具体的物理结构设计与实现、数据导入与升迁、高光谱数据库典型方法开发、高光谱典型模型开发作了阐述。

## 4.1 高光谱数据库结构建设

数据库结构建设需要将原有数据库进行升迁，并对原有数据结构进行分析和提炼之后，将其优化到新的数据库结构中。原有数据库的光谱数据存储主要有两种结构，一是大型表对象存储的所有数据的大表结构，另一是关联的多表存储数据表群结构。

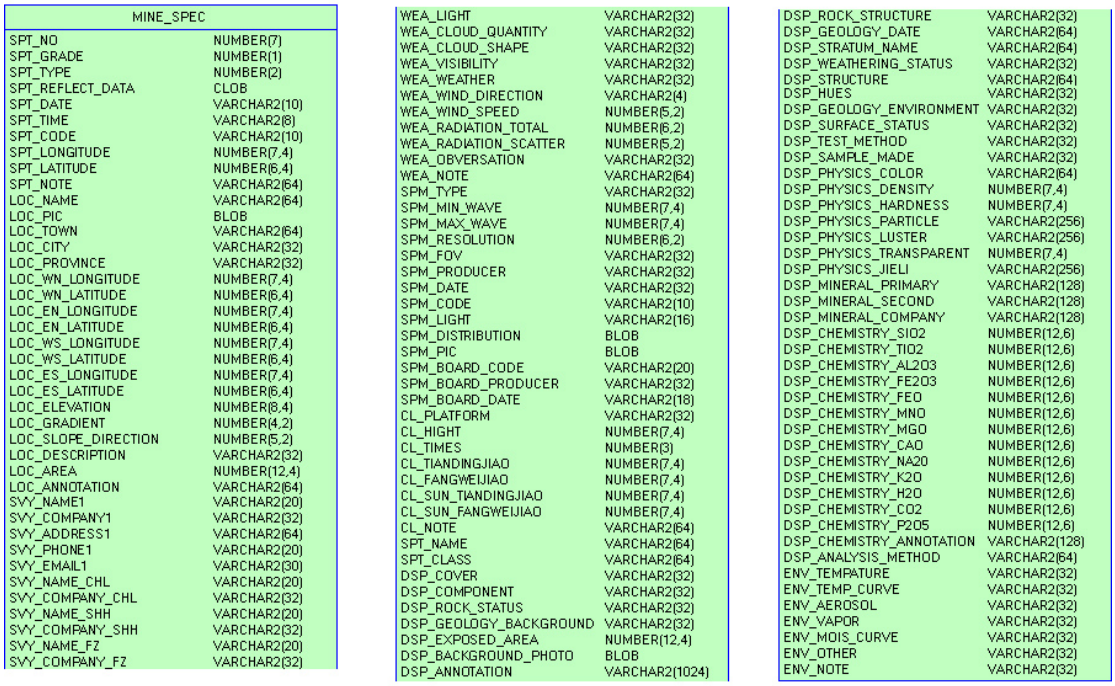
以原有岩石矿物数据为例，在原有 FOXPRO 数据库中，岩石属性数据是以 ROCK, ROCK_ANA, ROCKNAME, STRU_CON, WEATHERI, CONDITIO, CONSTRUC, DECADE 八个表存储的，而通过 DATANO 关键字与光谱数据表相关联，从而形成对光谱数据和属性数据的联合存储，如图 4.1。这种表群存储可以明显地节省空间，最大程度的降低数据的冗余度，但是，他对数据查询检索带来了麻烦，而且，数据表之间的关联太多，在增加新数据和删除数据的时候，维护各种约束会导致效率降低。从下图看出，ROCK 表几乎把大部分单一字段提取出来作为对象存储在关联表中，在现在的存储条件大大得以提升的情况下，这种结构可以进一步改进，以减少关联操作带来的时间损耗。




图表 4.1 表群结构

数据库表存储的另外一种方式，是把所有的数据放入一张大表中，通过对频繁字段建立索引实现对数据快速检索。如图 4.2，整张 MINE_SPEC 表有 124 个字段。

大表结构在数据添加或者大批量导入时比较方便，而且维护起来简单，只有一张大表，每一行记录完全独立的代表一个光谱测量数据，在查询检索的时候，不许要做表的链接操作，效率较高。但是，这种结构最直接的缺点就是数据冗余度太大，在同样测量环境下测量的数据属性都要被重复存储。



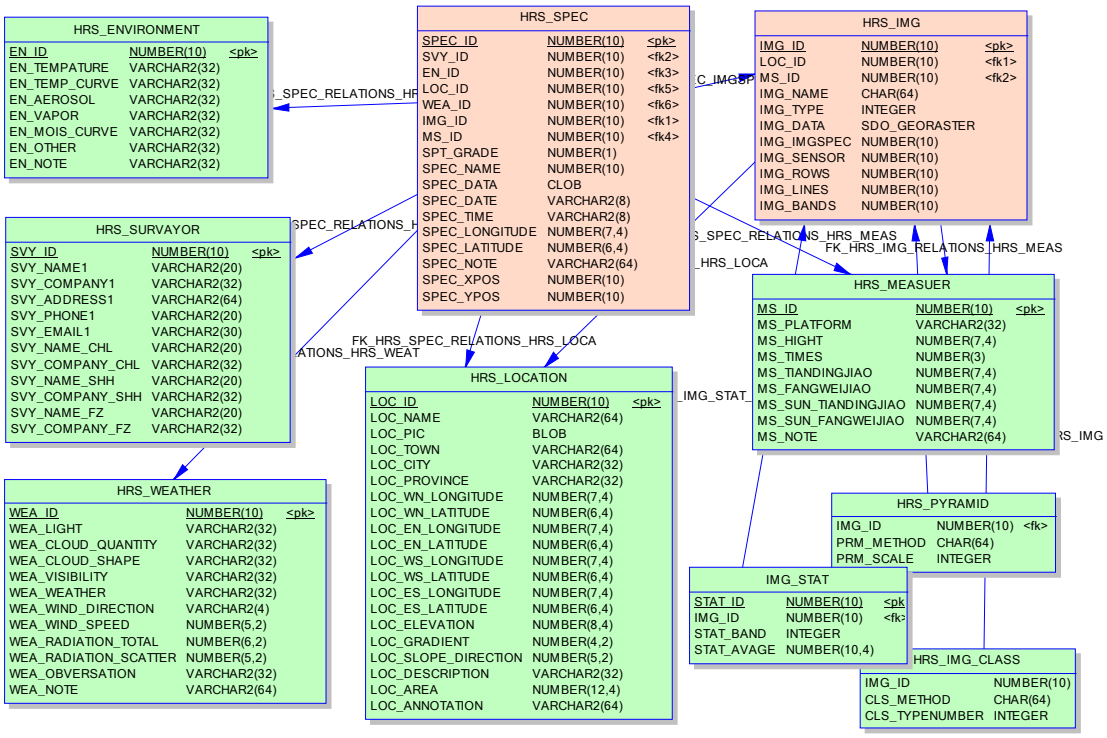
| MINE_SPEC | |
|---|---|
| SPT_NO | NUMBER(7) |
| SPT_GRADE | NUMBER(1) |
| SPT_TYPE | NUMBER(2) |
| SPT_REFLECT_DATA | CLOB |
| SPT_DATE | VARCHAR2(10) |
| SPT_TIME | VARCHAR2(8) |
| SPT_CODE | VARCHAR2(10) |
| SPT_LONGITUDE | NUMBER(7,4) |
| SPT_LATITUDE | NUMBER(6,4) |
| SPT_NOTE | VARCHAR2(64) |
| LOC_NAME | VARCHAR2(64) |
| LOC_PIC | BLOB |
| LOC_TOWN | VARCHAR2(64) |
| LOC_CITY | VARCHAR2(32) |
| LOC_PROVINCE | VARCHAR2(32) |
| LOC_WN_LONGITUDE | NUMBER(7,4) |
| LOC_WN_LATITUDE | NUMBER(6,4) |
| LOC_EN_LONGITUDE | NUMBER(7,4) |
| LOC_EN_LATITUDE | NUMBER(6,4) |
| LOC_WS_LONGITUDE | NUMBER(7,4) |
| LOC_WS_LATITUDE | NUMBER(6,4) |
| LOC_ES_LONGITUDE | NUMBER(7,4) |
| LOC_ES_LATITUDE | NUMBER(6,4) |
| LOC_ELEVATION | NUMBER(8,4) |
| LOC_GRADIENT | NUMBER(4,2) |
| LOC_SLOPE_DIRECTION | NUMBER(5,2) |
| LOC_DESCRIPTION | VARCHAR2(32) |
| LOC_AREA | NUMBER(12,4) |
| LOC_ANNOTATION | VARCHAR2(64) |
| SVY_NAME1 | VARCHAR2(20) |
| SVY_COMPANY1 | VARCHAR2(32) |
| SVY_ADDRESS1 | VARCHAR2(64) |
| SVY_PHONE1 | VARCHAR2(20) |
| SVY_EMAIL1 | VARCHAR2(30) |
| SVY_NAME_CHL | VARCHAR2(20) |
| SVY_COMPANY_CHL | VARCHAR2(32) |
| SVY_NAME_SHH | VARCHAR2(20) |
| SVY_COMPANY_SHH | VARCHAR2(32) |
| SVY_NAME_FZ | VARCHAR2(20) |
| SVY_COMPANY_FZ | VARCHAR2(32) |

| | |
|---|---|
| WEA_LIGHT | VARCHAR2(32) |
| WEA_CLOUD_QUANTITY | VARCHAR2(32) |
| WEA_CLOUD_SHAPE | VARCHAR2(32) |
| WEA_VISIBILITY | VARCHAR2(32) |
| WEA_WEATHER | VARCHAR2(32) |
| WEA_WIND_DIRECTION | VARCHAR2(4) |
| WEA_WIND_SPEED | NUMBER(5,2) |
| WEA_RADIATION_TOTAL | NUMBER(6,2) |
| WEA_RADIATION_SCATTER | NUMBER(5,2) |
| WEA_OBVERSATION | VARCHAR2(32) |
| WEA_NOTE | VARCHAR2(64) |
| SPM_TYPE | VARCHAR2(32) |
| SPM_MIN_WAVE | NUMBER(7,4) |
| SPM_MAX_WAVE | NUMBER(7,4) |
| SPM_RESOLUTION | NUMBER(6,2) |
| SPM_FOV | VARCHAR2(32) |
| SPM_PRODUCER | VARCHAR2(32) |
| SPM_DATE | VARCHAR2(32) |
| SPM_CODE | VARCHAR2(10) |
| SPM_LIGHT | VARCHAR2(16) |
| SPM_DISTRIBUTION | BLOB |
| SPM_PIC | BLOB |
| SPM_BOARD_CODE | VARCHAR2(20) |
| SPM_BOARD_PRODUCER | VARCHAR2(32) |
| SPM_BOARD_DATE | VARCHAR2(18) |
| CL_PLATFORM | VARCHAR2(32) |
| CL_HIGHT | NUMBER(7,4) |
| CL_TIMES | NUMBER(3) |
| CL_TIANDINGJIAO | NUMBER(7,4) |
| CL_FANGWEIJIAO | NUMBER(7,4) |
| CL_SUN_TIANDINGJIAO | NUMBER(7,4) |
| CL_SUN_FANGWEIJIAO | NUMBER(7,4) |
| CL_NOTE | VARCHAR2(64) |
| SPT_NAME | VARCHAR2(64) |
| SPT_CLASS | VARCHAR2(64) |
| DSP_COVER | VARCHAR2(32) |
| DSP_COMPONENT | VARCHAR2(32) |
| DSP_ROCK_STATUS | VARCHAR2(32) |
| DSP_GEOLOGY_BACKGROUND | VARCHAR2(32) |
| DSP_EXPOSED_AREA | NUMBER(12,4) |
| DSP_BACKGROUND_PHOTO | BLOB |
| DSP_ANNOTATION | VARCHAR2(1024) |

| | |
|---|---|
| DSP_ROCK_STRUCTURE | VARCHAR2(32) |
| DSP_GEOLOGY_DATE | VARCHAR2(64) |
| DSP_STRATUM_NAME | VARCHAR2(64) |
| DSP_WEATHERING_STATUS | VARCHAR2(32) |
| DSP_STRUCTURE | VARCHAR2(64) |
| DSP_HUES | VARCHAR2(32) |
| DSP_GEOLOGY_ENVIRONMENT | VARCHAR2(32) |
| DSP_SURFACE_STATUS | VARCHAR2(32) |
| DSP_TEST_METHOD | VARCHAR2(32) |
| DSP_SAMPLE_MADE | VARCHAR2(32) |
| DSP_PHYSICS_COLOR | VARCHAR2(64) |
| DSP_PHYSICS_DENSITY | NUMBER(7,4) |
| DSP_PHYSICS_HARDNESS | NUMBER(7,4) |
| DSP_PHYSICS_PARTICLE | VARCHAR2(256) |
| DSP_PHYSICS_LUSTER | VARCHAR2(256) |
| DSP_PHYSICS_TRANSPARENT | NUMBER(7,4) |
| DSP_PHYSICS_JIELI | VARCHAR2(256) |
| DSP_MINERAL_PRIMARY | VARCHAR2(128) |
| DSP_MINERAL_SECOND | VARCHAR2(128) |
| DSP_MINERAL_COMPANY | VARCHAR2(128) |
| DSP_CHEMISTRY_SIO2 | NUMBER(12,6) |
| DSP_CHEMISTRY_TIO2 | NUMBER(12,6) |
| DSP_CHEMISTRY_AL2O3 | NUMBER(12,6) |
| DSP_CHEMISTRY_FE2O3 | NUMBER(12,6) |
| DSP_CHEMISTRY_FEO | NUMBER(12,6) |
| DSP_CHEMISTRY_MNO | NUMBER(12,6) |
| DSP_CHEMISTRY_MGO | NUMBER(12,6) |
| DSP_CHEMISTRY_CAO | NUMBER(12,6) |
| DSP_CHEMISTRY_NA2O | NUMBER(12,6) |
| DSP_CHEMISTRY_K2O | NUMBER(12,6) |
| DSP_CHEMISTRY_H2O | NUMBER(12,6) |
| DSP_CHEMISTRY_CO2 | NUMBER(12,6) |
| DSP_CHEMISTRY_P2O5 | NUMBER(12,6) |
| DSP_CHEMISTRY_ANNOTATION | VARCHAR2(128) |
| DSP_ANALYSIS_METHOD | VARCHAR2(64) |
| ENV_TEMPATURE | VARCHAR2(32) |
| ENV_TEMP_CURVE | VARCHAR2(32) |
| ENV_AEROSOL | VARCHAR2(32) |
| ENV_VAPOR | VARCHAR2(32) |
| ENV_MOIS_CURVE | VARCHAR2(32) |
| ENV_OTHER | VARCHAR2(32) |
| ENV_NOTE | VARCHAR2(32) |

图表 4.2 大表结构

由于这两种结构都存在着不足之处，因此需要对数据结构进行优化。在建立新数据库结构之前，对表空间进行了划分。TBS_HRSDATA 用于存放实体表数据，TBS_HRSIDX 用于存放在这些表上建立的索引，而 TBS_HRSRASTER 用于存放 GEORASTER 对象数据表和一些 BLOB 数据。如果将表数据和索引数据放在一起，表数据的 I/O 操作和索引的 I/O 操作将产生影响系统性能的 I/O 竞争，降低系统的响应效率。将表数据和索引数据存放在不同的表空间中，并在物理层面将这两个表空间的数据文件放在不同的物理磁盘上，就可以避免这种竞争。同时，索引和主体表分开表空间存储对于日后的数据库迁移也带来方便：可以只迁出表数据的表空间降低数据大小，在目标数据库中通过重建索引的方式就可以生成索引数据。而对于影像数据而言，其主体表在 TBS_HRSDATA 中，但是 GEORASTER 对象是存放在相应的数据表中的，GEORASTER 的本质存储方式仍然为 LOB 形式，而 LOB 类型的数据在物理存储结构的管理上和一般数据的策略有很大的不同，将其放在一个独立的表空间中，可方便地设置其物理存储参数。




图表 4.3 优化的数据库结构

## 4.2 高光谱数据库数据建设

在对数据库结构进行优化设计之后，我们便在数据库内建立该结构表群。并通过数据库的升迁工具，将原有的各种数据源数据导入到 ORACLE 数据库中，并通过数据库的 SQL 语言将数据转换到新的数据模型中。高光谱数据库目前主要搜集了六个部分的数据，包括岩矿地面测量光谱数据、农作物地面测量数据、水体光谱数据、城市地物光谱数据、岩矿像元波谱数据、多时相 MODIS AVI 数据。在本节中将主要介绍这六部分数据状况。

### 4.2.1 岩矿地面测量光谱数据

这部分数据中收集了三千多条岩矿地面测量光谱数据。主要有两部分组成：一部分为原有的搜集数据，另一部分为 863 岩矿波谱库项目测量数据。

原有的搜集数据主要的来源有 3 个方面：

第一部分是由中科院安徽精密光学机械研究所提供，绝大部分为岩石光谱。主要岩性类涵盖范围较广，包括橄榄角闪岩，辉榄玄武岩，辉榄玄武岩，辉长岩，辉绿岩，二长岩，闪长岩，正长岩，安山岩，次粗面岩，凝灰岩，花岗闪长岩，流纹岩，伟晶岩等各类岩浆岩；砾岩，杂砾岩，杂砾岩，砂岩，长石石英砂岩，粉砂岩，泥岩，页岩，石灰岩，白云岩等各类沉积岩；以及板岩，千枚状，石英岩，大理岩，片岩，片麻岩，混合花岗岩等各类变质岩；另外也包括一些如蛇纹石岩等的蚀变岩

和矿石如黄铁矿矿石，磁铁矿矿石，赤铁矿矿石，重晶石矿石等。除 38 条数据为其他岩性类（矿石、蚀变岩等）外，共包括 37 种大的岩性类。这部分数据共 261 条。其中矿物光谱 8 条，岩石光谱 253 条。岩石光谱中，属岩浆岩的有 88 条，属变质岩的有 61 条，属沉积岩的有 66 条，属其他岩性类（矿石等）的有 38 条。因此此部分数据岩性类上的分布是比较均匀的。此部分数据中有 142 条为室内测定的光谱，其余 119 条为野外光谱。所采用的光谱仪主要有 101W 野外光谱辐射计（0.4－1.1um）、H-10 野外光谱仪（0.4－1.1 um），UV340（0.4－2.4 um），SRM-1200 型野外光谱仪（0.39－1.15 um）等 4 种。其中 UV340 是主要的室内测定光谱仪。这批数据基本上都配有：岩石或矿物的化学成份分析结果、采样点的经、纬度记录。而且野外光谱数据基本都有太阳方位角等记录。这批数据整体来说测试较规范，数据质量较高，欠缺的是按现有波谱库数据规范要求的有些必要的配套数据和说明项因在当时没有记录而欠缺，另外光谱采样间隔也相对较大。

第二部分为中科院遥感所所提供的光谱数据，绝大部分为野外测定的岩石光谱，共 862 条。只有少量矿物光谱或室内测定的光谱。光谱数据来源比较广泛，测定时间为从上世纪 80 年代初到本世纪初。其中矿物光谱为 12 条，其余 850 条为岩石光谱。岩石的岩性类函盖比较广泛，共包括 41 个编码方案中编例的岩性类，主要有：安山玢岩，安山岩，白云岩，板岩，粗面岩，大理岩，二长岩，粉砂岩，硅质岩，黑云母，花岗岩，灰岩，辉长岩，辉绿岩，混合岩，角闪石岩，砾岩，流纹斑岩，流纹岩，泥灰岩，泥岩，泥质岩，凝灰岩，片麻岩，片岩，千枚岩，熔结岩，砂岩，闪长玢岩，闪长岩，蛇纹岩，石灰岩，石英岩，伟晶岩，矽卡岩，响岩，玄武岩，页岩，英安斑岩，英安岩，正长岩等。另外还包括一些矿石等归列为其他岩性类的样本。岩石样品中，有岩浆岩样 105 条，变质岩样 71 条，沉积岩样 579 条，其余为其他类岩石。因此此批样品的岩性在分布较广泛齐全的同时，沉积岩样相对偏多。样品的测量使用了 11 种左右的光谱仪，主要有 GER－MARK5 等（0.4-2.5um）。由于数据来源较杂，这批数据的配套参数齐全程度不一，除了其中 238 条（约占 28%）具有化学成份分析结果外，其余样品都缺少这方面的数据。其余的必须配套参数也多有缺失，主要是由于当时这部分波谱的采集目的是为了图像数据定标或检验之用。这批数据整体来说质量较高，但野外测定样本不一定完全附合规范，稍显粗糙。欠缺的是按现有波谱库数据规范要求，有较多的必要配套数据和说明项缺少。

第三部分则收集自美国 USGS 的公开矿物光谱数据集，共 481 条。此部分数据只包括光谱数据本身，除了样品名称属性外缺失其他任何按现有波谱库数据规范要求的必要的配套数据和说明项。但这部分数据质量较高，光谱仪是 Beckman spectrometer，波谱波长范围是 0.3951－2.56um，光谱分辨率在可见光－近红外区约为 2nm，短波区约为 10nm 左右，波段数 420。这部分数据对象全部为矿物，除部分（如水）外全部是矿物粉末样，并在实验室内测定。其中包含矿物定名 219 个（包括水等），分属 134 个矿物种，按编码方案则包括 74 类（族）矿物。

第二部分测量数据从全国各地采集并搜集了我国典型岩矿样本，对所有样本进行了规范的标本制作，及物理化学参量测定，同时主要采用室内分光光度计为主要仪器对标准样品进行了波谱测量，以及标准化的数据预处理之后获取的光谱数据及配套参数。

样品采集区总共 181 个，主要分布于以下我国 27 个省市自治区：北京、天津、河北、山西、内蒙古自治区、辽宁、吉林、黑龙江、江苏、浙江、安徽、福建、江西、山东、河南、湖北、湖南、广东、四川、贵州、云南、西藏、陕西、甘肃、青海、宁夏、新疆。

此次地面波谱数据采集总共获取了 399 个岩石样本，1386 条光谱数据及其配套参数，其中岩石样品主要包括：变质岩 84 种、沉积岩 128 种、岩浆岩 106 种。同时，采集了 123 个矿物样本，获取了 238 条光谱数据及其配套参数，其中矿物样本包括了层状硅酸盐、岛状硅酸盐、环状硅酸盐、架状硅酸盐、链状硅酸盐、磷酸盐、硫化物、硫酸盐、氯化物、碳酸盐、硝酸盐、氧化物、氢氧化物和自然元素等。所采集的岩矿样品覆盖了我国主要的岩石和矿物类型。总共收入地面岩矿光谱数据为 1624 条，其中野外岩石 286 条，室内岩石 1100 条，矿物 238 条。表 4.4 是岩矿对应的配套参数表结构。

HRS_ROCK

| | | |
|---|---|---|
| ROCK_ID | NUMBER(10) | <pk> |
| ROCK_COVER | VARCHAR2(32) | |
| ROCK_COMPONENT | VARCHAR2(32) | |
| ROCK_ROCK_STATUS | VARCHAR2(32) | |
| ROCK_GEOLOGY_BACKGROUND | VARCHAR2(32) | |
| ROCK_EXPOSED_AREA | NUMBER(12,4) | |
| ROCK_BACKGROUND_PHOTO | BLOB | |
| ROCK_ANNOTATION | VARCHAR2(1024) | |
| ROCK_ROCK_STRUCTURE | VARCHAR2(32) | |
| ROCK_GEOLOGY_DATE | VARCHAR2(64) | |
| ROCK_STRATUM_NAME | VARCHAR2(64) | |
| ROCK_WEATHERING_STATUS | VARCHAR2(32) | |
| ROCK_STRUCTURE | VARCHAR2(64) | |
| ROCK_HUES | VARCHAR2(32) | |
| ROCK_GEOLOGY_ENVIRONMENT | VARCHAR2(32) | |
| ROCK_SURFACE_STATUS | VARCHAR2(32) | |
| ROCK_TEST_METHOD | VARCHAR2(32) | |
| ROCK_SAMPLE_MADE | VARCHAR2(32) | |
| ROCK_PHYSICS_COLOR | VARCHAR2(64) | |
| ROCK_PHYSICS_DENSITY | NUMBER(7,4) | |
| ROCK_PHYSICS_HARDNESS | NUMBER(7,4) | |
| ROCK_PHYSICS_PARTICLE | VARCHAR2(256) | |
| ROCK_PHYSICS_LUSTER | VARCHAR2(256) | |
| ROCK_PHYSICS_TRANSPARENT | NUMBER(7,4) | |
| ROCK_PHYSICS_JIELI | VARCHAR2(256) | |
| ROCK_MINERAL_PRIMARY | VARCHAR2(128) | |
| ROCK_MINERAL_SECOND | VARCHAR2(128) | |
| ROCK_MINERAL_COMPANY | VARCHAR2(128) | |
| ROCK_CHEMISTRY_SIO2 | NUMBER(12,6) | |
| ROCK_CHEMISTRY_TIO2 | NUMBER(12,6) | |
| ROCK_CHEMISTRY_AL2O3 | NUMBER(12,6) | |
| ROCK_CHEMISTRY_FE2O3 | NUMBER(12,6) | |
| ROCK_CHEMISTRY_FEO | NUMBER(12,6) | |
| ROCK_CHEMISTRY_MNO | NUMBER(12,6) | |
| ROCK_CHEMISTRY_MGO | NUMBER(12,6) | |
| ROCK_CHEMISTRY_CAO | NUMBER(12,6) | |
| ROCK_CHEMISTRY_NA2O | NUMBER(12,6) | |
| ROCK_CHEMISTRY_K2O | NUMBER(12,6) | |
| ROCK_CHEMISTRY_H2O | NUMBER(12,6) | |
| ROCK_CHEMISTRY_CO2 | NUMBER(12,6) | |
| ROCK_CHEMISTRY_P2O5 | NUMBER(12,6) | |
| ROCK_CHEMISTRY_ANNOTATION | VARCHAR2(128) | |
| ROCK_ANALYSIS_METHOD | VARCHAR2(64) | |

图表 4.4 岩矿光谱属性表

### 4.2.2 农作物地面测量光谱数据

农作物地面测量光谱数据主要是收集了北京精准农业波谱数据库中的测试数据 45 条光谱数据和完备的配套数据，以及小汤山精准农业基地全生命周期的小麦光谱曲线 151 条及相关的 LAI 数据。农作物光谱配套参数表如图 4.5 所示，涵盖了土壤数据、植株理化组分参数、播种和肥水管理信息、植株结构和结构参数等等。

| HRS_VEGE | | |
|---|---|---|
| VEGE_ID | NUMBER(10) | <pk> |
| VEGE_ROWDIS | NUMBER(4) | |
| VEGE_STATUS | VARCHAR2(40) | |
| VEGE_DENSITY | NUMBER(4) | |
| VEGE_LEAFAGE | VARCHAR2(20) | |
| VEGE_LEAFCOLOR | VARCHAR2(8) | |
| VEGE_PLANTHEIGHT | NUMBER(4) | |
| VEGE_COVER | NUMBER(8,2) | |
| VEGE_STEMDIAMETER | NUMBER(3,1) | |
| VEGE_LAD | NUMBER(3) | |
| VEGE_LEAFDIREC | NUMBER(3) | |
| VEGE_SEEDTIME | VARCHAR2(20) | |
| VEGE_IRRIGATIONTIME1 | VARCHAR2(20) | |
| VEGE_IRRIGATION1 | NUMBER(4) | |
| VEGE_FERTILIZATIONTIME1 | VARCHAR2(20) | |
| VEGE_FERTILIZATIONNAME1 | VARCHAR2(20) | |
| VEGE_FERTILIZATION1 | NUMBER(6,2) | |
| VEGE_IRRIGATIONTIME2 | VARCHAR2(20) | |
| VEGE_IRRIGATION2 | NUMBER(4) | |
| VEGE_FERTILIZATIONTIME2 | VARCHAR2(20) | |
| VEGE_FERTILIZATIONNAME2 | VARCHAR2(20) | |
| VEGE_FERTILIZATION2 | NUMBER(6,2) | |
| VEGE_IRRIGATIONTIME3 | VARCHAR2(20) | |
| VEGE_IRRIGATION3 | NUMBER(4) | |
| VEGE_FERTILIZATIONTIME3 | VARCHAR2(20) | |
| VEGE_FERTILIZATIONNAME3 | VARCHAR2(20) | |
| VEGE_FERTILIZATION3 | NUMBER(6,2) | |
| VEGE_LAI | NUMBER(4,2) | |
| VEGE_SLA | NUMBER(6,3) | |
| VEGE_LWC | NUMBER(5,2) | |
| VEGE_RWC | NUMBER(5,2) | |
| VEGE_PWC | NUMBER(4,2) | |
| VEGE_LCHLA | NUMBER(4,2) | |
| VEGE_LCHLB | NUMBER(4,2) | |
| VEGE_LCHLAB | NUMBER(4,2) | |
| VEGE_LTN | NUMBER(4,2) | |
| VEGE_LPROTEIN | NUMBER(4,2) | |
| VEGE_LAMYLUM | NUMBER(4,2) | |
| VEGE_LCELLULOSE | NUMBER(4,2) | |
| VEGE_LLIGNIN | NUMBER(4,2) | |
| VEGE_LANTHOCYANIN | NUMBER(4,2) | |
| VEGE_LXANTHIN | NUMBER(4,2) | |
| VEGE_PN | NUMBER(5,3) | |
| VEGE_TR | NUMBER(5,3) | |
| VEGE_GS | NUMBER(5,3) | |
| VEGE_SCHLA | NUMBER(4,2) | |
| VEGE_SCHLB | NUMBER(4,2) | |
| VEGE_SCHLAB | NUMBER(4,2) | |
| VEGE_STN | NUMBER(4,2) | |
| VEGE_SPROTEIN | NUMBER(4,2) | |
| VEGE_SAMYLUM | NUMBER(4,2) | |
| VEGE_SCELLULOSE | NUMBER(4,2) | |
| VEGE_SLIGNIN | NUMBER(4,2) | |
| VEGE_SOORGANIC | NUMBER(4,2) | |
| VEGE_SONNITROGEN | NUMBER(4,2) | |
| VEGE_SOANITROGEN | NUMBER(4,2) | |
| VEGE_SOWATER1 | NUMBER(4,2) | |
| VEGE_SOWATER2 | NUMBER(4,2) | |
| VEGE_SOPHOSPHOR | NUMBER(4,2) | |
| VEGE_SOKALIUM | NUMBER(4,2) | |
| VEGE_SOEC | NUMBER(4,2) | |
| VEGE_DIGPICTURE | BLOB | |

图表 4.5 农作物（植被）数据属性表

### 4.2.3 城市地物光谱数据

| HRS_ARTIFIC | | |
|---|---|---|
| ARTI_ID | NUMBER(10) | <pk> |
| ARTI_NAME | VARCHAR2(10) | |
| ARTI_TYPE | VARCHAR2(10) | |
| ARTI_DATANO | NUMBER(10) | |
| ARTI_LENGTH | NUMBER(9) | |
| ARTI_WIDTH | NUMBER(9) | |
| ARTI_AREA | NUMBER(8,2) | |
| ARTI_COLOR | VARCHAR2(20) | |
| ARTI_ENVIRONMENT | VARCHAR2(60) | |
| ARTI_SAMPLE | VARCHAR2(1) | |
| ARTI_CURVENO | NUMBER(9) | |
| ARTI_REC | NUMBER(9) | |
| ARTI_SHAPE | VARCHAR2(32) | |
| ARTI_SLOPEANGLE | VARCHAR2(10) | |
| ARTI_SLOPDIRE | VARCHAR2(4) | |
| ARTI_CHARACTER | VARCHAR2(10) | |

| HRS_SOIL | | |
|---|---|---|
| SOIL_ID | NUMBER(10) | <pk> |
| SOIL_DATANO | VARCHAR2(10) | |
| SOIL_NAME | VARCHAR2(5) | |
| SOIL_COVER_PER | NUMBER(8,2) | |
| SOIL_VEGE_NAME | VARCHAR2(6) | |
| SOIL_WATERLEVEL | NUMBER(8,2) | |
| SOIL_COLOR | VARCHAR2(20) | |
| SOIL_SAMPLE | VARCHAR2(5) | |
| SOIL_CURVE_NO | NUMBER(9) | |
| SOIL_QUALITY | VARCHAR2(4) | |
| SOIL_UTILITY | VARCHAR2(14) | |
| SOIL_DRAINCON | VARCHAR2(4) | |
| SOIL_EROSINTE | VARCHAR2(10) | |
| SOIL_EROSTYPE | VARCHAR2(10) | |
| SOIL_HUMIDITY | VARCHAR2(4) | |
| SOIL_LANDFORM | VARCHAR2(12) | |
| SOIL_ORIGINRO | VARCHAR2(10) | |
| SOIL_ORIGINSO | VARCHAR2(10) | |
| SOIL_ROUGH | VARCHAR2(8) | |

| HRS_WATER | | |
|---|---|---|
| WATER_ID | NUMBER(10) | <pk> |
| WATER_DATANO | VARCHAR2(10) | |
| WATER_COLOR | VARCHAR2(20) | |
| WATER_TEMPERATUR | NUMBER(8,1) | |
| WATER_DEPTH | NUMBER(8,2) | |
| WATER_TRANSPAREN | NUMBER(8,2) | |
| WATER_SAND_PER | NUMBER(8,1) | |
| WATER__SCHLO_PER | NUMBER(8,4) | |
| WATER_ANIMAL | VARCHAR2(3) | |
| WATER_WAVEHEIGHT | NUMBER(9) | |
| WATER_BUBBLE | VARCHAR2(1) | |
| WATER_SAMPLE | VARCHAR2(1) | |
| WATER_CURVE_REC | NUMBER(9) | |
| WATER_CURVENO | NUMBER(9) | |
| WATER_WATERNAME | VARCHAR2(14) | |
| WATER_POLLU_CO | VARCHAR2(8) | |
| WATER_BOTTOM | VARCHAR2(4) | |
| WATER_POLLU_IN | VARCHAR2(4) | |
| WATER_WATER_CO | VARCHAR2(10) | |

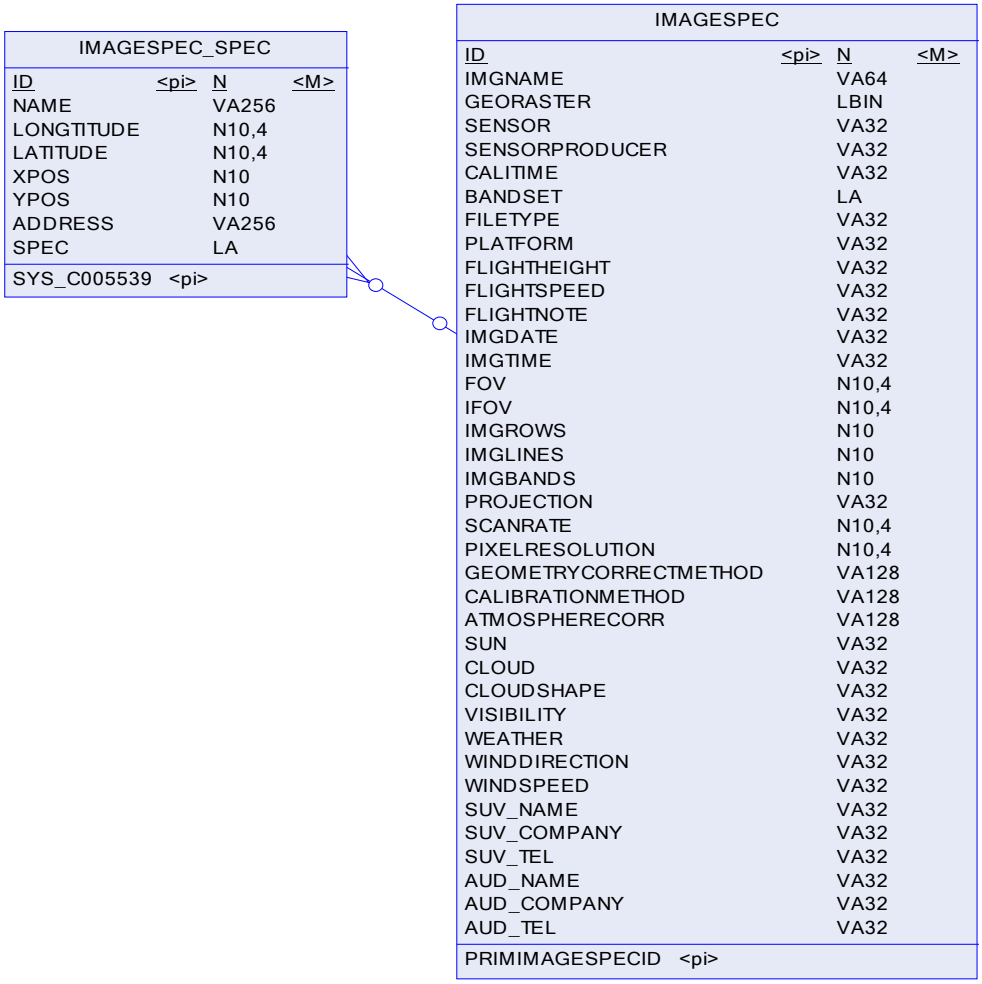
图表 4.6 城市地物数据人工地物、土壤、水体属性表

城市地物光谱数据是从原有 FOXPRO 城市地物光谱数据库中升迁而来的数据，总共包含数据 3252 条，而具体的数据类别包括岩石矿物数据 876 条，人工地物数据 194 条，土壤数据 594 条，植被数据 1432 条，水体数据 156 条。岩石矿物和植被数据属性表如图 4.4，4.5 所示，人工地物、土壤、水体数据属性表如图 4.6 所示。

### 4.2.4 岩矿像元波谱数据

岩矿像元波谱数据包括高光谱影像块和像元光谱数据两部分。数据的来源包括3个方面： 第一部分是航空高光谱传感器HyMap在新疆东天山的飞行实验数据。共计包含影像块53个，像元光谱曲线64条；第二部分是航空高光谱传感器HyMap在澳大利亚的飞行实验数据。共计包含影像块28个，像元光谱曲线28条；第三部分是航空高光谱传感器AVIRIS在美国的飞行实验数据。共计包含影像块5个，像元光谱曲线8条。

从这些影像块上提取像元光谱有2种办法，一种是利于同步地面观测点的经纬度信息，在经过了几何纠正的影像块上找到这个经纬度对应的像元，有61%的像元光谱是用这种办法获得的，这些像元光谱的命名、及其相关属性的确定由同步的地面观测点的数据决定。另一种提起像元光谱的办法是利于现有的岩矿光谱库，将影像块上的像元光谱曲线和光谱库中的光谱曲线进行匹配，从而得到匹配度很好的像元光谱，这些像元光谱的命名及其相关属性由光谱库中的相应信息决定。有31%的像元光谱是用这种办法得到的。像元波谱的关系表结构如图4.7所示。




图表 4.7 像元波谱数据表

### 4.2.5 多时相 MODIS AVI 产品

这是作为高光谱影像数据的案例存储在数据库中。这是存储了 2003 年 23 个时间点的华北地区 AVI 指数数据，也就是一个 23 波段的时间光谱影像。这和存储高光谱影像数据时完全相同的。在存储高光谱影像数据时，和数据库的批量数据导入以及数据库升迁不一样，由于涉及到具体的影像数据读取以及在数据库中以特定模型存储，因此需要利用存储过程实现影像数据导入的自动化。首先需要把高光谱数据转换为 BIP 格式的 tif 文件，然后运行以下 sql 文件，便可以实现对影像数据的自动化入库。对该程序略作修改，便可以实现网络客户端的影像数据上载。

```
DEFINE SYSPWD = '***' –sys 用户密码，用于对数据授权
DEFINE IMGFILE = &&1 –需要载入的影像文件参数，在运行程序后输入
DEFINE USERNAME = '***'--数据管理员账号
DEFINE USERPWD = '***'--数据管理员账号密码
--对文件进行授权
CONNECT SYS/&SYSPWD AS SYSDBA;
exec dbms_java.grant_permission （ 'MDSYS','SYS:java.io.FilePermission', '&IMGFILE', 'read' ） ;
exec dbms_java.grant_permission （ '&USERNAME','SYS:java.io.FilePermission', '&IMGFILE', 'read' ） ;
--更换用户登录
CONNECT &USERNAME/&USERPWD;
DECLARE
geo SDO_GEORASTER;
i number（10）;
BEGIN
--确定记录号
select count（*） into i from TEST_IMG;
i:=i+1;
INSERT INTO TEST_IMG VALUES （ i,0,0,'',2,SDO_GEOR.INIT （'TEST_IMG_RDT'）,0,0,0,0,0）;
--导入数据
SELECT IMG_DATA INTO geo FROM TEST_IMG WHERE IMG_ID=i FOR UPDATE;
SDO_GEOR.IMPORTFROM（geo,'blocksize=（512,512）','TIFF','file','&IMGFILE'）;
UPDATE TEST_IMG SET IMG_DATA = geo WHERE IMG_ID=i;
END;
```

在上传数据完毕之后，可以设定几何坐标参考系并生成影像金字塔。

## 4.3 高光谱数据库的典型方法

### 4.3.1 地物光谱数据转换方法

为了提高数据存储效率，地物光谱数据是以 CLOB/BLOB 方式存放在数据库中的。根据第三章的分析，在对光谱数据进行处理分析时，需要提供一个数据转换方法来把 CLOB/BLOB 格式的地物光谱数据转换为表数据，从而利用数据库内嵌的统计分析方法来实现对地物光谱数据的应用分析。本文用 PL/SQL 实现了这一方法。输入参数为数据表/视图名称，间隔采样步长 k，波段数 bandnumber。通过对数据表创建符合查询条件的视图，我们可以对数据记录进行快速选择；通过光标的操作，我们可以对属性数据进行选择；因此，该方法能够在数据记录和属性字段实现灵活的操作，并将我们需要研究的光谱曲线和属性数据转换到同一数据表中，方便我们进行分析。

```
declare
  cursor cr_para is select datano, chl_a, spectrum from watertest;
type datatype is table of float index by binary_integer;
  type datatable is table of datatype index by binary_integer;
  refdata datatype;
  wave datatype;
lnumPos       NUMBER := 1;
  lnumLength NUMBER;
  lstrChunk    VARCHAR2（32767）;
  cnumBiteSize CONSTANT NUMBER := 32767;
  e_clob clob;
  teststring varchar2（32767）;
  readlength integer;
  firstloc integer;
  lastloc integer;
  i integer;
  j integer;
  m integer;
  n integer :=1;
  k constant integer :=1;
```

```
 bandnumber integer:=191;
begin
--创建光谱数据表
   execute immediate 'drop table watertab';
   execute immediate 'create table watertab（datano number, chl_a number）';
   for i in 1..bandnumber loop
        execute immediate 'alter table watertab add （band'||i||' number）';
   end loop;
   commit;
--对原有数据逐行读取光谱
for r1 in cr_para
loop
   --convert blob to clob
   lnumLength := DBMS_LOB.GETLENGTH（r1.spectrum）;
   if lnumLength<32767 then
      lstrChunk := UTL_RAW.CAST_TO_VARCHAR2（DBMS_LOB.SUBSTR
（r1.spectrum,lnumLength,lnumPos））;
      e_clob:=to_clob（lstrChunk）;
   else
   WHILE lnumPos <= lnumLength LOOP
            lstrChunk := UTL_RAW.CAST_TO_VARCHAR2
（DBMS_LOB.SUBSTR（r1.spectrum,cnumBiteSize,lnumPos））;
            dbms_lob.append（e_clob,to_clob（lstrChunk））;
            lnumPos := lnumPos + cnumBiteSize;
         END LOOP;
   end if;
i:=1;
j:=1;
--open the lob object
dbms_lob.open（e_clob,dbms_lob.lob_readonly）;
   loop
   --get one line
        firstloc:=dbms_lob.instr（e_clob,chr（13）,1,i）;
        lastloc:=dbms_lob.instr（e_clob,chr（13）,1,i+1）;
        readlength:=lastloc-firstloc-1;
        teststring:=to_char（dbms_lob.substr（e_clob,readlength,firstloc+1））;
```

```
--将数据转换为自定义数组
    teststring:=trim（both chr（10） from teststring）;-- trim half-enter
    wave（j）:=to_number（substr（teststring,1,instr（teststring,chr（9））-1））;
    refdata（j）:=to_number（substr（teststring,instr（teststring,chr（9））+1））;
    i:=i+k;--间隔 k 个波段选择一个数值
    j:=j+1;
   if lastloc=0 then
   dbms_lob.close（e_clob）;
   exit;
   end if;
 end loop;
 insert into watertab（datano,chl_a） values （r1.datano, r1.chl_a）;
--将光谱数据转换到数据表中
 for j in 1..190 loop
    execute immediate 'update watertab set band'||j||'='||refdata（j）||' where datano='||r1.datano;
 end loop;
end loop;
commit;
end;
```

### 4.3.2 影像与数据表转换方法

对高光谱影像的相关波谱转换和地物光谱操作类似。这是针对单个影像操作，即把一幅高光谱影像转换为由于影像数据在高光谱数据库中是以 GEORASTER 对象存放，因此可以调用 sdo_geor.getCellValue 方法实现对影像数据单个像元值得提取。本方法将高光谱影像转换为影像数据表有两种方式：第一、将单波段影像转换为（行、列）形式的数据表，第二、将多波段影像转换为（像元、波段）形式的数据表。在对影像数据进行分析处理的时候，这两种方式分别侧重的是光谱维信息和空间维信息。对高光谱影像数据的空间信息进行统计分析时，可以采用（行、列）式数据表，而对高光谱影像的光谱维信息进行分析时，则采用（像元、波段）式数据表。以下代码为将单波段影像转换为（行、列）形式数据表，需要输入的参数有：影像表名、影像记录号、影像行数、影像列数，后两个参数可以直接从影像数据模型中获取，输出则为影像数据表。

```
declare
 i integer;
```

```
    j integer;
    rowsnumber integer:=10;
    linesnumber integer:=10;
    geor sdo_georaster;
begin
    execute immediate 'drop table dcimgtabtest';
    execute immediate 'create table dcimgtabtest （lineno number）';
    select img_data into geor from test_img where img_id=11;--原始数据
    for i in 1..rowsnumber loop
    execute immediate 'alter table dcimgtabtest add （row'||i||' number）';
    end loop;
    for i in 1..linesnumber loop
        insert into dcimgtabtest （lineno） values （i）;
        for j in 1..rowsnumber loop
        execute immediate 'update dcimgtabtest set row'||j||'='||sdo_geor.getCellValue
（geor,0,i,j,0）;
        end loop;
        end loop;
    commit;
end;
```

对于多波段影像数据存储的（像元，波段）方式，需要输入的参数有：影像表名、影像记录号、影像行数、影像列数、影像波段数，后三个影像特征参数都可以从影像数据模型中直接获取。输出则为影像数据表。

```
declare
    rowsnumber integer:=100;
    linesnumber integer:=100;
    bandnumber integer:=23;
    imgid integer:=11;
    i integer;
    j integer;
    k integer;
    geor sdo_georaster;
begin
    execute immediate 'drop table dcimgtabtest';
    execute immediate 'create table dcimgtabtest （pixno number）';
    select img_data into geor from test_img where img_id=imgid;--原始影像数据
```

```
    for i in 1..bandnumber loop
    execute immediate 'alter table dcimgtabtest add （band'||i||' number）';
    end loop;
    for i in 1..linesnumber loop
      for j in 1..rowsnumber loop
          insert into dcimgtabtest （pixno） values （i*j）;
           for k in 1..bandnumber loop
          execute immediate 'update dcimgtabtest set row'||j||'='||sdo_geor.getCellValue
（geor,0,i,j,k）;
          end loop;
          end loop;
          end loop;
    commit;
  end;
```

### 4.3.3 包络线去除方法

包络线去除方法主要是针对一维向量，也就是光谱曲线进行的操作。将它以像元作为循环，便可以对整幅图像进行包络线去除操作。该方法的核心输入参数为光谱曲线的波长和对应的反射率。而核心算法则是获取包络线节点，在获取包络线节点之后，对节点之间的曲线点作线形变换即可。

```
Declare
Type datatype is table of float index by binary_integer;
Newspec datatype;
Nodesspec datatype;-- 保存节点光谱
Nodeswave datatype;-- 保存节点波段
    Nodesspec（1）=spec[0];//第一个节点
    nodeswave[0]=wave[0];
    int i=0;
    int z=1;
    while （i<specbands）
    {
      for （int j=i+1;j<specbands;j++）
      {
        float k=（spec[j]-spec[i]）/（wave[j]-wave[i]）;
        float   p=spec[j]-wave[j]*k;
```

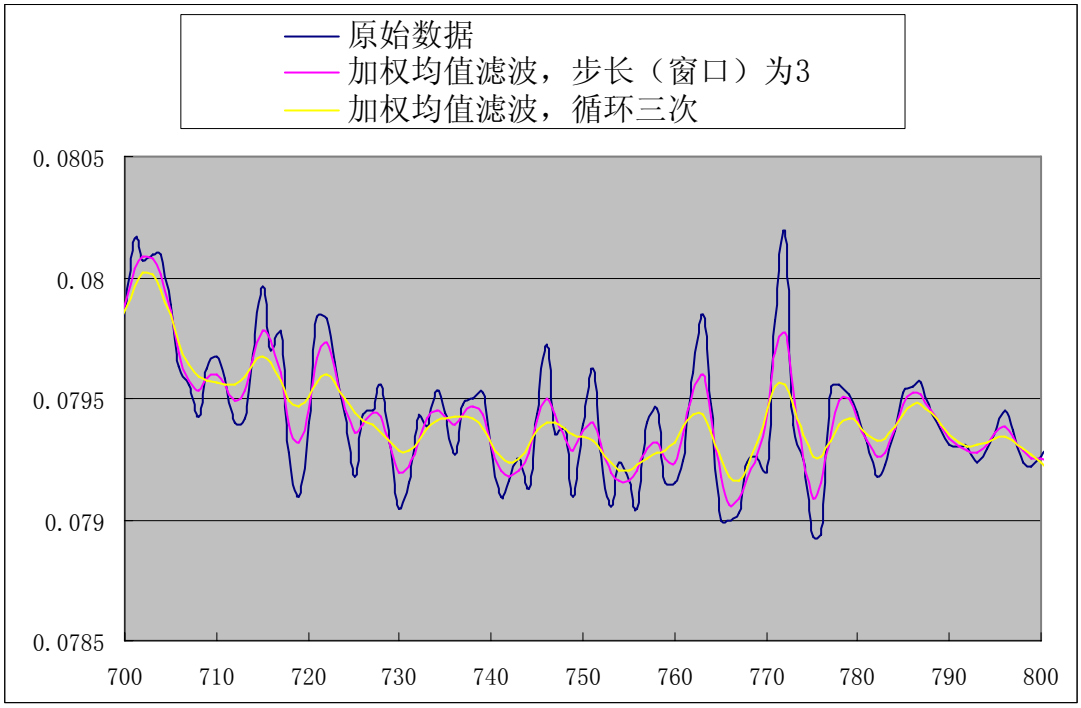
```
        for （int m=j+1;m<specbands;m++）
        {
          if （（wave[m]*k+p）>=spec[m]）
          {
            nodesspec[z]=spec[j];
            nodeswave[z]=wave[j];
            z++;
            i=j;
            break;
          }
        }
      }
    }
```

### 4.3.4 加权均值滤波方法

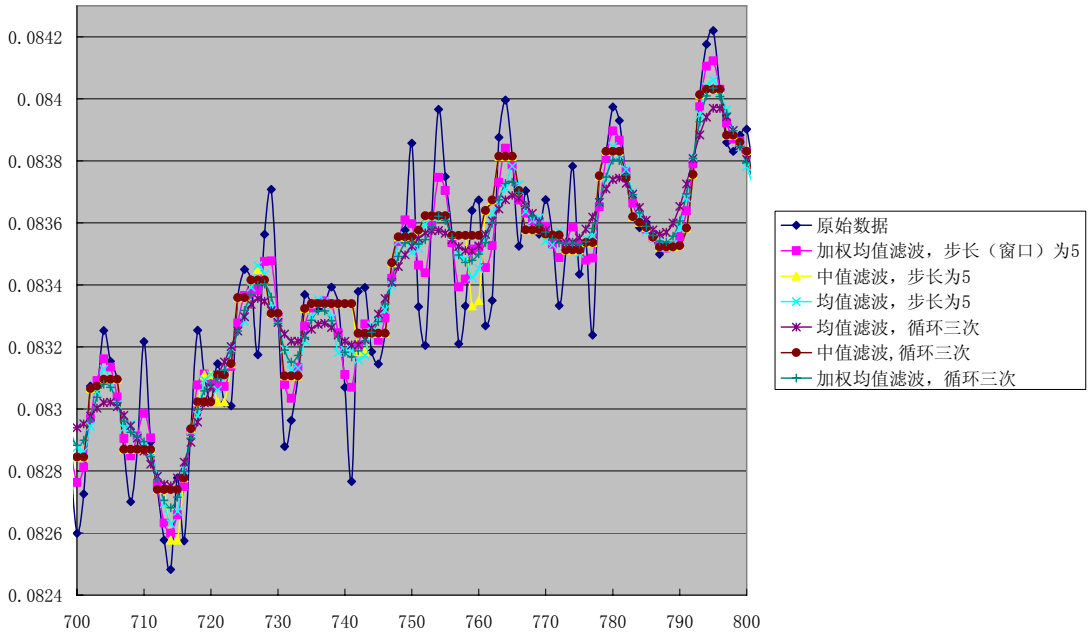
滤波算法有很多种，在光谱数据库中实现，主要是将光谱转换为数组，并利用PLSQL 语言实现具体的算法。本节列举了窗口（步长）为 3 的加权均植滤波算法的PLSQL 代码。从表 4.8，4.9 中可以看出，加权均值滤波相对均值滤波、中值滤波而言，能够较好的保持数据的边缘信息。

```
circle:=60;
newdata.trim（newdata.count）;
temp.trim（temp.count）;
newdata.extend;
newdata（1）:=data（1）;
temp.extend;
temp（1）:=data（1）;
for j in 2..data.count-1 loop
    newdata.extend;
    temp.extend;
    newdata（j）:=（data（j-1）+2*data（j）+data（j+1））/4;
    temp（j）:=newdata（j）;
end loop;
newdata.extend;
newdata（j+1）:=data（data.count）;
temp.extend;
```

```
temp（j+1）:=data（data.count）;
for i in 2..circle loop
    for j in 2..data.count-1 loop
        newdata（j）:=（temp（j-1）+2*temp（j）+temp（j+1））/4;
    end loop;
    for j in 2..data.count-1 loop
        temp（j）:=newdata（j）;
    end loop;
end loop;
```




图表 4.8 加权均值滤波算法平滑结果




图表 4.9 三种滤波算法和不同迭代次数的处理结果

## 4.4 高光谱数据库的典型应用模型

### 4.4.1 典型矿物波谱识别模型应用实例

- 输入数据：

被匹配光谱曲线：refl_mineral_Hematite_20040602024322_A1371.txt（赤铁矿）

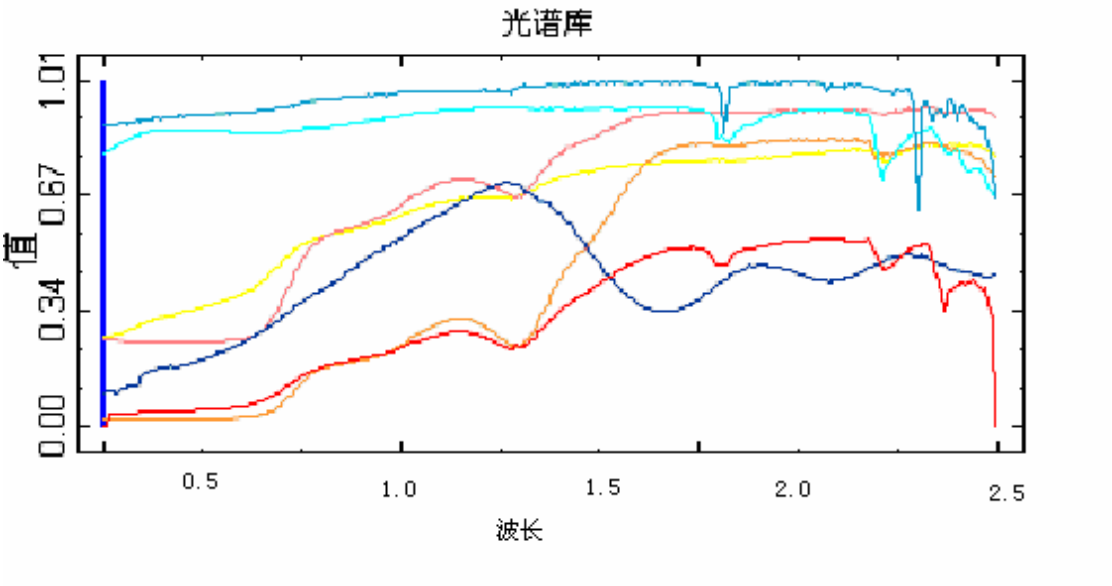
匹配光谱数据源：USGS 标准岩矿光谱库

- 测试结果：

| | |
|---|---|
| hematit1.spc Hematite 2%+98%Qtz GDS76, | 匹配度：0.9033 |
| hematit2.spc Hematite GDS27, | 匹配度：0.8977 |
| barite.spc Barite HS79.3B, | 匹配度：0.8805 |
| perthite.spc Perthite HS415.3B, | 匹配度：0.8544 |
| hematitb.spc Hematite WS161, | 匹配度：0.8459 |
| andalusi.spc Andalusite NMNHR17898, | 匹配度：0.8448 |
| goethit4.spc Goethite WS220, | 匹配度：0.8160 |
| cinnabar.spc Cinnabar HS133.3B, | 匹配度：0.8141 |
| microcl4.spc Microcline HS108.3B, | 匹配度：0.8026 |
| sphene.spc Sphene HS189.3B, | 匹配度：0.7949 |
| microcl1.spc Microcline HS82.3B, | 匹配度：0.7849 |

图 4.10 显示的是匹配算法中用到的部分矿物光谱。其中红色的是被匹配的物质

光谱曲线，橙色、淡橙色、黄色是匹配度排在前三的物质光谱曲线，淡蓝色、蓝色、深蓝色是匹配度居于最后三位的物质光谱曲线。从光谱曲线的波形上可以看出，匹配效果最高的，波形以及对应的光谱特征都很吻合。而相反，匹配度低的三条光谱曲线不论在光谱形状还是在征吸收位置上面，和待匹配光谱曲线都相差很远。因此，该矿物波谱识别模型能够很好地将矿物识别出来。

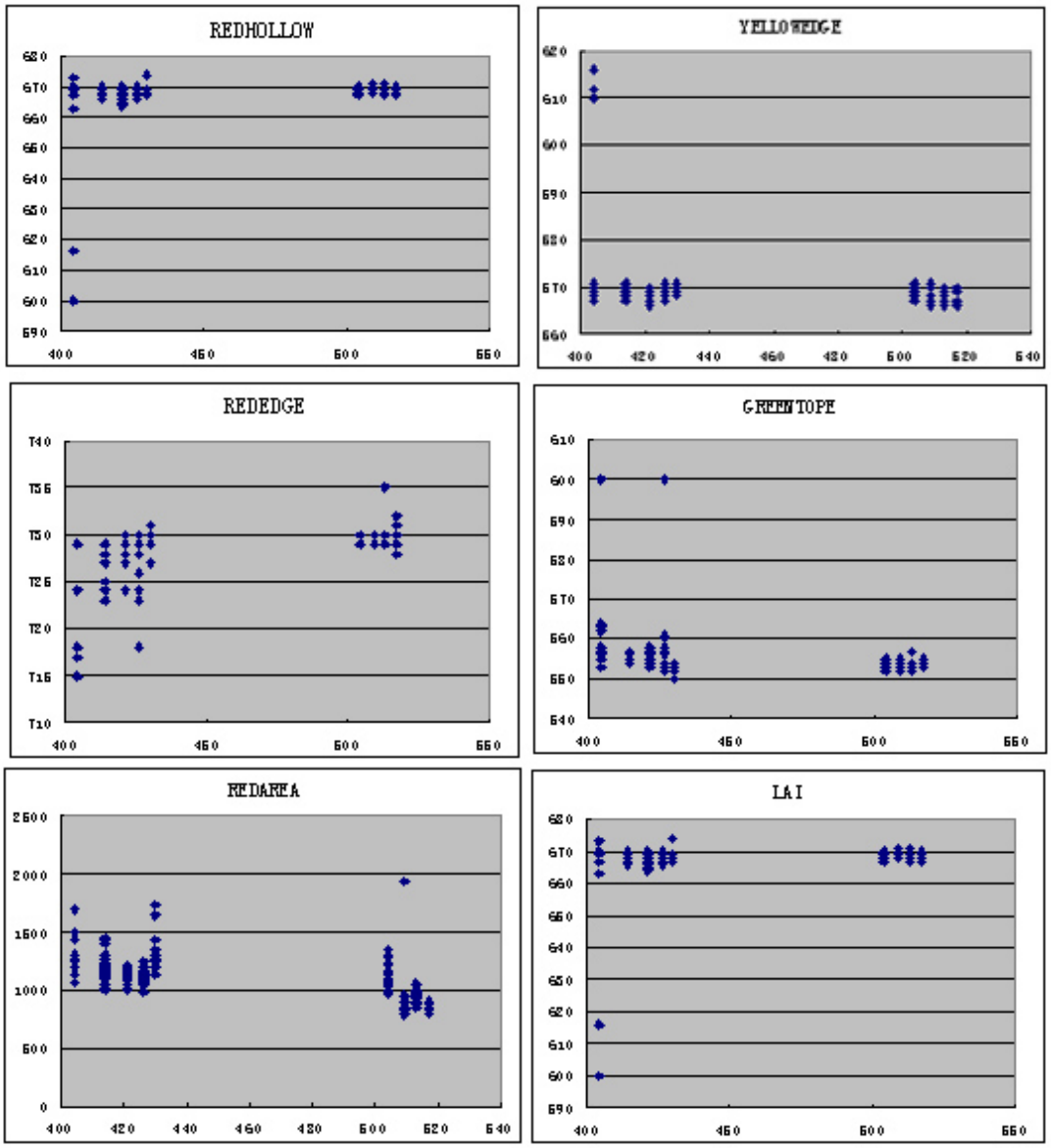



图表 4.10 矿物识别模型应用结果光谱曲线

refl_mineral_Hematite_20040602024322_A1371.txt, 红色
hematit1.spc Hematite 2%+98%Qtz GDS76, 橙色
hematit2.spc Hematite GDS27, 淡橙色
barite.spc Barite HS79.3B, 黄色
marialit.spc Marialite NMNH126018-2, 淡蓝色
topaz1.spc Topaz Wigwam_Area_A_#10, 蓝色
spessar2.spc Spessartine HS112.3B, 深蓝色

### 4.4.2 植被光谱维特征提取应用实例

对于植被光谱维特征，在本次应用实例中，主要选择了 4 月上旬到 5 月中旬的小麦光谱的红谷、黄边、红边、绿峰、红边面积、以及 LAI 数据，按照时间作了一次对比分析。根据数据库的自动化参量提取功能，我们可以迅速将光谱特征和被研究目标（理化参数、组分含量等等）做一些回归分析或者相关分析，从而发掘光谱特征的潜在意义。




图表 4.11 植被光谱特征参数提取（4 月 4 日至 5 月 17 日）

### 4.4.3 岩石矿物组分分析模型应用实例

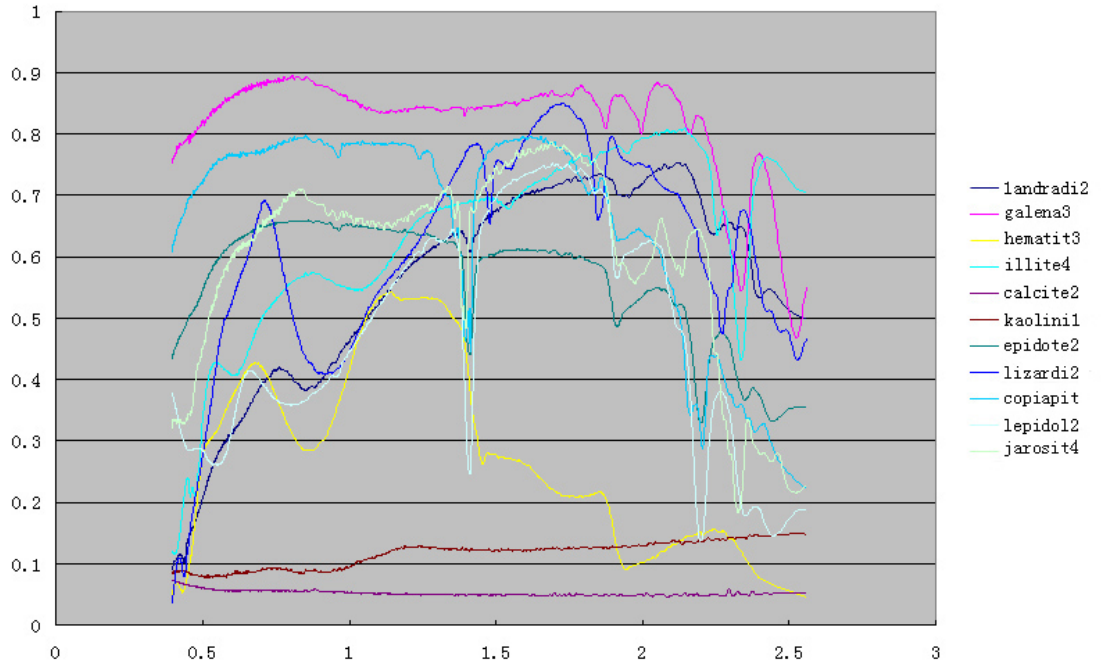
给定岩石光谱曲线 landradi 和其组分的光谱，通过组分分析模型，将其分解为组分光谱的百分含量，结果如表 4.1 示：

表格 4.1 岩石矿物组成分析应用结果

| 按组分排序 | 组分光谱名称 | 组分百分含量 |
|---|---|---|
| 1: | galena3 | 0.538183 |
| 2: | hematit3 | 0.236755 |
| 3: | illite4 | 0.0527622 |
| 4: | calcite2 | 0.0371711 |

| | | |
|---|---|---|
| 5: | kaolini1 | 0.0329593 |
| 6: | epidote2 | 0.0285074 |
| 7: | lizardi2 | 0.0272973 |
| 8: | copiapit | 0.0169116 |
| 9: | lepidol2 | 0.0164607 |
| 10: | jarosit4 | 0.0129924 |

由表可以看出，galena 组分和 hematit 组分占据了 landradi 的 76%，从岩石和矿物组分的光谱曲线也能够看出，galena 和 hematit 的光谱曲线在一定程度上基本控制了 landradi 的曲线形状。




图表 4.12 岩石光谱与其组分光谱曲线图

## 4.5 本章小结

本章主要落实了数据库的概念设计，在数据库物理层对数据库的数据结构、模型、方法进行了实现。通过对原有数据库结构的分析，将表群结构和大表结构，转换成为以光谱数据和影像数据为核心的双中心星型结构。实现了岩矿、水体、土壤等多源数据在数据库装存储，并针对高光谱影像数据在数据库中的存取作了自动化处理。实现了光谱数据和影像数据在存储方式和分析方式两种状态中的结构转换方法。对数据库中方法作了代码级别的实例，并对数据库中的三个模型作了具体应用阐述。通过对数据库建设，我们发现，数据库技术可以大幅度提高研究效率，通过数据库中的数据存储和分析自动化技术，可以有效地帮助我们做好数据准备和分析工作。

# 第五章 光谱数据挖掘

地物光谱数据主要包含光谱数据和属性数据。在获取数据之后，我们传统的工作模式是根据这两部分数据的基本统计数据，根据人的先验知识对二者之间的关系进行分析和发现。由于高光谱数据量大，信息丰富，数据之间关系复杂，传统的分析模式往往有一定的局限。数据仓库技术和数据挖掘技术可以将这些工作自动完成，并且深度挖掘数据之间的联系；如果有时间维上的数据积累，数据仓库还能够自动分析其发展趋势和规律。因此，不论从分析的自动化还是从信息利用的高效化和深层次化等各方面的考虑，数据挖掘技术，是高光谱遥感发展的必然需求，也将成为高光谱技术的重要组成之一。

## 5.1 光谱数据挖掘的定义与方法

数据挖掘技术从一开始就是面向应用研究的。目前，在很多领域，数据挖掘（data mining）都是一个很时髦的词，尤其是在如银行、电信、保险、交通、零售（如超级市场）等商业领域。数据挖掘所能解决的典型商业问题包括：数据库营销（Database Marketing）、客户群体划分（Customer Segmentation & Classification）、背景分析（Profile Analysis）、交叉销售（Cross-selling）等市场分析行为，以及客户流失性分析（Churn Analysis）、客户信用记分（Credit Scoring）、欺诈发现（Fraud Detection）等等。同样的思路被应用到科学研究之中，对于记录的分类、异常提取等等，这些都是有着共同的思想内涵。

光谱数据挖掘，其含义就是充分利用精细的光谱信息和辅助信息，发掘地物光谱曲线中微妙的变化并进行特征化、参量化，去粗取精，去伪存真，深入理解光谱信息与辅助参数的关系。对于光谱数据的挖掘，主要体现在挖掘和研究光谱曲线本身的特征和光谱曲线与对应的属性参数之间的关系。一般的研究可以有三个方向：一是通过光谱曲线及光谱特征来获取属性信息；二是通过属性信息来模拟光谱特征或者光谱曲线；第三，则是通过光谱曲线及特征和一部分属性信息来获取另外的属性信息。借助于数据库和数据仓库平台，我们可以利用数据库中的方法和模型来实现对光谱数据的信息挖掘。

本章将通过以下几个方面来阐述利用高光谱数据库进行光谱数据挖掘的应用模式。

## 5.2 岩石矿物光谱数据的模拟

区域地质制图和矿产勘探是高光谱技术主要的应用领域之一，地质是高光谱遥感应用最成功的一个领域（Vane Gregg，1993）。岩石矿物的波谱特性作为一种物理量来研究是从 50 年代起始，特别是 60 年代中，美国、日本等国家的一些实验室系统地、大量地测定了一些矿物、岩石的可见光至短波红外的光谱特性。自 70 年代起，美国一些学者陆续发表了包括矿物和岩石在内的可见光短波红外光谱特性专著，较

完整、系统地研究了岩石、矿物的光谱特性。各种矿物和岩石在电磁波谱上显示的诊断性光谱特征可以帮助人们识别不同的矿物成分（Crosta A P， 1997）。光谱数据库的存在对于航空、航天成像光谱数据的模拟提供了重要的数据支持。对于植被光谱的模拟，sails、prospect 模型及其相关的改进模型能够利用一些基础的生化参量实现对光谱曲线的模拟，而在岩矿领域，对于光谱的模拟研究还研究不多。

岩石的光谱特征相对矿物较为复杂，其波谱特征与矿物组成成分、结构、构造、风化等因素有关，往往岩石的光谱特征并不像矿物的光谱特征那样具有可鉴定的清晰的光谱特征。除去岩石的矿物组成成分和大气环境因素对岩石光谱反射率的影响之外，对于正常获取的岩石地面光谱数据而言，影响岩石光谱特性的因素还有以下三个方面：

（1）风化对岩石光谱反射率的影响

风化对原岩的成分、结构的改变是显而易见的。作用岩石受风化剥蚀的碎屑由水化作用生成水化物或多或少地残留覆盖岩石的表面。就沉积岩而言，由于风化后岩石的成分变化不大，风化面与新鲜面的光谱差异，主要表现在光谱反射率大小上。而在波谱形态上，由于 $Fe^{3+}, Fe^{2+}$ 的影响，在可见光部分变化略大，而在其它部分变化较小。对于透明物质，具有典型意义的是，减小粒度，反射率就会增大，光谱特征的对比度则减小。

（2）岩石表面结构对光谱反射率的影响

岩石表面结构对岩石光谱反射率有一定影响，在矿物成分基本相同时，矿物颗粒的粒度尺寸减小会导致光谱反射强度的增高，这是因为粒度愈小，它对入射光的散射愈强并减少了消光作用。在通常斜入射的情况下，细粒的矿物颗粒的微阴影覆盖的面积会变得更小，这样也提高了该表面的反射强度。

（3）岩石表面颜色对光谱反射率的影响

岩石的颜色是矿物成分、金属杂质及有机质含量的集中表现。不同种类的岩石由不同的矿物所组成，它们在颜色上是有差别的。一般来说，岩石颜色越深，说明以暗色矿物为主或含某些有机质（如炭质）杂质，则反射率亦低；岩石颜色越浅，说明以浅色矿物为主，含有机质少，则反射率亦高。岩石中的杂质成分往往反映在岩石的颜色上，进而影响该岩石的光谱反射率，有时甚至压抑掉该岩石的光谱特征。

中科院遥感所对新疆柯坪地区几类沉积岩的光谱特征研究表明（陈述彭、童庆禧，1998）：岩石的矿物成分与含量构成了岩石光谱特征的决定因素；岩石的赋成环境、表面物理性质及测量的时空条件等，都对其光谱反射率产生影响；同一种岩石虽然由于表面结构、颜色、粒度及化学成分等因素不一样，使其光谱反射率产生一定的差异，但其反射率光谱曲线形态基本不变。因此，通过岩石的矿物组分光谱及其含量实现岩石的光谱模拟是可行的。

遥感器所获取的地面反射或发射光谱信号是以像元为单位记录的。它是像元所

对应的地表物质光谱信号的综合。图像中每个像元所对应的地表，往往包含不同的覆盖类型，他们有着不同的光谱响应特征。而每个像元则仅用一个信号记录这些“异质”成分。若该像元仅包含一种类型，则为纯像元，它所记录的正是该类型的光谱响应特征或光谱信号；若该像元包含不止一种类型，则成为混合像元，它记录的是所对应的不同土地覆盖类型光谱响应特征的综合。从理论上讲，混合光谱的形成主要有以下原因：（1）单一成分物质的光谱、几何结构，及在像元中的分布；（2）大气传输过程中的混合效应；（3）遥感仪器本身的混合效应。其中：（2）、（3）为非线性效应，（2）大气的影响可以通过大气纠正加以部分克服；（3）仪器的影响可以通过仪器的校准、定标加以部分克服。第一部分原因造成的像元在光谱特性上的差异是光谱数据挖掘研究的内容之一。

光谱混合从本质上分可以分为线性混合和非线性混合两种模式。线性模型是假设物体间没有相互作用，每个光子仅能“看到”一种物质，并将其信号叠加到像元光谱中。而物体间的多次散射可以被认为是一个迭代乘积过程，是一个非线性过程；物体的混合和物理分布的空间尺度大小决定了这种非线性的程度；大尺度的光谱混合完全可以被认为是一种线性混合，而小尺度的内部物质混合是微非线性的。线性光谱混合模型包括物理学模型、数学模型和几何学模型（张兵，2002）：

线性光谱混合的物理学模型是指像元内部各物质成分的“纯”光谱的面积加权平均，$X = A\times\alpha + B\times\beta + C\times\gamma$，如图 5.1 所示，其中像元 $X$ 由 $A,B,C$ 三种物质混合而成。实际上，这里的“纯”也只是一个相对的概念，在一定空间尺度内被认为其物质组成是单一的。

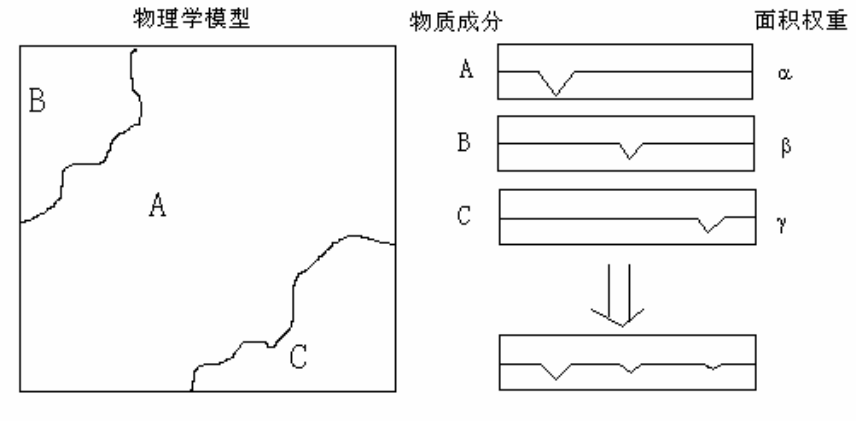



图 5.1 线性光谱混合模型的物理学描述

混合光谱的代数模型是指像元光谱矢量$\mathbf{C}$，是其所含所有端元光谱矩阵$\mathbf{A}$与各端元光谱丰度$\mathbf{B}$矢量的乘积。如图 5.2，像元光谱矢量$\mathbf{C}$就是像元光谱，为已知量。$\mathbf{A}$为整幅图像所有端元光谱组成的端元光谱矩阵，可以通过地面光谱实测或光谱库获得。矢量$\mathbf{B}$是未知量，反映的是一个光谱像元内各种端元所占面积比率。m 和 n 分别是波段数和端元数。

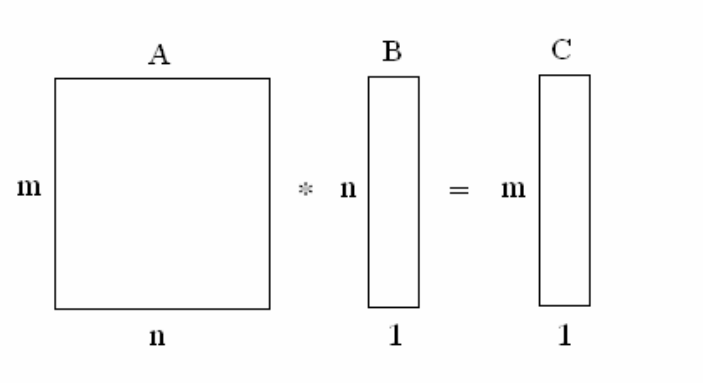



图 5.2 线性光谱混合模型的代数学描述

混合光谱的几何模型是由 Boardman 首先提出，他认为光谱数据在其特征空间（波段空间）呈现单形体的结构，继而引入了凸面几何学（Convex Geometry in N-Space）的分析方法，从特征空间对光谱线性混合进行诠释。以二维光谱为例，如图 5.3a 轴方向为波段方向，Y 轴为反射率方向，它显示了 A、B、C 三种不同端元所对应的 $i_,j$ 两个波段反射率值。图 5.3b 中 X 轴方向为波段 j 方向，Y 轴为波段 i 方向，A、B、C 三种物质在 $i_,j$ 两维光谱空间的分布构成一个三角形。可以看出，如果任何一个混合像元只要由这三种端元光谱组成，其光谱空间位置必然落在这个三角形区域（包括边线和顶角）。也就是说，三角形区域内的点均可被 A、B、C 三个点所数学表达。随着光谱维数的增加，就意味着光谱空间顶点的增加，就构成以端元为顶点的凸面几何空间。

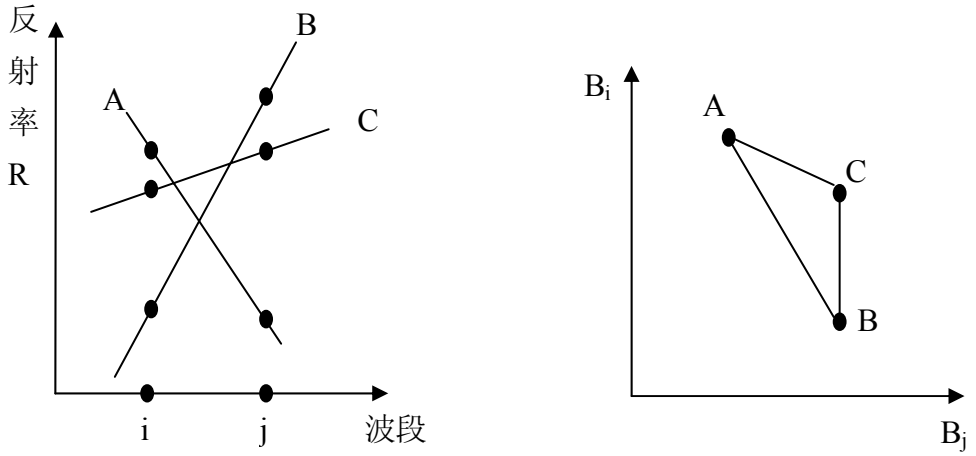



图 5.3：a 三种物质所对应的波段反射率，b.三种端元物质在两维光谱空间的分布

中科院遥感所在实验室和野外实验验证了线性混合光谱模型，均匀光照明、表面比较光滑的情况下，实验结果很好地符合线性混合光谱模型，对野外地面物体来说，由于其表面状态复杂、地面与大气以及地物之间的多次散射、阴影，还有仪器视场的不均匀等原因产生非线性效应，偏离线性模型，但基本上是符合的（陈述彭、童庆禧，1998）。

根据线性混合光谱理论（Adams, 1986）：如果每个光子与像元中的单个端元相

互作用，则混合模型是线性的，否则为非线性的。从广义的尺度角度出发，地面光谱仪的测量高度（距离）相对航空遥感和航天遥感而言，尺度更为细小，在一般的应用尺度而言，它所测量的光谱数据可以作为端元光谱。但从更加细微的尺度来考虑，岩石的光谱便可看作由更为精细的矿物光谱形成的混合像元光谱。对风化状岩石来说，其各种矿物的混合均基本保留各自吸收特征，在各种尺度和稳定视场内以线性关系表现。因此，我们把现行混合光谱理论扩展到岩石的地面测量光谱的模拟中来，即：把岩石的组成矿物作为端元，把其百分含量作为面积权重，通过其现行组合来模拟岩石的光谱曲线。

岩矿光谱数据模拟模型计算公式如下：

$$\begin{aligned} &w_1 + w_2 + \ldots + w_n = 1; \\ &ref_{out} = w_1 * ref_{in1} + w_2 * ref_{in2} + \ldots + w_n * ref_{inn}; \end{aligned} \qquad \text{公式 5.1}$$

其中，$w_i$ 为第 i 种矿物的百分含量，$ref_{ini}$ 是第 i 种矿物的光谱曲线（向量），$ref_{out}$ 是混合后岩石的光谱曲线。

为了对该模型进行验证，我们随机从波谱库中获取了两条具有矿物组成含量的地面测量波谱曲线。该岩石样本的矿物组成分别如表 1：

表 5.1 岩石样品的矿物组成成分

| 样品一：新疆哈密星星峡的斜长角闪岩 | 样品二：新疆哈密星星峡的石英闪长玢岩 |
|---|---|
| 【主要矿物名称】角闪石 80%、斜长石 15%<br>【次生矿物名称】石英、黑云母<br>【伴生矿物名称】磁铁矿 | 【主要矿物名称】角闪石 25%、石英 15%、斜长石 50%<br>【次生矿物名称】黑云母、白云母<br>【伴生矿物名称】褐铁矿、磁铁矿、榍石 |

同时，我们从标准的 USGS 数据库中找到角闪石，斜长石，石英三种矿物的光谱曲线作为组分光谱，如图 5.4，并将石英的含量暂定为 5%，忽略伴生矿物以及黑云母的影响，对该岩石标本进行光谱模拟试验。

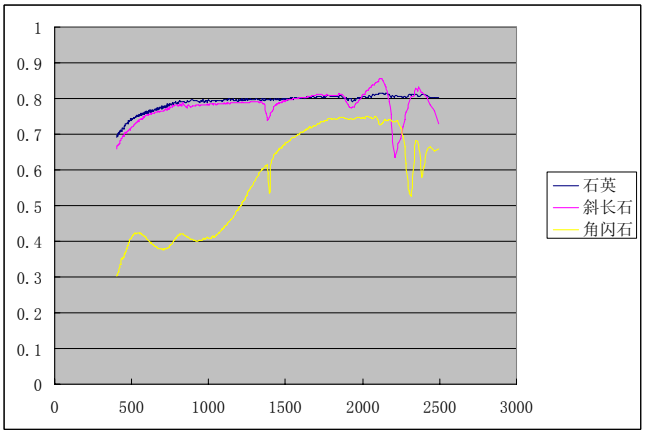



图 5.4 矿物组分石英、斜长石、角闪石的光谱曲线

模拟结果如图 5.5 所示，对模拟结果和原始光谱进行回归分析和两个向量之间的夹角计算，得到分析结果如表 5.2，

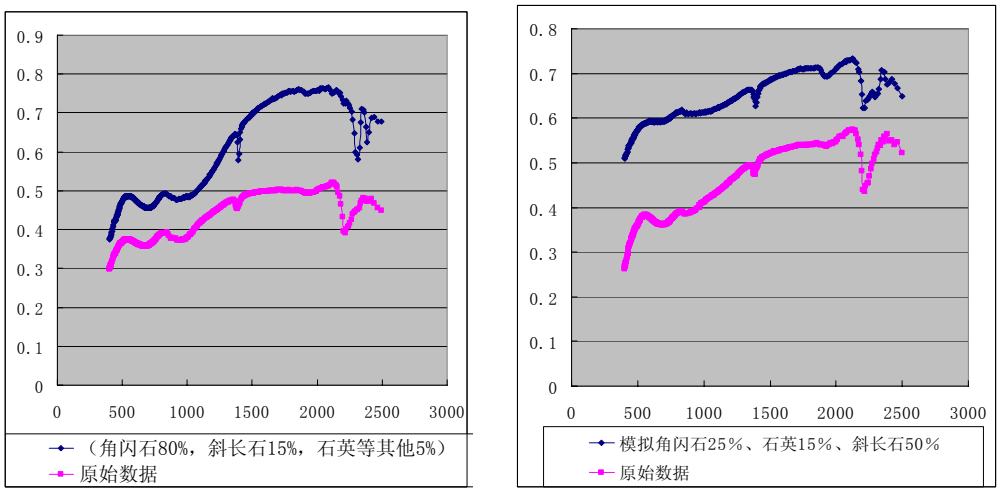



图 5.5 模拟试验，左图为斜长角闪岩，右图为石英闪长玢岩

通过分析模拟结果，可以看出模拟曲线具有以下几个特点：

第一：原始反射率与模拟反射率在波形上一致，二者向量夹角余弦在 0.99 以上，相似度高。

第二：模拟反射率在各波段的反射率值普遍比原始反射率数值高。岩石表面的颜色和粒度可能是造成此现象的主要原因。此外，也有可能与测量时的条件有关。

第三：模拟反射率数据在 Fe3+谱带（700nm），Fe2+谱带（1000nm），水谱带（1900nm），羟基震动谱带（2200nm，2400nm）附近的吸收特征保存完好。

通过本实验表明，通过对斜长角闪岩和石英闪长玢岩光谱曲线的模拟，基于线性混合光谱理论的岩矿光谱模拟模型能够较好的模拟岩石光谱曲线，同时并且能够保存各个矿物组分的吸收特征。

表 5.2 模拟数据与原始数据之间的相似分析

| 样品 | 斜长角闪岩 | 石英闪长玢岩 |
|---|---|---|
| 相关系数 | 0.950435228 | 0.980273021 |
| R 方差 | 0.903327123 | 0.960935196 |
| 调整 R 方差 | 0.903093048 | 0.960840608 |
| 标准误差 | 0.036543935 | 0.010326997 |
| 观测值 | 415 | 415 |
| 向量夹角余弦 | 0.99652583 | 0.99506400 |

本文通过线性混合模型来模拟岩矿的光谱曲线，并通过波谱库地面测量光谱数据及其对应的配套参数对模拟效果进行了解释与验证。岩石的矿物组成成分来模拟其光谱曲线具有可行性，通过实验表明，基于线性混合光谱理论的岩矿光谱模拟模型能够在一定程度上保存组成矿物的吸收特征，模拟的岩石光谱曲线和实地测量光

谱曲线能够反映出岩石的光谱特性。本模型仅仅考虑了矿物组成成分对岩石光谱的影响来对岩石光谱曲线进行模拟，对于其他因素，诸如风化状况、表面结构、颜色等等对岩石光谱具有重要影响的因素没作考虑，这是本模型需要改进的方向。由于光谱的复杂性和“同物异谱”、“同谱异物”现象的存在，通过对数据库内矿物组成含量对于岩石光谱曲线的模拟，还有待今后的进一步研究与完善。

## 5.3 蒙皂石含量与膨胀土光谱吸收参量相关关系挖掘

膨胀土是“隐藏的地质灾害”，主要造成轻型建筑物的损害，年均灾害损失达数亿元。40 多年来，国内外在膨胀土粘土矿物成分测试和研究、蒙皂石晶层胀缩规律、工程性质评价、膨胀土判别与分类和工程治理方面开展了大量研究与实践，形成了不同的膨胀土判别、分类方法。国内外大量研究结果表明, 蒙皂石晶层间的失水收缩、吸水膨胀是膨胀土胀缩性的根源，蒙皂石是膨胀土胀缩性的物质基础，蒙皂石和混层粘土矿物的识别和量化是膨胀土判别的关键因素之一。由于组成膨胀土的粘土矿物在可见光-近红外波段（0.4-2.5μm）都有诊断性吸收峰，目前，国际上在率先运用传统的技术方法，鉴别膨胀土的粘土矿物成分和判别膨胀势，并研究膨胀土的矿物成分与膨胀指标的相关关系后开展了膨胀土识别、填图的遥感技术研究, 以检验方便、快捷的遥感技术在膨胀土识别、填图方面的应用效能。（燕守勋，2005）

通过岩矿光谱吸收参数提取模型，在数据库中可以自动计算吸收位置，吸收深度，吸收宽度，吸收对称性和吸收面积等 5 个参量，同时把这些数据保存到数据库中的参量表，也可以将数据输出到外部文本文件，以方便非数据库应用人员使用。

根据 XRD 测量计算的有效蒙皂石含量划分的膨胀势，将剧、强、中、弱膨胀土的光谱进行归类（图 5.6， 图 5.7， 图 5.8， 图 5.9），发现不同类型的膨胀土的光谱特征显著不同，1900nm, 1400nm, 2200nm 吸收带的强度差异可以区分膨胀土类型：剧膨胀土 3 个吸收带的强度特征是：1900>1400>2200（图 5.6）；强膨胀土 3 个吸收带的强度特征是：1900>1400≈2200（图 5.7）；中膨胀土 3 个吸收带的强度特征是：1900>2200>1400（图 5.8）；弱膨胀土 3 个吸收带的强度特征是：1900≥2200 极>1400（图 5.9），且 1400nm 峰变宽平。从剧膨胀土到弱膨胀土，随蒙皂石含量的减少，1400nm 吸收峰强度逐渐减弱，2200nm 吸收峰强度逐渐增强，尽管 1900nm 吸收峰一直最强，但到弱膨胀土其强度与 1900nm 吸收峰已相近。

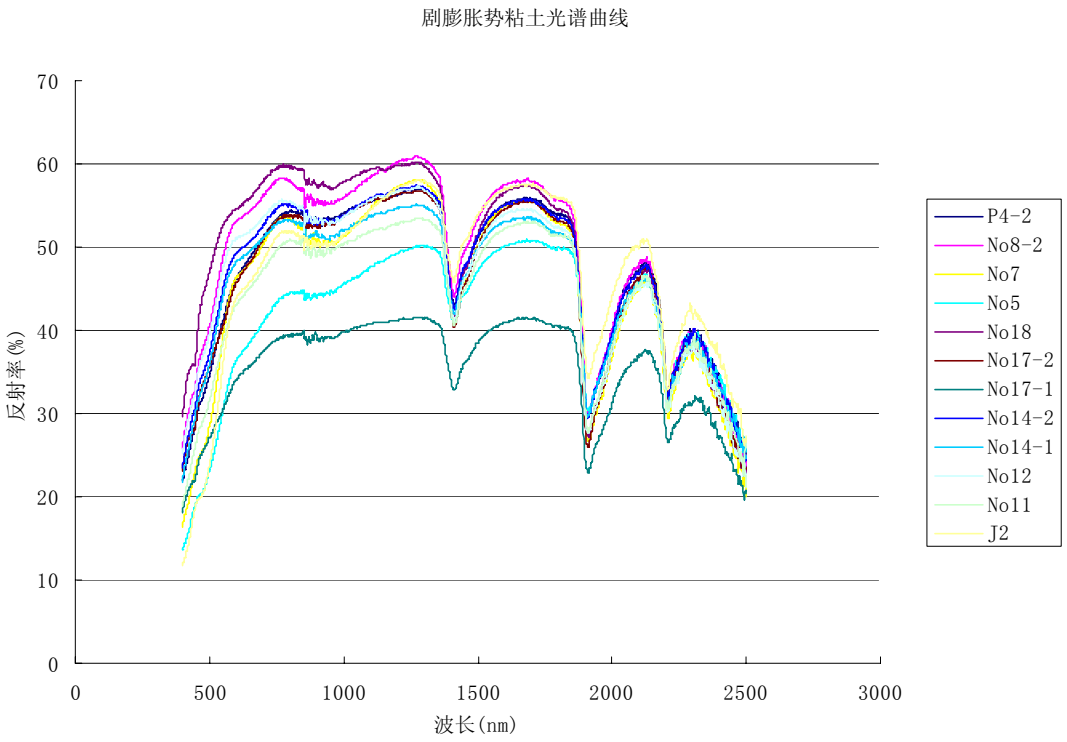



图 5.6 剧膨胀土光谱曲线

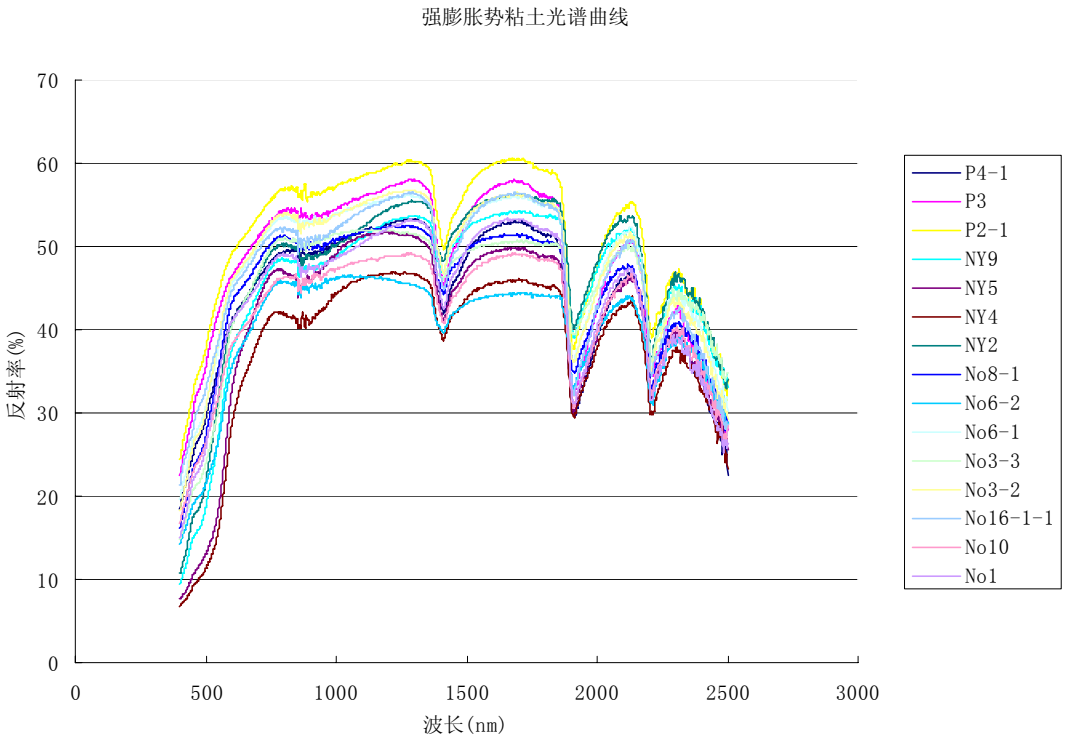



图 5.7 强膨胀土光谱曲线

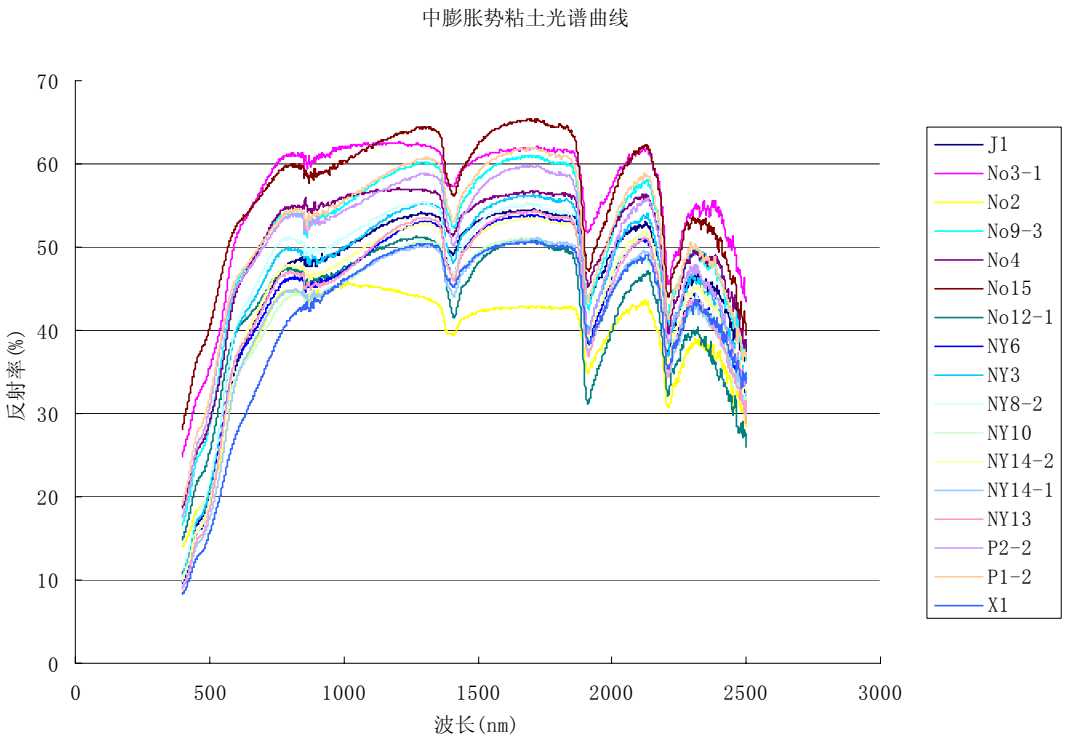



图 5.8 中膨胀土光谱曲线

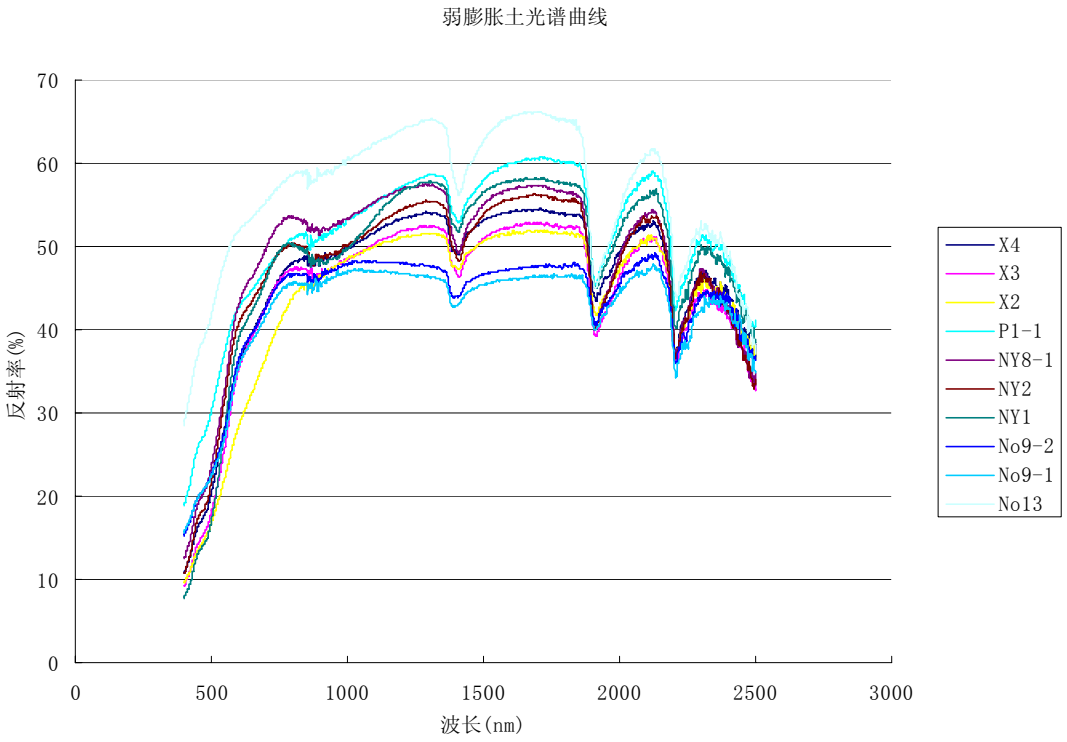



图 5.9 弱膨胀土光谱曲线

### 5.3.1 吸收位置分析

吸收位置是鉴定矿物的首要标志。测量样品出现 1400nm, 1900nm, 2200nm 吸收，是蒙皂石的诊断性波谱特征。统计表明，1400nm 吸收峰变化于 1390-1414nm 之间，相对稳定地集中在 1400nm 附近; 1900nm 吸收峰变化于 1910-1916nm 之间，相对稳定地集中在 1900nm 附近；2200nm 吸收峰变化于 2207-2215nm 之间，主要相对稳定地集中在 2210nm 和 2208nm 处；这 3 个吸收峰的位置变化与蒙皂石含量之间没有明显的变化规律。Kariuki P C（1999）的试验研究表明，在混合矿物中，端元矿物含量决定其吸收特征：在蒙皂石、伊利石、高岭石混合样品中，高岭石含量达到 40%以上，才出现其 2.16-2.17nm 和 2200nm 诊断性双吸收峰和 1400nm 处的双吸收峰；伊利石含量大于 50%，才出现其 2340-2350nm 诊断性吸收峰。由于本研究测量的样品高岭石含量为 0-19%，伊利石含量为 6-65%，没有出现高岭石的光谱特征，仅伊利石含量为 65%的样品 X2 出现了 2346nm 弱峰，而且与噪声难以区分；伊蒙混层与伊利石异物同谱，没有其光谱特征；绿泥石含量为 0-16%，其在 2300nm 处的吸收特征也未出现。本研究测量的样品仅明显表现了蒙皂石的光谱特征。

### 5.3.2 吸收深度分析

吸收深度是外包络线 Rc 与吸收谷 Rb 之间的差：D=Rc-Rb, 其取决于吸收矿物丰度和粒度。本研究的样品粒度是均一的小于 0.5mm 的粘土，因此，吸收深度主要与成分有关。用回归分析了 2200nm、1900nm、1400nm 吸收深度与蒙皂石含量和胶粒、粘粒含量之间的关系。结果表明：随蒙皂石含量、胶粒、粘粒含量的增加，2200nm 吸收深度稳定集中在 0.23-0.3nm 之间，没有明显变化规律；1900nm 吸收深度与蒙皂石含量之间呈明显的线性关系（图 5.10），拟合关系式为 Y=0.0082x+0.1212, 决定系数 R2=0.7627；与胶粒含量之间呈指数关系，拟合关系式为 Y=0.1e0.0261x, R2=0.6073; 与粘粒含量之间呈线性关系，拟合关系式为 y = 0.0071x - 0.0151，R2 = 0.6063；1400nm 吸收深度与蒙皂石含量之间呈明显的线性关系（图 5.11），拟合关系式为 y = 0.0055x + 0.036，R2 = 0.8015；与胶粒含量之间呈指数关系，拟合关系式为 Y=0.05e0.0268x, R2 =0.6204; 与粘粒含量之间呈线性关系，拟合关系式为 y = 0.0047x - 0.0572，R2 = 0.6255。

### 5.3.3 吸收宽度分析

吸收宽度是吸收谷的半高宽。统计结果表明，随蒙皂石含量、胶粒、粘粒含量的增加，2200nm 吸收宽度稳定集中在 60-70nm 之间，无明显变化规律；1900nm 吸收宽度与蒙皂石含量之间呈明显的线性关系（图 5.12），拟合关系式为 y = 0.5876x +

91.331，R2 = 0.66；与胶粒含量之间呈相关性小的线性关系，拟合关系式为 Y=0.4701+85.288, R2=0.4433; 与粘粒含量之间呈相关性小的线性关系，拟合关系式为 y = 0.4781x + 83.005，R2 = 0.4638；1400nm 吸收深度与蒙皂石含量之间呈明显的线性关系（图 5.13），拟合关系式为 y = 0.0055x + 0.036，R2 = 0.8015；与胶粒含量之间呈决定系数小的线性关系，拟合关系式为 Y=0.734x+67.056, R2 =0.6586; 与粘粒含量之间呈决定系数小的线性关系，拟合关系式为 y = 0.5747x +72.911，R2 = 0.4289。比较上述相关关系可知，1900nm 吸收宽度和 1400nm 吸收宽度与蒙皂石含量之间呈明显的线性关系，1900nm 吸收宽度比 1400nm 吸收宽度与胶粒含量和粘粒含量之间相关性小。

### 5.3.4 吸收对称性分析

吸收对称性是描述吸收波形的指标，其是吸收峰左半面积与右半面积的比值，当左半面小于右半面时，对称性小于 1；反之则大于 1。在高岭石、伊利石、蒙皂石混合波谱中，吸收对称性是鉴别不同膨胀图类型的指标 1。本研究中，仅为蒙皂石的光谱特征，统计结果表明，2200nm 吸收峰对称性小于 1，稳定集中在 0.25-0.45 之间，这与该吸收峰的右偏峰特征一致；2200nm 吸收峰对称性与蒙皂石含量、胶粒、粘粒含量之间无明显变化规律；1900nm 吸收对称性集中稳定在 0.15-0.25 之间，与蒙皂石含量、胶粒、粘粒含量之间无明显变化规律；1400nm 吸收对称性集中稳定在 0.1-0.2 之间，与蒙皂石含量、胶粒、粘粒含量之间无明显变化规律。2200nm、1900nm、1400nm 吸收对称性全部小于 1，这与蒙皂石的右偏峰鉴定标志相一致。

### 5.3.5 吸收面积分析

吸收带的面积是吸收形状的描述参量。统计结果表明，2200nm 吸收面积稳定集中在 250-350 之间，离散度小，与蒙皂石含量之间呈弱的负相关，相关性公式为：y=-1.1584x+303.54, R2=0.1101; 1900nm 吸收面积离散度大，与蒙皂石含量之间呈弱的负相关，相关性公式为：y=-1.3262x+513.89, R2=0.372（图 5.14）; 1400nm 吸收面积稳定集中在 600-700 之间，离散度小，与蒙皂石含量之间呈弱的负相关，相关性公式为：y=-2.6822x+717.22, R2=0.2347。总之，2200nm、1900nm、1400nm 吸收面积与蒙皂石含量之间呈弱的负相关关系，2200nm 、1400nm 吸收面积相对稳定集中，1900nm 吸收面积相对离散度大。它们与胶粒、粘粒含量之间具有类似的关系。

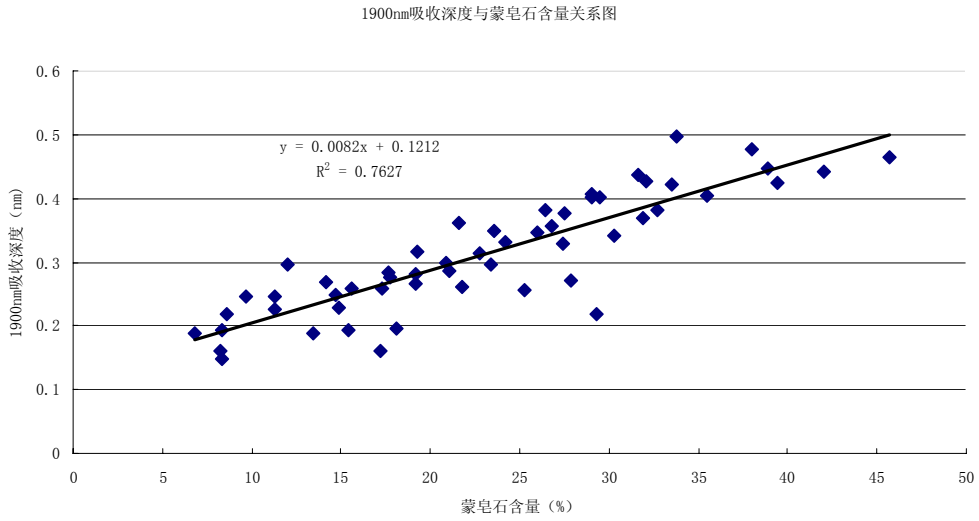



图 5.10 膨胀土 1900nm 吸收深度与蒙皂石含量相关关系图

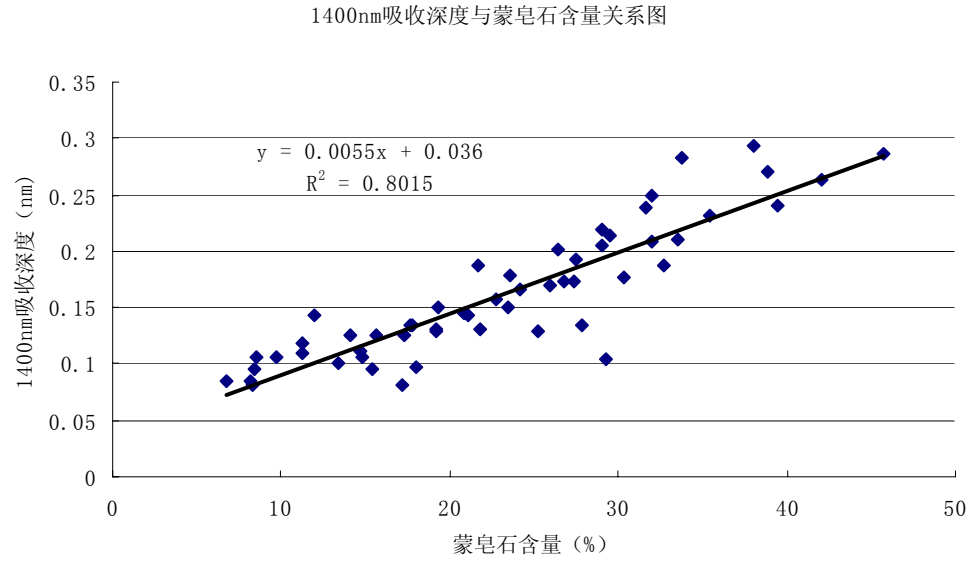



图 5.11 膨胀土 1400nm 吸收深度与蒙皂石含量相关关系图

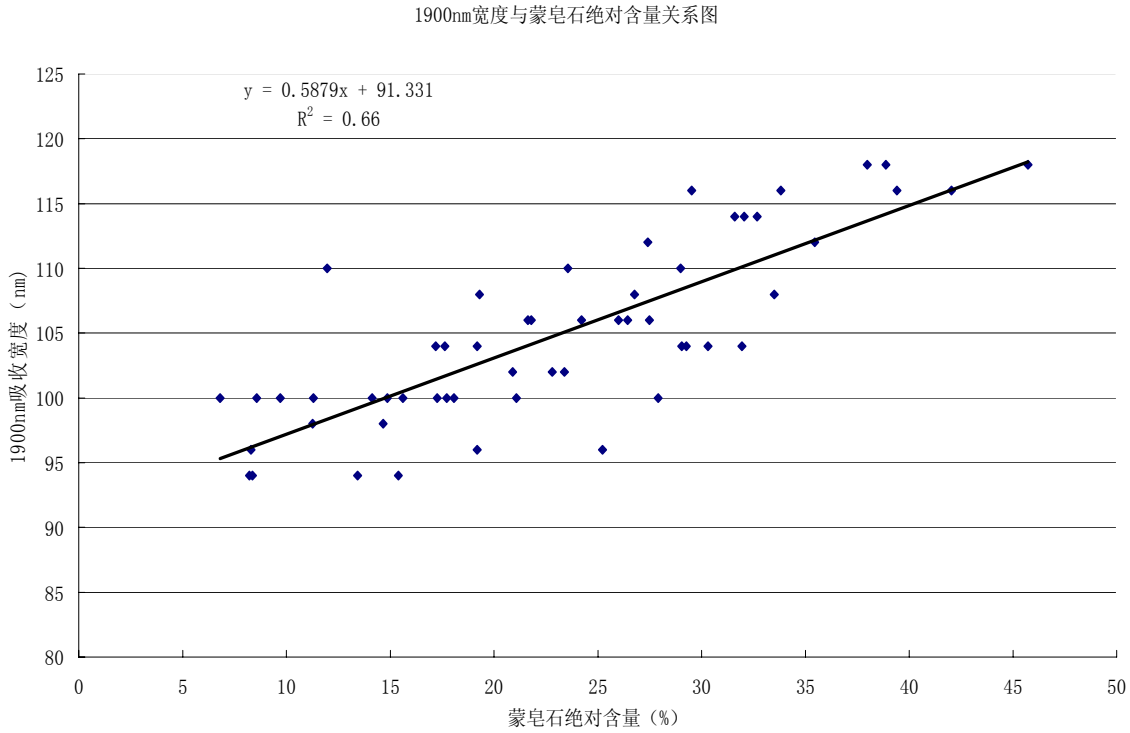



图 5.12 膨胀土 1900nm 吸收宽度与蒙皂石含量相关关系图

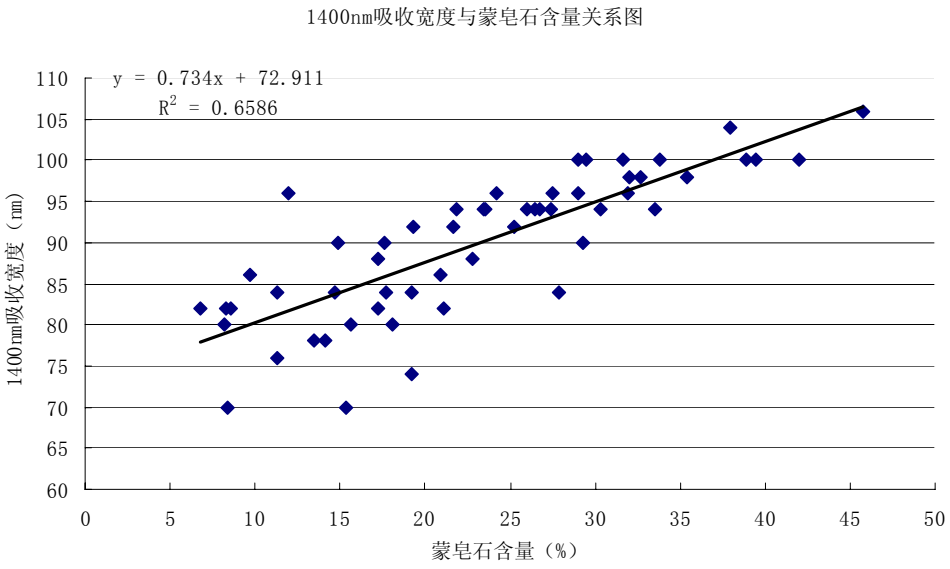



图 5.13 膨胀土 1400nm 吸收宽度与蒙皂石含量相关关系图

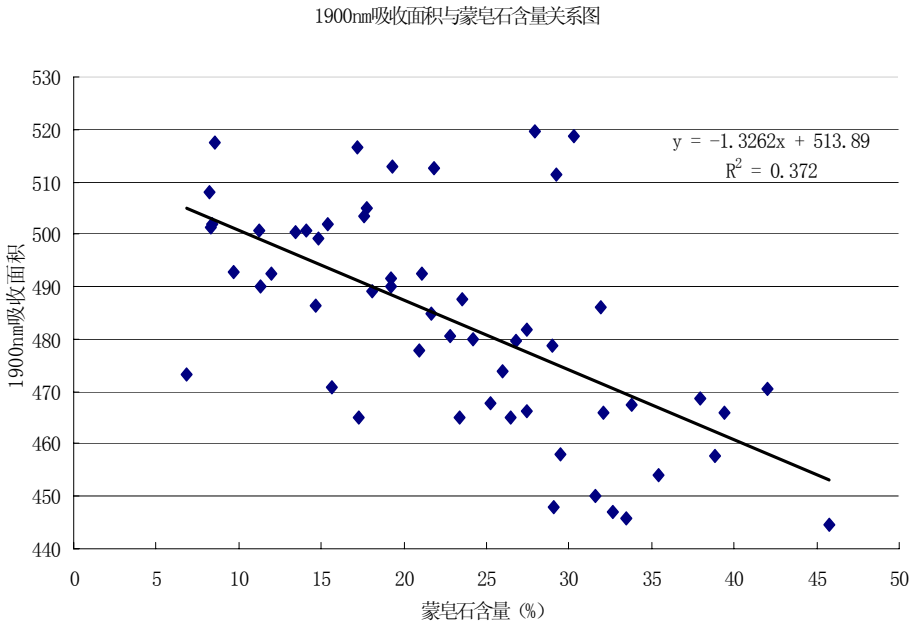



图 5.14 吸收面积与蒙皂石含量相关关系图

本文膨胀土实验室光谱研究表明，膨胀土光谱曲线主要反应决定膨胀性的蒙皂石的光谱特征，不出现含量小的其它矿物，如高岭石、伊利石、绿泥石等的吸收特征；根据蒙皂石 2200nm、1900nm、1400nm 吸收的相对强度，可以划分不同的膨胀土类型。从剧膨胀势到弱膨胀势粘土，随蒙皂石含量的减少，1400nm 吸收峰强度逐渐减弱，2200nm 吸收峰强度逐渐增强，尽管 1900nm 吸收峰一直最强，但到弱膨胀土其强度与 1900nm 吸收峰已相近。1900nm 是水的吸收，1400nm 是水和羟基的吸收，2200nm 是羟基吸收。膨胀土蒙皂石含量的变化在光谱上表现为水和羟基吸收强度的变化；1900nm、1400nm 吸收深度和吸收宽度与蒙皂石含量之间具有明显的变化规律，因此，根据膨胀土实验室光谱 1900nm、1400nm 吸收深度和吸收宽度可以估算膨胀土的蒙皂石含量; 根据膨胀土实验室光谱 1900nm、1400nm 吸收深度还可以估算膨胀土的胶粒、粘粒含量。而且，1900nm 与 1400nm 吸收深度之间呈对数关系，关系式为：y = 0.2518Ln（x） + 0.7858，R2 = 0.9851（图 5.15）；1900nm 与 1400nm 吸收宽度之间呈线性关系，关系式为：y = 0.7099x + 41.115，R2 = 0.7871（图 5.16）。根据膨胀土实验室光谱 1900nm、1400nm 吸收深度和吸收宽度估算蒙皂石含量及胶粒、粘粒含量，两者可以相互印证。

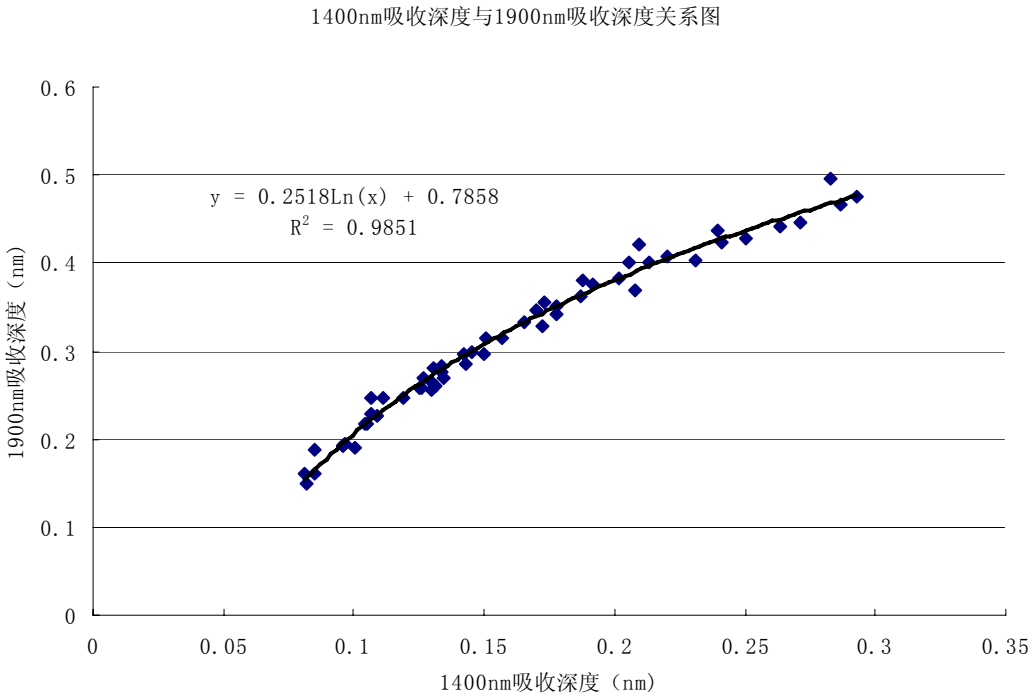



图 5.15. 膨胀土 1400nm 吸收深度与 1900nm 吸收深度相关关系图

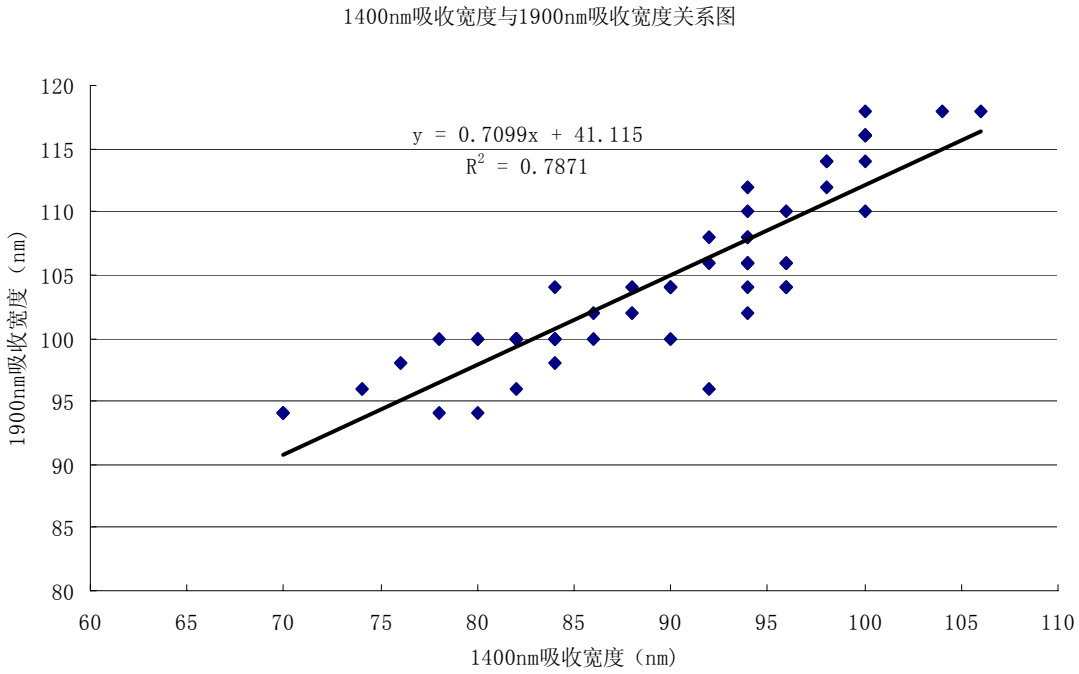



图 5.16. 膨胀土 1400nm 吸收宽度与 1900nm 吸收宽度相关关系图

膨胀土蒙皂石 2200nm 、1900nm、1400nm 吸收对称性均小于 1，这是蒙皂石右偏峰的典型鉴定标志；吸收对称性不能运用于与蒙皂石含量及胶粒、粘粒含量的光谱估算。

## 5.4 光谱波段组合自动优化

随着高光谱遥感应用的不断深化，各种传统的光谱指数也面临着重新定义。以归一化植被指数 NDVI 为例，NDVI 最早被定义为：$NDVI = (R_{NIR} - R_{Red})/(R_{NIR} + R_{Red})$，其中 RNIR, RRED 分别代表的是物体在近红外和红光波段的反射率。然而，随着光谱分辨率的提高，红波段和近红外波段也得以细化，在红光波段和近红外波段的反射率值可以由数十甚至数百个至来综合反映。高光谱遥感在大幅度提高地物信息量的同时，也对我们提取光谱指数带来了更大的工作量和更多需要考虑的问题。在众多的波段中，如何选择最合适的波段组合来构造光谱指数，也是高光谱遥感应用研究的一个重要问题。

在高光谱数据库中，我们可以利用数据库技术中的自动数据处理方法，对波段组合进行穷举，构造出所有可能的波段组合，然后根据得到各种指数对应的结果，和对应的参量进行回归分析，并返回关联最好的波段组合，然后形成对光谱指数的波段组合推荐方案。

本节利用小汤山精准农业基地小麦的 151 条光谱曲线和对应的 LAI 来寻找和优化最能反映小麦 LAI 的 NDVI 指数。通过数据库的存储过程，我们只需要通过指定数据表的样本记录和采样间隔，就能够实现对波段组合的优化，并能够将优化结果输入到指定的结果数据表中。以下存储过程代码实现了从数据库中读取数据、转换光谱数据为表数据、获取波段组合并计算 NDVI，计算 NDVI 和 LAI 的相关系数，将计算结果存储到结果表中的所有过程。

通过指定不同的间隔采样效率，我们可以得到在不同分辨率下的最佳波段组合前十名和所需要的运算时间。样本数据：小汤山小麦光谱数据 151 条，波长范围 350-2500，光谱分辨率：1nm，设定红光波段范围（620，760）、短波红外波段范围（780，1100）。数据挖掘结果如以下五个表所示：

表格 5.3 20NM 间隔采样，运算时间为 9.718 秒，136 个组合

| 排名 | 波段组合 | 相关系数 |
|---|---|---|
| 1 | 711-791 | 0.60463902 |
| 2 | 711-811 | 0.59951442 |
| 3 | 731-791 | 0.59685848 |
| 4 | 711-831 | 0.59518729 |
| 5 | 711-851 | 0.59254444 |
| 6 | 711-871 | 0.58874996 |
| 7 | 711-891 | 0.58439778 |
| 8 | 731-811 | 0.58218550 |
| 9 | 711-911 | 0.57787505 |
| 10 | 731-831 | 0.57032682 |

表格 5.4　10NM 间隔采样，运算时间为 29.344 秒，495 个组合

| 排名 | 波段组合 | 相关系数 |
|---|---|---|
| 1 | 721-781 | 0.61473468 |
| 2 | 721-791 | 0.61049682 |
| 3 | 711-781 | 0.60728458 |
| 4 | 721-801 | 0.60679181 |
| 5 | 731-781 | 0.60476738 |
| 6 | 711-791 | 0.60463902 |
| 7 | 721-811 | 0.60249925 |
| 8 | 711-801 | 0.60229266 |
| 9 | 711-811 | 0.59951442 |
| 10 | 721-821 | 0.59818387 |

表格 5.5　5NM 间隔采样运算时间为 95.922 秒，1885 个组合

| 排名 | 波段组合 | 相关系数 |
|---|---|---|
| 1 | 721-781 | 0.61473468 |
| 2 | 726-781 | 0.61351278 |
| 3 | 716-781 | 0.61314575 |
| 4 | 721-786 | 0.61256755 |
| 5 | 716-786 | 0.61147071 |
| 6 | 726-786 | 0.61063750 |
| 7 | 721-791 | 0.61049682 |
| 8 | 716-791 | 0.60987826 |
| 9 | 721-796 | 0.60858748 |
| 10 | 716-796 | 0.60839944 |

表格 5.6　2NM 间隔采样，运行时间 480.547 秒，11431 个组合

| 排名 | 波段组合 | 相关系数 |
|---|---|---|
| 1 | 719-781 | 0.61568469 |
| 2 | 719-783 | 0.61501255 |
| 3 | 721-781 | 0.61473468 |
| 4 | 723-781 | 0.61439045 |
| 5 | 717-781 | 0.61431135 |
| 6 | 719-785 | 0.61419327 |
| 7 | 725-781 | 0.61400588 |
| 8 | 721-783 | 0.61398246 |
| 9 | 717-783 | 0.61370606 |

| | | |
|---|---|---|
| 10 | 723-783 | 0.61355179 |

表格 5.7　1NM 间隔采样，运行时间 2002.704 秒，45261 个组合

| 排名 | 波段组合 | 相关系数 |
|---|---|---|
| 1 | 719-780 | 0.61596899 |
| 2 | 719-781 | 0.61568469 |
| 3 | 720-780 | 0.61557604 |
| 4 | 718-780 | 0.61542175 |
| 5 | 719-782 | 0.61533446 |
| 6 | 720-781 | 0.61527549 |
| 7 | 718-781 | 0.61515301 |
| 8 | 721-780 | 0.61505269 |
| 9 | 719-783 | 0.61501255 |
| 10 | 720-782 | 0.61490436 |

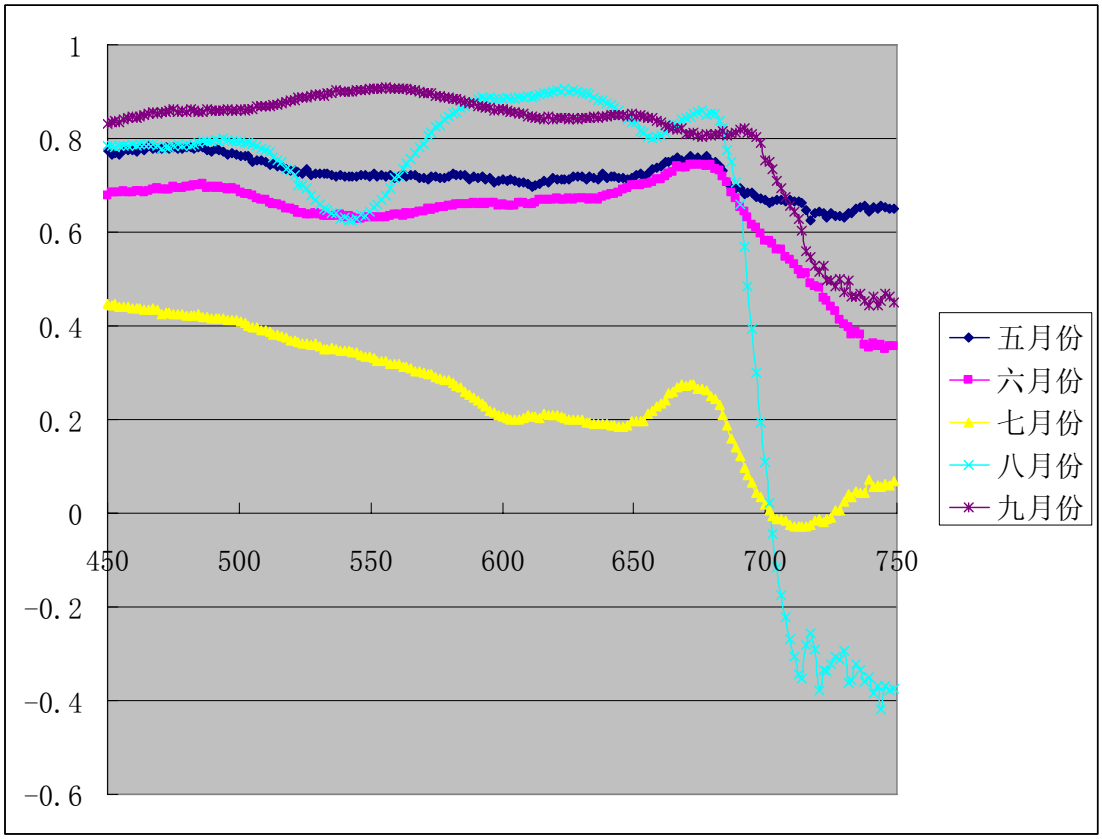
由以上五个不同光谱分辨率的数据挖掘结果表可以看出，随着光谱分辨率越高，获取的 NDVI 和 LAI 的相关系数也越大，而运行时间也呈几何级数增加。组合集中元素个数由 136 个增加到 45261 个，而运行时间也由 9.718 秒增加到 33.34 分钟，NDVI 与 LAI 的最佳相关系数从 0.60463902 上升到 0.61596899，最佳波段组合由 711-791 优化为 719-780。因此，LAI 对于光谱分辨率的变化是敏感的，在光谱分辨率提高的同时，最优的 NDVI 指数对于 LAI 的反映也更加准确。

同样，我们在进行水质参数反演时，同样需要做波段选择和模型优化的工作。

本文通过对太湖地区水质光谱和叶绿素 A 含量的相关关系挖掘，来实现对反演模型的优化。首先，我们将太湖地区测量的五个不同月份光谱数据输入数据库（通常，把保存在 EXCEL 表中 100 条以内的光谱数据导入到数据库中，全部过程不超过五分钟），在导入光谱数据的同时，我们也把样本数据对应测量的叶绿素 a 数据导入到数据表中。第二，我们确定数据挖掘的题目为：对光谱全波段和叶绿素 a 做相关关系分析，从而获得最能够反映叶绿素 a 含量变化的光谱波段或者光谱波段组合。确定挖掘目标之后，我们要对数据进行准备，即将 blob 格式存放的光谱数据转为光谱数据表，同时建立涵盖光谱数据表和叶绿素 a 的数据视图。第三，在数据库中，调用相关关系分析方法，抽取样本数据和叶绿素 a 进行相关分析，将结果输出到相关关系结果表。图 5.17 为数据挖掘处理的初步结果。从图中可以看出，七月份的数据存在异常，而八月份的数据部分波段存在异常。经过验证，在数据测量的七八月份，太湖地区均为阴雨天气。从五、六、九月份的数据可以看出，波长 500，680 均为能够反映叶绿素 a 的特征波段，对于五、六月份而言，545 是相关关系变化的拐点，而对九月份数据而言，555，620，660，700 都是相关关系变化的拐点。因此，在分析太湖地区水质参数反演的时候，在不同月份，光谱曲线对于叶绿素 a 的相应

程度最强烈的波段不尽相同。

在本次数据挖掘中，最开始时将所有五个月份数据作挖掘分析，结果发现相关性并不是很好，于是按月份查询数据并建立数据视图，通过逐月分析，发现相关性不明朗的原因。正是由于数据库查询检索的便捷性以及集成化方法的优越性，我们对于数据的分析变得简单、高效而多元。




图表 5.17 光谱全波段数据与叶绿素 a 的相关关系图

由以上两个应用案例可以看出，通过数据库，我们可以大大缩短光谱波段组合的优化过程，直观的获取最优波段组合，同时，也能够从不同的数据集中获取潜在的信息，来为我们科学研究提供新的线索，从而为不同行业应用提供直接的决策支持。

## 5.5 本章小结

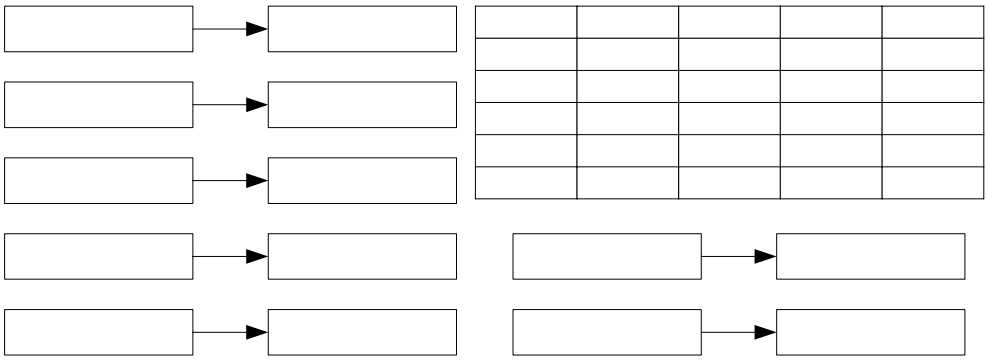
本章阐述了光谱数据挖掘的定义与方法，并通过三个方面可以对光谱数据挖掘进行应用研究。从属性到光谱方向的研究，本章通过对矿物组分光谱数据模拟岩石光谱作了尝试；从光谱到属性方向的研究，本章通过蒙皂石含量和膨胀土光谱吸收参量的相关关系分析做了一些应用；而对于光谱和属性综合研究，则通过了植被 LAI 来控制 NDVI 的波段组合，水体叶绿素含量和波段选择两个研究案例来作一定的探讨一些自动的优化方案提取。

# 第六章 高光谱影像数据挖掘

高光谱影像数据有四个基本的参数：空间分辨率、时间分辨率、光谱分辨率、辐射分辨率。在高光谱传感器制作完成后，辐射分辨率和光谱分辨率基本固定。如果是星载，则空间分辨率和时间分辨率一般固定。如果是机载，空间分辨率和时间分辨率则会随着飞行的高度和频率发生变化。高光谱在经过辐射校正和几何校正，以及根据需求在空间维和光谱维重采样等一般的预处理过程后，数据一般不会再发生变化，而与高光谱数据同步或者后期测量的其他参数数据也不会发生变化。高光谱数据的相对稳定性和数据的高维性为高光谱数据库的建设和数据挖掘提供了必要的基础。

## 6.1 高光谱影像数据挖掘的定义与方法

高光谱影像数据和其他遥感数据最显著的区别是：高光谱影像波段数成百上千。而这一典型的数据特点，成为了高光谱影像数据挖掘的根本优势。我们把高光谱影像数据和数据库的数据表做一个映射，我们便可以得到如图 6.1 所示的一个映射情况：



图表 6.1 高光谱影像数据与数据表的对应关系

当我们把一幅高光谱影像数据当作一张数据表来看待时，一幅高光谱影像便成为了由成百上千个字段组成，包含上百万条记录的一张二维数据表。每一个像元对应一条记录，每一个波段对应一个字段。而像元之间的匹配运算变成了记录之间的相似性查找，影像的光谱特征提取变成了数据表的特征提取，影像的降维变成了数据表的降维，影像的分类聚类变成了数据表的分类聚类。更加重要的是，而这些记录的值具有严格的结构化属性，即统一的数值型。这为将数据库领域的数据挖掘方法应用到影像数据挖掘，提供了良好的前提条件。在具体数据挖掘过程中，我们可以根据像元号把图像数据表和对应的应用数据表创建关联视图，从而进行针对各种不同应用目的的数据挖掘研究，同时，我们也可以把各种数据挖掘方法应用到遥感影像数据中，实现对遥感影像数据的深度信息挖掘。本章将把传统数据挖掘领域的属性重要性评价、特征提取和分类方法应用到高光谱影像数据中，实现对高光谱影像数据的几个自动化处理过程。

高光谱影像 数据表



像 元 记 录

## 6.2 基于最小描述长度模型的高光谱影像波段选择

高光谱影像中光谱波段是具有具体的物理含义的。随着遥感器技术的进步，高光谱影像的波段数也不断增加，利用高光谱能够识别的地物精细信息也不断增多。高光谱遥感数据大量的光谱波段为我们了解地物提供了极其丰富的遥感信息，这必然有助于我们完成更加细致的遥感地物分类和目标识别，然而波段的增多也必然导致信息的冗余和数据处理复杂性增加。在高光谱数据的处理过程中，人们发现，高光谱数据大部分波段间的相关性很强，信息的冗余度很大，实际处理的过程中往往要进行波段选择：从所有的波段中选择一些信息含量大、信噪比好、使目标可分离性高的波段，再进行分类或识别处理。一方面减少了运算量，加快处理时间，克服了训练样本少的困难；另一方面，波段选择处理也提高了分类和识别的精度。波段选择往往分为从光谱空间进行和从几何空间进行两种。从光谱空间进行波段选择，往往是基于光谱吸收特征，利用一些光谱处理方法（如包络线去除等）和极值检索方法来获取光谱的吸收位置，以此作为特征波段。

在针对具体应用时，波段选择往往是通过线性回归方法来完成（刘伟东，2002），它包括了能够最好解释其它光谱波段信息的波段。整个过程始于选择一个与其它光谱波段最相关的波段具有最小的残差平方和，然后使用第一个入选波段与另一未知波段去表示其它波段，如果残差平方和最小该波段即选入的第二波段，重复该过程直到确信残差低于实验的不确定性。另外，张兵提出了基于光谱特征选择的图像波段合成（张兵，2003）。美国 West Virginia University 的 Timothy A. Warner 等（Timothy A. Warner et al.,1996）用影像空间自相关来做波段选择，即最后所选出的波段中的各地物集合区域灰度反差越大越好，各地物集合区域越集中越均匀越好，波段之间越独立越好。比较两两波段比值图像的空间自相关值，值越小，说明构成比值图像的两个原始波段图像信息量越大，两波段应入选。刘建平等分析了多光谱遥感数据最佳波段选择的联合熵、行列式值厦最佳指数等信息量计算方法的内在联系，提出了基于类间可分性的最佳波段选择原则和方法（刘建平，2001）

在数据库和数据仓库中，也同样存在“波段选择”的问题。在对数据库中的海量数据表进行分析决策时，面对的往往是大量的属性字段信息。在建立应用模型时，需要减少数据量，对属性字段进行选择，把研究重点放在“热点”字段上，由此就诞生了属性重要性这一概念。在对属性重要性的评价模型中，通过对各个属性字段的相对重要性，或者对指定目标字段的影响程度来对属性排序。我们把这种数据挖掘方法应用到高光谱影像中，针对具体应用对高光谱影像的各个波段进行重要性排序，从而实现对高光谱影像在具体某一应用方向的波段重要性评价。通过对波段的重要性评价来获取对应用方向影响最大的波段，从而实现波段选择。

本节采用了最小描述长度算法来对波段进行重要性评价。最小描述长度算法（MDL）是一个信息理论的模型选择原则。MDL 认为最简单简洁的数据描述能够

对数据进行最好、也是最可能的解释。我们在面对成百上千波段的高光谱影像时，总是希望能够通过降维、使用一个比较合理的模型来描述它。MDL 算法把每个波段看作对目标的简单预测模型，通过 MDL 评价分数来对这些简单预测模型进行比较和分级。

MDL 算法最早来源于 Chaitin 的算法信息理论（Chaitin, 1977,1987）,是指用二值符号对样本编码时的最小码长。在传输编码时，当样本比较多时，对样本逐个编码再传输则总编码很长,如果能够知道他的某些内在规律(产生该样本序列的模型)，则不必对所有样本进行编码，只要将模型加以描述并编码，就可以大大压缩要传送的最小码长。因此，为了压缩码长，才用某种模型对其编码进行压缩，然后再保存压缩后的数据，同时为了以后正确恢复这些数据，将所用的模型也保存起来。所以需要保存的总描述长度（比特数）等于这些事例数据进行编码压缩后的长度加上保存模型所需要的数据长度。最小描述长度（MDL）原理就是选择总描述长度最小的模型。

在 MDL 算法里，对于一个高光谱影像像元 X=|x1,x2,…,xn|，xi 和 xj 可能不是相互独立的，数据间存在着内在的规律性。我们借助模型来反映这种规律，设模型是 M,对模型的描述是 DL（M），利用模型对数据 X 的描述为 DL（X|M），则对原数据 X 的描述就转化为：

$$DL（X,M）=DL（M）+DL（X|M）$$ 公式 6.1

用|*|表示描述长度，则总的描述长度为：| DL（X,M）|=| DL（M）+DL（X|M）|。

Rissanen 提出了 MDL 准则来对模型论域 M=（M1,M2,…,Mk）中确定最合适的模型：能够给出最短描述长度的模型 M*是最佳模型，即满足：

$$M^{*} = \arg\min_{Mi \in M}(|DL(Mi)| + |DL(X \mid Mi)|)$$ 公式 6.2

从该公式可以看出，最佳模型实际上是求得了模型复杂性|DL（M）|和模型对数据的描述能力|DL（X|M）|间的最佳平衡。一般来说，模型越复杂，|DL（M）|越大，对数据的描述能力越强，从而|DL（X|M）|越小，反之，模型越简单，|DL（M）|小了，但是|DL（X|M）|会变大。

利用贝叶斯网络来进行数据编码。贝叶斯网络由网络结构和条件概率分布两部分组成。Bs = （V, E） 是有向无环图，即贝叶斯网络结构；V 是节点集，节点 vi 的值域记作 val（vi）；E 是有向边集；每一条边表示节点之间的直接依赖关系，而依赖程度由条件概率决定；每一个节点 vi 都有相应于该节点的父节点集 parent（vi）和条件概率分布表。在具体应用时，贝叶斯网络结构的构造是要根据具体问题的知识原理，将直接相互依赖的变量之间用有向线段连结，形成网络结构；

在具体应用到高光谱遥感影像，对高光谱影像像元描述结构的 MDL 评分函数：

$$Score_{MDL} = L_{model} + L_{data}$$ 公式 6.3

$L_{model}$ 是模型描述长度；$L_{data}$ 是已知模型时恢复原始数据时所需信息的编码。

对于$L_{model}$：贝叶斯网络模型，需要记录每个节点的父结点组合（即网络结构的编码长度）和条件概率标的编码长度。对于一个有 n 个节点的给定网络模型，一个节点的父节点数据为 k，描述着个节点的父节点编码长度为 $k\log_2$（n）。对于一个像元变量 Xi，用|Pai|表示其父节点集合的取值组合数目，用 Si 表示像元变量的取值数目，则其条件概率标所包含的项数为|Pai|Si。根据概率的归一性原理，只需要保存其中的|Pai|（Si-1）项就可以了。对于条件概率表中的每一项，Fireman 通过研究指出需要 $\frac{\log N}{2}$ 个比特数来表示（Fireman,1997），其中 N 是像元数据集中像元的个数。所以，

$$L_{model} = \sum_{i=1}^{n} k_i \log_2(n) + \frac{\log N}{2}(\sum_{i=1}^{n}(S_i - 1 | Pa_i |))$$ 公式 6.4

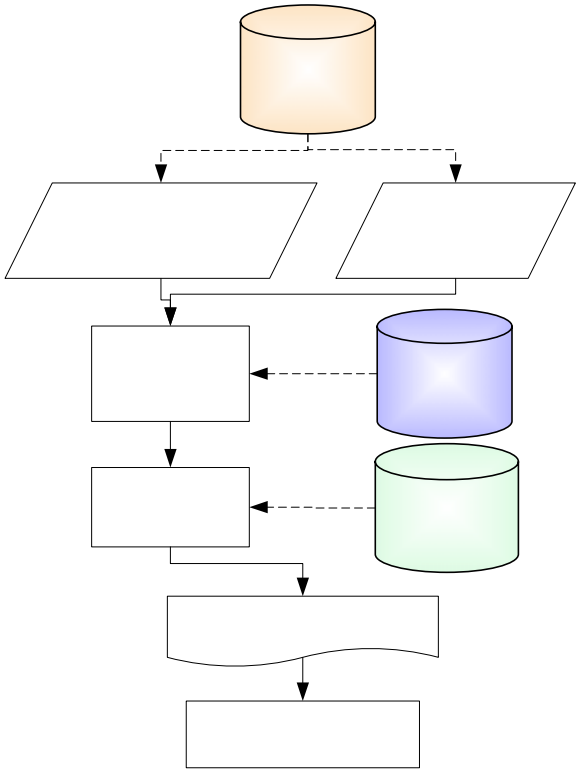
对于$L_{data}$：采用哈夫曼编码，对于每一个像元，根据它的分布（出现频率）来确定编码长度，其概率越大，编码长度越小；概率越小，编码长度越大。于是，像元数据编码压缩后的长度为：

$$L_{data} = N\sum_{i=1}^{n} H(X_i | Pa_i)$$ 公式 6.5

其中，$H(X|Y) = -\sum P(x,y)\log P(x|y)$ 称为变量 X 对于变量 Y 的条件熵，对于完整的像元数据集合，可以通过对像元数据集中像元的个数统计来求出。因此：

$$Score_{MDL} = \sum_{i=1}^{n} k_i \log_2(n) + \frac{\log N}{2}(\sum_{i=1}^{n}(S_i - 1 | Pa_i |)) + N\sum_{i=1}^{n} H(X_i | Pa_i)$$ 公式 6.6

以下通过对 MODIS 的全年 23 个 EVI 数据和对应的土地利用覆盖分类结果（张霞，2006）进行波段选择试验。数据操作流程图如下：



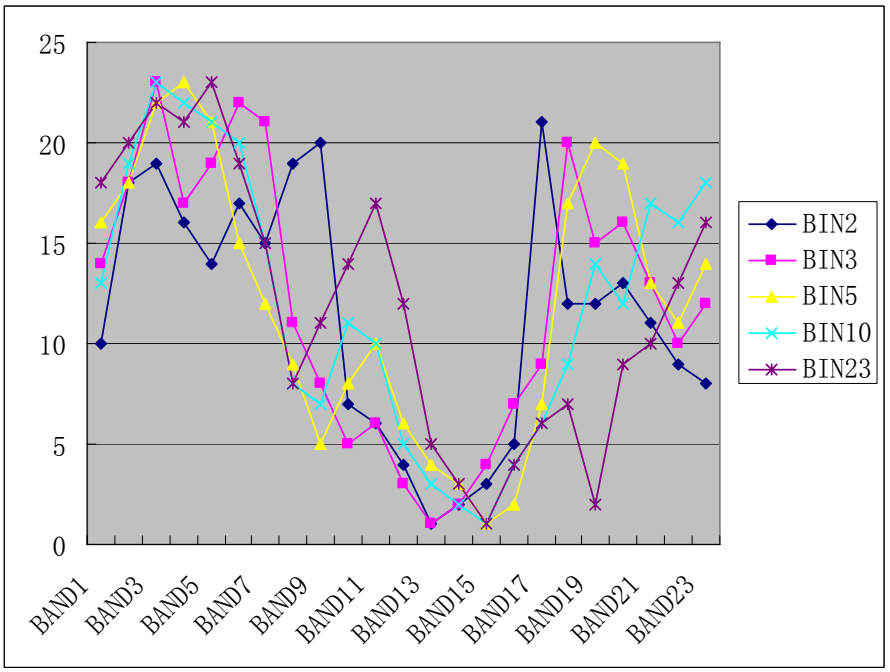
图表 6.2 基于 MDL 的高光谱影像波段选择

首先通过将数据库中遥感影像和分类结果通过 GEORASTER 操作读取 100*100 的空间数据出来，并保存到临时影像数据表 IMGTAB 中。IMGTAB 表的属性由波段、分类结果组成，而 10000 条记录对应着每一个像元。通过数据库查询语言间隔五个像元对图像和分类结果获取样本数据。然后，对数据表的波段数据进行标准化，从而提高模型的执行效率。第三，通过数据挖掘接口建立属性重要性评价模型设置表。第四，在模型表中插入模型参数记录，将 MDL 算法作为属性重要性的应用方法。第五，应用模型，并通过查询语句查询个波段评分结果。最后，在不同的模型参数下（设置不同的参数数量 K）得到不同的排序结果，通过对其累计加和得到波段重要性指数。具体试验结果如表格 6.1。在这里，K=2 表示如果使用两个参数来建立像元与其土地利用类型的模型，则这 23 个波段对于这两个参数的重要性（相对影响程度）分别是多少。通过对不同 K 值下的波段重要性进行排序，如图表 6.3，将其排名进行累计便得到了各波段的重要性综合排名。

表格 6.1 不同 k 值下的波段重要性及综合排名

| MDL 评分 | K=23 | K=2 | K=3 | K=10 | K=5 | 重要性综合排名 |
|---|---|---|---|---|---|---|
| BAND1 | 0.1337425 | 0.0008407 | -0.003265 | 0.1693514 | 0.0502443 | 71 |
| BAND2 | 0.1017451 | -0.004 | -0.004964 | 0.1042737 | 0.0137431 | 93 |
| BAND3 | 0.077929 | -0.004025 | -0.009298 | 0.0628106 | -0.01043 | 109 |

| | | | | | | |
|---|---|---|---|---|---|---|
| BAND4 | 0.0957113 | -0.00347 | -0.004799 | 0.0669374 | -0.012471 | 99 |
| BAND5 | 0.0687211 | -0.003148 | -0.005055 | 0.0698002 | -0.008992 | 98 |
| BAND6 | 0.1081451 | -0.003827 | -0.007157 | 0.0837091 | 0.0787174 | 93 |
| BAND7 | 0.1625354 | -0.003316 | -0.0064 | 0.1541271 | 0.1302953 | 78 |
| BAND8 | 0.2316836 | -0.004025 | -0.000689 | 0.2468098 | 0.1961817 | 55 |
| BAND9 | 0.2116548 | -0.004797 | 0.1557438 | 0.2491559 | 0.2347411 | 51 |
| BAND10 | 0.169084 | 0.0139492 | 0.1972028 | 0.2009726 | 0.1990013 | 45 |
| BAND11 | 0.1408627 | 0.0469628 | 0.1946682 | 0.2068655 | 0.192606 | 49 |
| BAND12 | 0.2110455 | 0.1271332 | 0.2103971 | 0.2790788 | 0.2330375 | 30 |
| BAND13 | 0.2451423 | 0.159522 | 0.2273393 | 0.3088954 | 0.2791826 | 14 |
| BAND14 | 0.260479 | 0.1282568 | 0.2144237 | 0.3168037 | 0.3184284 | 12 |
| BAND15 | 0.2686047 | 0.1276005 | 0.210021 | 0.3316385 | 0.3396825 | 10 |
| BAND16 | 0.2473596 | 0.0823028 | 0.1834897 | 0.3000383 | 0.321462 | 22 |
| BAND17 | 0.2428067 | -0.004981 | 0.0688556 | 0.2558933 | 0.2079666 | 49 |
| BAND18 | 0.2424915 | -0.002426 | -0.005269 | 0.2168989 | 0.048927 | 65 |
| BAND19 | 0.2637906 | -0.002426 | -0.003467 | 0.1678892 | -0.002268 | 63 |
| BAND20 | 0.2223875 | -0.002761 | -0.004639 | 0.1850805 | 0.0040071 | 69 |
| BAND21 | 0.22161 | -0.002326 | -0.002078 | 0.1173321 | 0.1106404 | 64 |
| BAND22 | 0.1719524 | 0.0011648 | 0.0039447 | 0.1486533 | 0.1468949 | 59 |
| BAND23 | 0.1486087 | 0.0023757 | -0.001403 | 0.1087835 | 0.094395 | 68 |

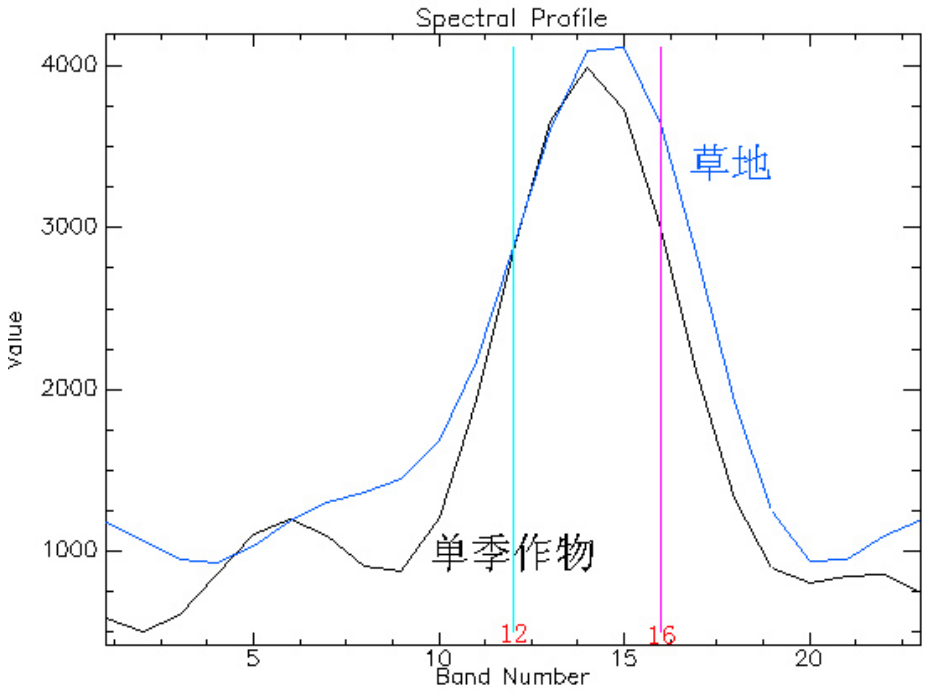



图表 6.3 不同 K 值下的波段重要性排名

从图表 6.3 种可以更加详细地看出该综合排名反映了不同 K 值下的波段重要性分布，可以明显地看出，重要性排名靠前的集中在了 15、14、13、16、12 这五个波段，我们从样本的地物类别可以看出，如表格 6.2，像元主要类别是草地和单季作物，占据了像元总数的 84.36%，当再次查看两种地物的光谱曲线，就可以发现在 12-16 这五个波段中，二者具有很高的相似性。如图 6.4 所示。因此，我们只需要选择这五个波段，甚至排位靠前面的三个波段进行分类或者其他操作，就可以实现把草地和单季作物的组合类型区分开来。

表格 6.2 获取的影像样本数据分类图

| 像元个数 | 类别代码 | 地物类别 |
|---|---|---|
| 409 | 1 | 水体 |
| 8 | 4 | 双季旱作 |
| 1 | 6 | 水旱两作 |
| 4380 | 7 | 草地 |
| 1089 | 8 | 稀疏草地 |
| 57 | 9 | 城市 |
| 4056 | 12 | 单季作物 |




图表 6.5 单季作物与草地的时间谱曲线

MDL 准则不需要计算其参数的先验分布，计算较简单，且由于明确地将结构复杂性作为一个指标而倾向于选择较简单的网络结构，计算结果更容易被接受。计算时间复杂度较低，运行效率比较高。但由于没有用到先验知识，其学习结果的正确性完全依赖于样本数据集合，故要求样本数据数量特别大且不能出现偏差。表 6.3 统计了在准备好样本数据之后对不同 K 值运算生成波段重要性运行时间（样本数为 2000）都在 10 秒以内：

表格 6.2 波段重要性提取程序运行时间

| | K=2 | K=3 | K=5 | K=10 | K=23 |
|---|---|---|---|---|---|
| 运行时间 | 8.578 秒 | 6.656 秒 | 8.891 秒 | 5.734 秒 | 9.016 秒 |

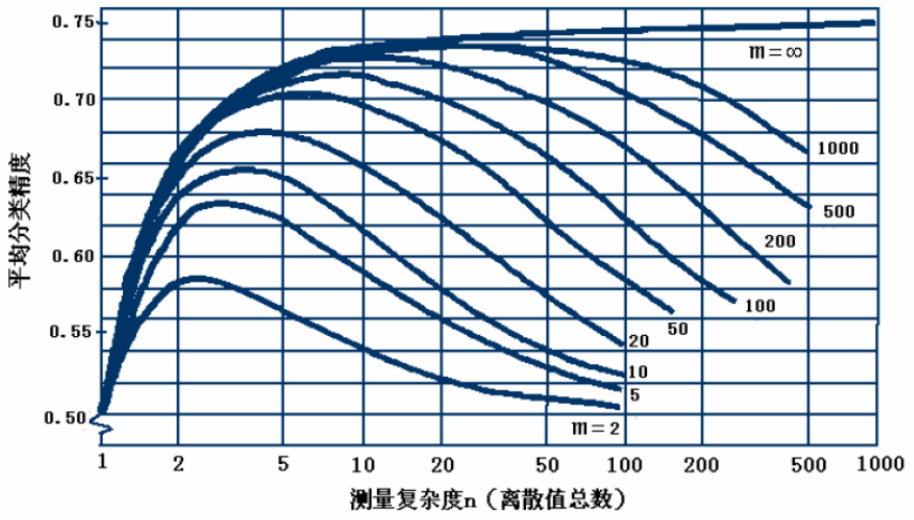
由于在数据库中操作，不仅仅可以对图像数据，也可以对光谱曲线进行查询操作，因此，我们可以保证数据完整性的前提下，通过构建不同类型组合的样本数据表，来实现快速波段选择。

## 6.3 利用非负矩阵分解进行高光谱影像特征提取

对于高光谱遥感数据，当光谱波段数增加时，其特征组合成指数方式增加。假设原始光谱波段数为 N，优选后的光谱波段是 M，N>M，则光谱特征组合的数目为：N！/（N-M）！M!。这个数目是巨大的，并导致运算效率的下降，为此，光谱特征空间的减少和优化显得十分重要（张兵，2003）。为了消除这种现象，各种参量化方法和特征提取手段都被应用到高光谱遥感中。

波段数的减少不一定会导致分类模型或者预测模型的准确性降低。波段数太多，由于噪声的影响，反而会使得模型降低准确性。因此，挖掘出最少的数据维数来表征数据特点，能够大量的提高计算效率，同时也能为建立模型提供更好的支持。

Hughes 在对特征数和分类精度的研究中发现，在样本数一定的情况下，错误率首先随着特征数增加而下降，然后上升，说明分类精度在某个特征数达到极大；而对于特定的特征数，训练样本数增多会提高分类精度（Hughes,1986, 沈清，1991）。如图，




图表 6.6 Hughes 现象

在数据库及数据挖掘领域，在进行决策时，同样需要剔除海量数据表中样本信息存在的冗余成分，提取关键信息。一般来说，从数据库中选取的样本信息具有信息完备、决策属性的模式类型确定、条件属性大多为连续型等特点，但由于每个条件属性重要性不一，且常存在冗余。因此，经常进行属性降维。为了追求最佳属性

降维，人们提出了不少算法。Pawlak 等人提出了先求核（CORE）、然后逐步扩展求出相对属性约简的思想（Pawlak,1994），Jelonek 等人提出了基于属性重要性程度的逐步扩展的算法（Jelonek, 1995）。

由此可以看出，数据挖掘领域中的属性降维与高光谱影像中的降维是具有很强的相似性。而降维的过程，也是特征参量提取的过程。本节拟采用非负矩阵分解方法来实现对影像数据的特征提取。

在科学研究中，讨论利用矩阵分解来解决实际问题的分析方法很多，如 PCA（主成分分析）、ICA（独立成分分析）、SVD（奇异值分解）、VQ（矢量量化）等。在所有这些方法中，原始的大矩阵 V 被近似分解为低秩的 V=WH 形式。这些方法的共同特点是，因子 W 和 H 中的元素可为正或负，即使输入的初始矩阵元素是全正的，传统的秩削减算法也不能保证原始数据的非负性。在数学上，从计算的观点看，分解结果中存在负值是正确的，但负值元素在图像中往往是没有意义的。

《Nature》于 1999 年刊登了两位科学家 D.D.Lee 和 H.S.Seung 对数学中非负矩阵研究的突出成果。该文提出了一种新的矩阵分解思想——非负矩阵分解（Non-negative Matrix Factorization，NMF）算法，即 NMF 是在矩阵中所有元素均为非负数约束条件之下的矩阵分解方法（D.D.Lee, 1999）。NMF 的基本思想可以简单描述为：对于任意给定的一个非负矩阵 A，NMF 算法能够寻找到一个非负矩阵 U 和一个非负矩阵 V，使得满足，从而将一个非负的矩阵分解为左右两个非负矩阵的乘积。由于分解前后的矩阵中仅包含非负的元素，因此，原矩阵 A 中的一列向量可以解释为对左矩阵 U 中所有列向量的加权和，而权重系数为右矩阵 V 中对应列向量中的元素。这种基于基向量组合的表示形式具有很直观的语义解释，它反映了人类思维中“局部构成整体”的概念。研究指出，非负矩阵分解是个最优化问题，可以划为优化问题用迭代方法交替求解 U 和 V。NMF 算法提供了基于简单迭代的求解 U、V 的方法，求解方法具有收敛速度快、左右非负矩阵存储空间小的特点，它能将高维的数据矩阵降维处理，适合处理大规模数据。 （汪鹏，2004）

非负矩阵分解（NMF）问题可描述为：已知非负矩阵 Vn×m{Vij≥0, i=0…n-1, j=0…m-1}，求解两个非负矩阵 Wn×r {Wij≥0, i=0…n-1 j=0…r-1} 和 Hr×m {Wij≥0, i=0…n-1, j=0…r-1}，使得 V≈WH，其中 r 满足（n+m） r<n×m。显然 NMF 是用非负性约束来获取数据表示的一种方法，也即所获取的数据只允许是原始数据的加性组合，而不允许减运算，这一约束导致了基于“部分（part-based representation）”的表示结果。其中，W 被称为基向量，H 为权系数矩阵。V 可以看作是由 m 个 n 维空间中的观测数据向量所组成的矩阵，每一个列向量都是一组观测数据。通过 NMF 方法，V 中每一个列向量均为 W 中 r 个列向量的线性组合，H 中的向量则为相应的权值系数。这样，W 包含的列向量就可认为是对 V 进行线性估计而优化了的基。其结果是用 r 个基表示 m 个原始数据，从而达到了降维的效果。NMF 问题的目标函数有很多种，最常用的两种目标函数为 KL 散度（Kullback-Leibler divergence）和欧几里德距离

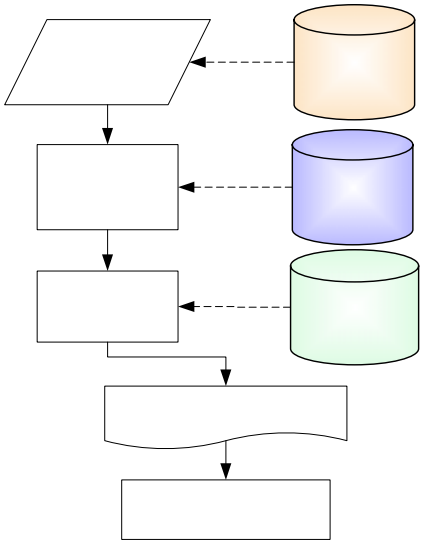
（Euclidean distance）（刘明宇，2006）。

本文采用 KL 散度作为目标函数，KL 散度可以表述为：

$$D(V \| WH) = \sum_{i,j} [Vi_j \log \frac{V_{ij}}{(WH)i_j} - V_{ij} + (WH)_{ij}]$$

公式 6.7

我们通过如下图的流程来实现对高光谱影像特征的提取。



图表 6.7 基于非负矩阵分解的特征提取

首先从图像数据表（记录数为 200*200）中创建样本数据表和应用数据表，间隔十个像元采样把 4000 条数据存储到样本数据表，另外将 10000 条数据存储到应用数据表。这样就产生了一个 4000*23 的非负矩阵作为建模数据，和一个 10000*23 的非负矩阵作为试验数据。初始化 W,H 为任一非负随机矩阵，然后按照公式 6.8 行迭代运算，更新完 W 中的一行以后，立即更新 H 中相应的列。根据公式 6.7 计算 V 和 WH 之间的散度，如果大于预定值，返回继续进行迭代运算，如果小于，则停止运算。目前，假设叠代次数为 50，收敛度为 0.1。

高光谱图像数据表

特征提取模型

$$\begin{cases} W_{ia} \leftarrow W_{ia} \sum_u \frac{V_{iu}}{(WH)_{iv}} H_{au} \\ W_{ia} \leftarrow \frac{W_{ia}}{\sum_j W_{ia}} \\ H_{au} \leftarrow H_{au} \sum_i W_{ia} \frac{V_{iu}}{(WH)_{iu}} \end{cases}$$

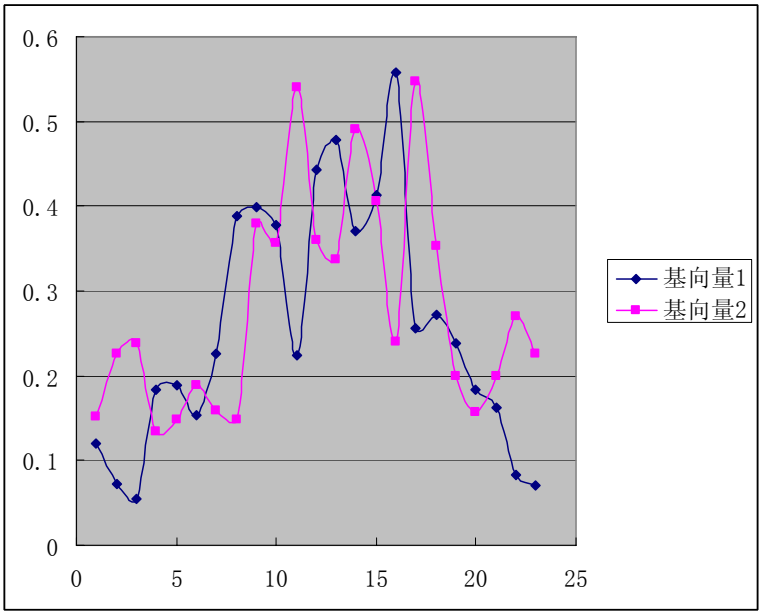
NMF算法

公式 6.8

通过对样本数据进行迭代运算，我们得到 NMF 模型基向量（23*2），如图 6.7 所示，通过这个基向量和试验数据进行相乘，便得到了试验数据的特征向量（2*10000）。而这两个权系数矩阵（2*10000），就是通过非负矩阵分解得到的特征。

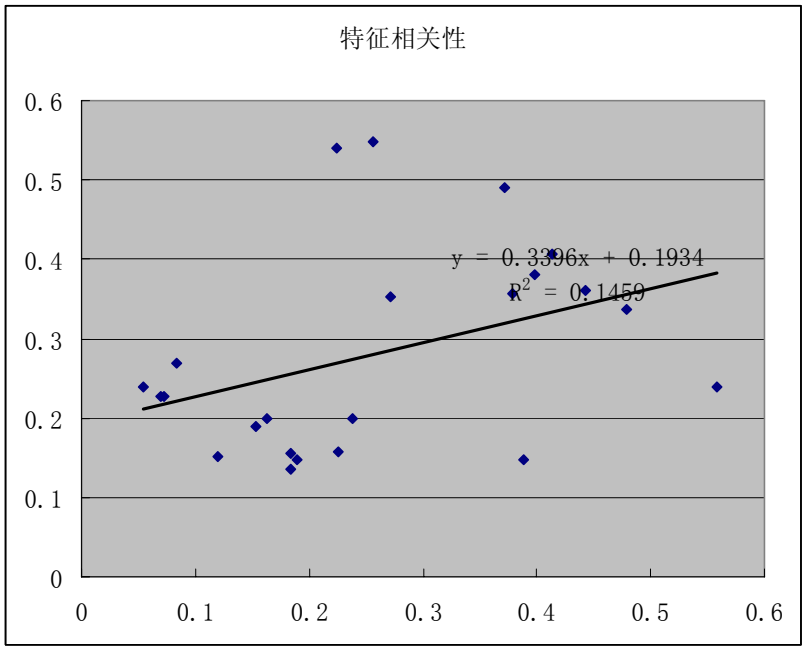
基向量和

特征




图表 6.8 NMF 模型基向量

通过对两个特征基向量进行统计分析发现，而这相关性较小，如图 6.8：两个向量的 R 方差仅为 0.1459，而他们的向量夹角达到了 50.5 度。



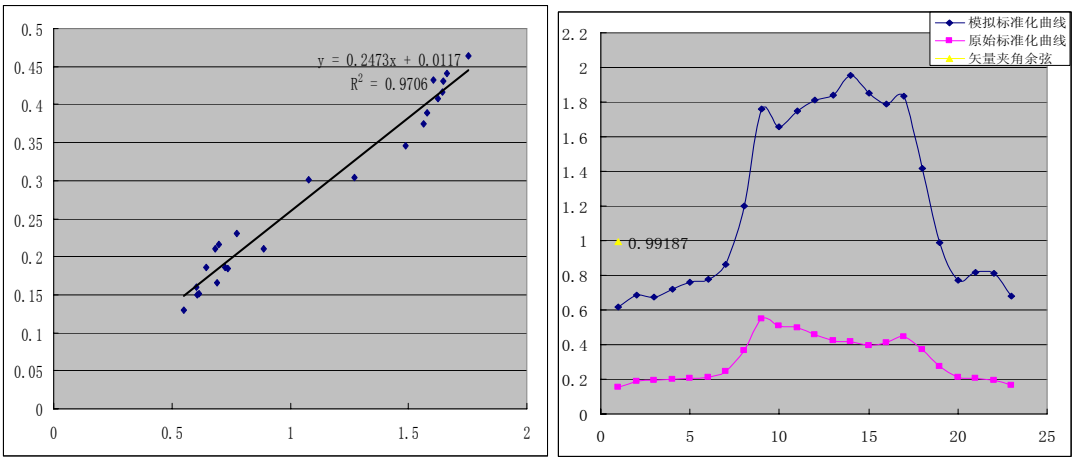

图表 6.9 基向量相关关系分析

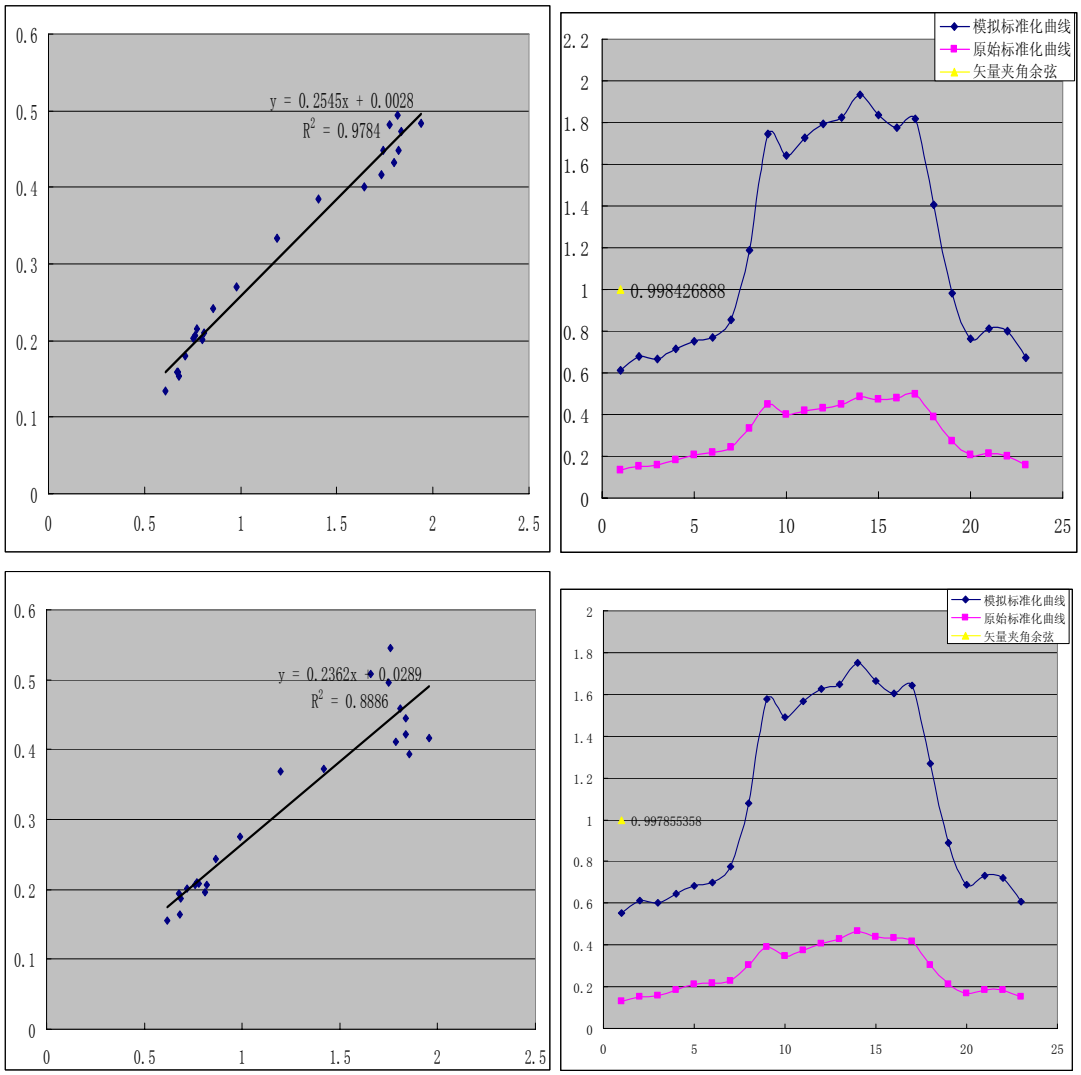
## 6.4 利用非负矩阵分解进行高光谱影像压缩

我们知道，通过非负矩阵变换和样本数据可以得出基向量，而通过基向量和图像数据进行运算，可以得出图像每个像元的权系数。我们把这个过程逆向做一个近似变换，就反推出图像每个像元的原始光谱。

这在高光谱遥感应用里也有实际意义。基向量可以被认为是端元光谱，而权系数矩阵则是该像元中各个端元光谱的百分含量。因此，非负矩阵变换，也可以被理解称为是一种高光谱图像端原提取方法，而非负矩阵逆变换则是通过端元光谱和像元中各个端元百分含量来模拟真实像元光谱的方法。因此，通过 NMF 变换和 NMF 逆变换便能够对高光谱图像进行压缩和解压。实际上，利用端元光谱作为高光谱影像压缩的切入点，在 2001 年被列入了 NASA 的研究项目中，Marsha 博士领导的研究小组提出利用凸矩阵分解（Convex Matrix Factorization）方法实现对高光谱影像端元提取和像元中各端元组分百分含量的获取。他们的研究表明，该方法可以达到 50 倍的压缩比。实际上，利用端元进行数据压缩的方法，其压缩比例主要取决于高光谱波段数和端元数量之比。在利用非负矩阵分解方法进行高光谱影像数据压缩，其压缩倍数同样取决于原始数据波段数和提取的基向量个数，即端元数。

根据这一思想，对 MODIS 的 23 波段影像数据进行了 NMF 变换，实现了把 100*100*23 的高光谱影像数据压缩成为 100*100*2+2*23 的数据。压缩比例为 11 倍。再通过逆变换，即通过两个基向量和对应的权系数矩阵来实现对原始影像的反算，由此得到 MODIS 的 23 波段解压数据。从原始数据和 NMF 逆变换数据中随机抽取了三条数据进行比较，他们的相关系数和光谱夹角如图 6.7 所示。由图可以看出，其中两条光谱曲线 R 方差达到了 0.97 以上，而另一条光谱曲线的 R 方差为 0.89；三对光谱曲线的矢量夹角余弦就达到了 0.99 以上。由此可知，通过 NMF 逆变换进行反算原始图像，是能够达到一定精度要求的。






图表 6.10 标准化真实光谱与 NMF 逆变换标准化光谱比较

然而，由于 NMF 变换是一个最优化过程，并不是一个可逆过程，因此，在通过 NMF 逆变换对高光谱影像进行解压时，必然无法完全保全数据的真实性。虽然说，这个方法在模拟光谱曲线的相似度有很好的效果，但是对于光谱矢量的模数增益，还需要作进一步的研究。

## 6.5 利用支持向量机对高光谱图像进行目标提取

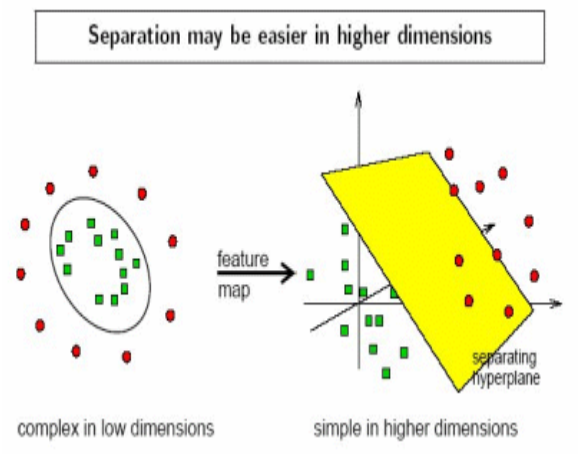
分类作为数据挖掘中一项非常重要的任务,目前在商业上应用最多（比如分析型 CRM 里面的客户分类模型，客户流失模型，客户盈利等等，其本质属于分类问题）。分类的目的是学会一个分类函数或分类模型（也常常称作分类器）,该模型能把数据库中的数据项映射到给定类别中的某一个，从而可以用于预测。目前，分类方法的研究成果较多，判别方法的好坏可以从三个方面进行：1）预测准确度（对非样本数据的判别准确度）；2）计算复杂度（方法实现时对时间和空间的复杂度）；3）模式的简洁度（在同样效果情况下，希望决策树小或规则少）。

近年来，对数据挖掘中分类算法的研究是数据领域中一个热点，对不同分类方

法都有许多对比研究成果。没有一个分类方法在对所有数据集上进行分类学习均是最优的。目前在数据挖掘软件中运用的最早也是最多的分类算法是神经网络，它具有对非线性数据快速建模的能力，通过对训练集的反复学习来调节自身的网络结构和连接权值，并对未知的数据进行分类和预测。但是由于神经网络是基于经验最小化原理，它有如下几个固有的缺陷：1）结构复杂（神经元的结构，还有输入层，隐含层，输出层组合起来的复杂结构）；2）容易陷入局部极小；3） 容易出现过学习问题，也就是训练出来的模型推广能力不强。为了克服传统神经网络的以上缺点，Vapnik 提出了一种新的基于统计学习理论的机器学习算法：支持向量机。该方法是建立在统计学习理论基础上的机器学习方法。通过学习算法，SVM 可以自动寻找出那些对分类有较好区分能力的支持向量，由此构造出的分类器可以最大化类与类的间隔，因而有较好的适应能力和较高的分准率。该方法只需要由各类域的边界样本的类别来决定最后的分类结果。

鉴于支持向量机扎实的理论基础，并且和传统的学习算法想比较（比如人工神经网络），SVM 通过提高数据的维度把非线性分类问题转换成线性分类问题，较好解决了传统算法中训练集误差最小而测试集误差仍较大的问题，算法的效率和精度都比较高。所以近年来该方法成为构造数据挖掘分类器的一项新型技术，在分类和回归模型中得到了很好的应用。但由于支持向量机出现的时间在 90 年代中期，人们对支持向量机的应用主要集中在模式识别方面，对于将支持向量机应用于数据挖掘的研究刚处于起步阶段。

支持向量机实现是通过某种事先选择的非线性映射（核函数）将输入向量映射到一个高维特征空间，在这个空间中构造最优分类超平面。我们使用 SVM 进行数据集分类工作的过程首先是通过预先选定的一些非线性映射将输入空间映射到高维特征空间（如图 6.12），使得在高维属性空间中有可能对训练数据实现超平面的分割，避免了在原输入空间中进行非线性曲面分割计算。SVM 数据集形成的分类函数具有这样的性质：它是一组以支持向量为参数的非线性函数的线性组合，因此分类函数的表达式仅和支持向量的数量有关，而独立于空间的维度。在处理高维输入空间的分类时，这种方法尤其有效。从图 6.12 我们可以直观地了解 SVM 算法：支持向量机算法的目的在于寻找一个超平面 H（d）,该超平面可以将训练集中的数据分开，且与类域边界的沿垂直于该超平面方向的距离最大，故 SVM 法亦被称为最大边缘（maximum margin）算法（陶卿,2000）。




图表 6.11 SVM 算法原理

遥感图象分析与处理是 SVM 应用一个热门的研究方向, 特别是在高光谱遥感分类中, SVM 优越性得到了充分体现（杜培军，2006）。Fabio Roli 等使用线性核、多项式核和 RBF 核, 选择不同参数进行试验, 平均分类精度在 89.4%～91.5%之间, 试验表明 SVM 的分类精度优于 k-NN、MLP 等（FABIO R,）。骆剑承等将 SVM 应用于遥感影像空间特征提取与分类, 说明 SVM 不但能获得比较高的分类精度, 而且在学习速度、自适应能力、特征空间维数不限制、可表达性等方面具有优势（骆剑承，2002）。J. A. Gualtieri 等对同一区域分别按照 4 类和 16 类问题用 SVM 进行高光谱分类, 分别获得了 96%和 87%的分类精度, 发现 SVM 用于高光谱分类的最大优点就是能够直接对高维数据进行处理,而不必经过降维处理, 从而保证了光谱信息的充分应用（Gualtieri,1998）。G. Camps- Valls 等采用 SVM 对高光谱图象进行作物分类, 证明 SVM 的分类效果优于神经网络, 而且在高维输入时神经网络无法训练但 SVM 可以正常使用, 并认为 SVM 在高维时可以避免噪声波段的影响（G. Camps）。C. A. Shah 等试验发现 SVM 用于分类不需要降维, 分类精度达到 97%, 不受 Hughes 现象的影响, 能够很好地应用于高光谱分类（Shah）。刘志刚等运用若干个核函数之和代替 SVM 中的核函数时,对不同物理意义的特征向量选择不同的核函数, 并应用于遥感影像土地利用分类中,取得了较好的效果（刘志刚，2003）。王凯峰等利用单类 SVM 对遥感影像进行目标探测，实验表明单类 SVM 在牺牲少量泛化性的同时能有效地降低误栓率，并提高检测速度（王凯峰，2005）。

本节把支持向量机方法引入到高光谱数据库中，对高光谱影像数据表进行基于支持向量机的数据挖掘分析。

图像像元样本数为 400，目标样本数为 22，目标所占比例为 5.5%，类别分布如表 6.4 所示；测试数据像元数为 10000，目标像元为 219 个，目标所占比例为 2.19%；样本所占数据比例为 4%。本次数据挖掘实验采用线性函数作为 SVM 的核函数。

表格 6.3 样本数据表目标类别分布

| 数据类别编码 | 1 | 7 | 8 | 12 | 13 |
|---|---|---|---|---|---|
| 目标数量 | 22 | 231 | 32 | 114 | 1 |

表格 6.4 基于 SVM 的目标提取结果

| 概率 | 识别数量 | 命中 | 遗漏 | 错分 | 命中率 |
|---|---|---|---|---|---|
| 0.05 | 6326 | 219 | 0 | 6107 | 100 |
| 0.1 | 2367 | 218 | 1 | 2149 | 99.543379 |
| 0.15 | 743 | 218 | 1 | 525 | 99.543379 |
| 0.2 | 406 | 218 | 1 | 188 | 99.543379 |
| 0.25 | 296 | 217 | 2 | 79 | 99.08675799 |
| 0.3 | 255 | 217 | 2 | 38 | 99.08675799 |
| 0.35 | 236 | 216 | 3 | 20 | 98.63013699 |
| 0.4 | 219 | 214 | 5 | 5 | 97.71689498 |
| 0.45 | 212 | 211 | 8 | 1 | 96.34703196 |
| 0.5 | 207 | 206 | 13 | 1 | 94.06392694 |
| 0.55 | 204 | 203 | 16 | 1 | 92.69406393 |
| 0.6 | 197 | 196 | 23 | 1 | 89.49771689 |
| 0.65 | 186 | 186 | 33 | 0 | 84.93150685 |
| 0.7 | 178 | 178 | 41 | 0 | 81.27853881 |
| 0.75 | 173 | 173 | 46 | 0 | 78.99543379 |
| 0.8 | 150 | 150 | 69 | 0 | 68.49315068 |
| 0.85 | 123 | 123 | 96 | 0 | 56.16438356 |
| 0.9 | 79 | 79 | 140 | 0 | 36.07305936 |
| 0.95 | 20 | 20 | 199 | 0 | 9.132420091 |

从分析结果可以看出，在以“是”的概率为 50%在作为“是”与“不是”的决策点时，219 个目标像元识别出了正确目标 206 个，识别率为 94.07%；错分目标 1 个，错分率为 0.46%；遗漏目标 13 个，漏分率为 5.93%。

由于支持向量机方法是建立在统计学习理论的 VC 维理论和结构风险最小原理基础上的，根据有限的样本信息在模型的复杂性（即对特定训练样本的学习精度）和学习能力（即无错误地识别任意样本的能力）之间寻求最佳折衷，以期获得最好的推广能力。支持向量机方法可以解决小样本情况下的机器学习问题，能够直接处理高维数据，可以利用不同的核函数解决非线性问题。

## 6.7 本章小结

本章利用数据挖掘方法对高光谱遥感影像进行了数据挖掘研究。本章仅仅是数据挖掘在高光谱领域应用的冰山一角。

# 第七章 结论与展望

本文针对高光谱数据库这一研究对象，提出了光谱数据模型与高光谱影像数据模型，并面对应用向导和数据挖掘，对高光谱数据库设计和建设进行了研究和实施，初步建立了集光谱数据和影像数据为一体的、支持数据挖掘的高光谱数据库原型系统。围绕高光谱数据挖掘这一主线，在数据存储、数据升迁、数据筛选、高光谱方法、数据挖掘模型等各方面进行了研究，把最小描述长度、非负矩阵分解、支持向量机等数据挖掘方法应用到影像数据挖掘中，并在 ORACLE 数据库平台上进行了部分实现。通过研究证明，通过对高光谱影像的转换，在数据库中将数据挖掘技术应用到高光谱遥感影像分析中，从另一个角度来分析遥感影像的光谱空间信息，能够很好的实现对光谱维信息的认识。

## 7.1 论文的特色与创新点

本文以数据库为背景，在进行项目总结和课题研究的同时，注意归纳总结，参考最新科研进展，在统一地面测量光谱数据和高光谱影像数据基础之上，构建高光谱数据库，并把数据库领域的数据挖掘技术应用到高光谱遥感影像分析中。本文主要的特色有以下几个方面：

- 本文总结了前人的工作成果，结合自己的研究实践，在构建高光谱数据库基础之上，提出了针对数据库存储方案的光谱数据模型和影像数据模型，通过将必要的高光谱方法和高光谱模型整合到数据库之中，实现了数据、方法、模型在高光谱数据库中的集成与统一。
- 本文将多源数据统一到一个数据库平台之中，将原有的大表结构和表群结构作转换，设计了以二元数据核心的星型数据库概念结构，并总结了通用的高光谱数据库方法和模型设计线路。
- 本文将数据库技术应用在光谱数据模拟和参量关系研究上，从属性和光谱两个方向阐述了光谱数据挖掘的应用。同时，利用数据库的自动分析功能，实现了对波段组合与波段选择的自动优化。
- 本文将高光谱影像数据空间映射到数据库表空间，从而实现借助于数据库平台对高光谱影像进行数据分析与信息挖掘。通过最小长度模型辅助高光谱影像波段选择；通过非负矩阵分解实现高光谱影像特征提取；通过非负矩阵分解的逆变换对高光谱数据压缩；通过支持向量机对高光谱图像进行目标提取。

## 7.2 高光谱数据库的发展和数据挖掘展望

本文主要对高光谱数据库光谱与影像数据存储模型、数据库概念结构设计及数据库建设、高光谱数据库方法和模型设计与开发作了系统研究，并对光谱数据挖掘和影像数据挖掘作了初步的探讨。本文主要研究重点侧重在数据库端，集中在数据

挖掘方向的应用开拓。高光谱数据库的发展，可以沿两个方向进行：横向延伸和纵向拓展。

所谓横向延伸，指的是侧重于数据和应用的广度。由于高光谱数据库设计之初的思想，便是针对网络应用而展开，因此，在数据库模型与方法设计时，对网络应用作了考虑。数据库端可以生成影像金字塔以满足将来的大空间多尺度浏览，同时数据库的模型和方法都以存储过程内嵌在数据库中，这对于客户端的平台等几乎没有什么限制，可以轻松的实现跨平台、跨操作系统。鉴于时间和研究重点地考虑，本文研究没有在前台的展示做深入的工作，但是，基于 JAVA 和 JSP 的开发可以轻易的将数据库后台的分析成果作展现。另外，在现有的数据结构下，动态的影像浏览也能够实现。因此，高光谱数据库横向延伸，可以拓宽高光谱数据库的可应用性和灵活性。

所谓纵向拓展，指的是侧重于数据和应用的深度。这也是本文主要向阐述的方向。将数据挖掘技术应用于高光谱，这为高光谱遥感科学研究开辟了一个新的道路。我们可以从另一个角度来看光谱曲线和高光谱影像，通过数据库中成熟的方法来对高光谱影像进行分析研究。本文在第六章提到的一些方法仅仅是数据挖掘算法中的沧海一粟。还有诸如“正交分区聚类/O-Clustering”， “高维数据自动子空间聚类/CLIQUE”，“ 基于自回归滑动平均的序列匹配方法/ARMA”等等，这些算法和研究思想都可以通过数据库和数据仓库平台应用到遥感影像的数据分析和信息挖掘之中。

随着高光谱数据日益增多，数据积累和应用逐步拓展，高光谱数据库及数据挖掘应用必将成为高光谱遥感应用与研究的重要手段。

# 参考文献

[1]. Adams J B, Smith M O, and Johnson P E. Spectral mixture modeling: A new analysis of rock and soil types at the Viking Lander I site[J]. J.Geophys. Res., 1986, 91, 8098-8112.

[2]. Arce G R, Foster R E. Detail-preserving ranked-order based filters for image processing. IEEE Trans, Acoust, Speech, Signal Processing, 1989, ASSP-37: 83～98

[3]. Aspinall R, Pearson D. Integrated Geographical Assessment of Environmental Condition in Water Catchments: Linking Landscape Ecology, Environmental Modeling and GIS. Journal of Environmental Management, 2000, Vol.59, PP. 299-319

[4]. Barnard, A., Zaneveld, J., Pegau, W., 1999. In situ determination of the remotely sensedreflectance andthe absorption coefficient: closure andinversion. AppliedOptics 38 （24）, 5108–5117.

[5]. Boardman, J.W.,1993,Autamating Spectral unming of AVIRIS Data Using Convex Geometry Concepts,4th Annual JPL Airborne Geo Science Workshop,JPL.Pub.93-26,1:11-14.

[6]. Boardman, Joseph W., 1995, “Analysis, understanding and visualization of hyperspectral data as convex sets in n-space”, SPIE Proceedings, Vol. 2480. pp. 14-22.

[7]. C. A. Bateson and B. Curtiss, “A tool for manual endmember selection and spectral unmixing,” in Summaries of the V JPL Airborne Earth Science Workshop, Pasadena, CA, 1993.

[8]. CAMPS - VALLS1 G, GOMEZ - CHOVA1L, CALPE- MARAVILLA1 J , et al. Support Vector Machines for Crop Classification Using Hyperspectral Data . http://citeseer.nj.nec.com

[9]. Canadian Space Agency, Hyperspectral Remote Sensing Applications, http://www.space.gc.ca

[10]. Ceccato P., Flasse S., Tarantola S., et al., 2001. Detecting vegetation leaf water content using reflectance in the optical domain, Remote Sens. Environ., 77: 22-33.

[11]. Chaitin G J. Algorithmic Information Theory, Combridge University Press, 1987

[12]. Chaitin G J. Algorithmic Information Theory, IBM J Res. Develop,1977, 21（4）, PP. 350-359

[13]. Chevrel S., Belocky R., Gr el K. - （2002） - Monitoring and assessing the environmental impact of mining in Europe using advanced Earth Observation techniques - MINEO, First results of hte Alpine test site. Environmental Communication in the Information Society, EnviroInfo Vinee 2002, W. Phillmann and K. Tochtermann Eds, part1, PP. 518- 526

[14]. CHEVREL S., KUOSMANNEN V., BELOCKY R., MARSH S., TAPANI T., MOLLAT H., QUENTAL L., VOSEN P., SCHUMACHER V., KURONEN E., AASTRUP P, （2001） - Hyperspectral airborne imagery for mapping

mining-related contaminated areas in various European environments - First results of the MINEO Project .5th International Airborne Remote Sensing Conference, San Francisco, California, 17-20 September 2001 （in press）

[15]. Clark, R., 1990. High spectral resolution reflectance spectroscopy of minerals. Journal of Geophysical Research 95 （B8）, 12 653–12 680.

[16]. Clark, R.N. and Swayze, G.A., Mapping Minerals, Amorphous Materials, Environmental Materials, Vegetation, Water, Ice and Snow, and Other Materials: The USGS Tricorder Algorithm. Summaries of the Fifth Annual JPL Airborne Earth Science Workshop, January 23- 26, R.O. Green, Ed., JPL Publication 95-1, p. 39-40, 1995.

[17]. Clark, R.N., A.J. Gallagher, and G.A. Swayze, Material Absorption Band Depth Mapping of Imaging Spectrometer Data Using a Complete Band Shape Least-Squares Fit with Library Reference Spectra, Proceedings of the Second Airborne Visible/Infrared Imaging Spectrometer （AVIRIS） Workshop. JPL Publication 90-54, 176-186, 1990.

[18]. Clark, R.N., G.A. Swayze, A. Gallagher, N. Gorelick, and F. Kruse, Mapping with Imaging Spectrometer Data Using the Complete Band Shape Least-Squares Algorithm Simultaneously Fit to Multiple Spectral Features from Multiple Materials, Proceedings of the Third Airborne Visible/Infrared Imaging Spectrometer （AVIRIS） Workshop, JPL Publication 91-28, 2-3, 1991.

[19]. Clementini E, Felice P D, Koperski K. Mining Multiple-level Spatial Association Rules for Objects with a Broad Boundary. Data and Knowledge Engineering, 2000, 34, PP. 251-270

[20]. CODASYL Development Committee, "An Information Algebra," Communications of the ACM, V, 4, April 1962, p. 190-204. （The information algebra, mainly the work of R. Bozak, is a calculus language for describing data manipulation. Abstract, concise. Mathematical.）

[21]. CODASYL, "Report of the DATA BASE TASK GROUP," April 1969 and October 1971, available from ACM. （Language specifications for a data description language and COBOL enhancements for data manipulation. Based on a network organization of information.）

[22]. Codd, E.F.,A relational model of data for large shared data banks, Communication of the ACM, Vol.13, No.6, June 1970, PP. 377-387

[23]. Collins,W. E. and S. H., Chang （1982） Application of geophysical environmental research （GER） airborne scanner data for detection of hydrothermal alteration in Nevada. In: Proceedings of the sixth thematic conference on remote sensing for exploration geology, ERIM , may 16-19 , Houston , Texas , USA.

[24]. Coyle E J, Lin J H. Stack filters and the mean absolute error criterion. IEEE Trans, Acoust, Speech, Signal Processing, 1988, 36（8）: 1244～1245

[25]. Crosta A P, Souza C R de F.Evaluating AVRIRIS hyperspectral remote sensing data for geological mapping in Laterized Terranes[C], Central Brazil, Proceedings of the Twelfth International Conference and Workshops on

Applied Geologic Remote Sensing, Denver, Colorado. 17-19 November 1997, Volume, 1997,II:II-430~II-437

[26]. Curran P.J., 1989. Remote Sensing of foliar chemistry. Remote Sens. Environ. 30: 271-278.

[27]. Curran P.J., Dungan J.L., Peterson D.L., 2001. Estimating the foliar biochemical concentration of leaves with reflectance spectrometry: testing the Kokaly and Clark methodologies. Remote Sens. Environ., 76: 349-359.

[28]. D.D.Lee, H.S.Seung, Learning the Parts of Objects by Non-Negative Matrix Factorization, Nature, 401, PP. 788-791

[29]. Declan Butler, Virtual globes: The web-wide world, Nature 439, PP. 776-778, 16 February 2006.

[30]. Edwin. Finding Boundary Shape Matching Relationships in Spatial Data. In: Scholl M, Voisard[e]ds. Proceedings of the 5th International Symposium on Spatial Databases（SSD'97）. Berlin: Springer-Verlag, 1997

[31]. Eklund P W, Kirby S D, Salim A. Data Mining and Soil Salinity Analysis. International Journal of Geographical Information Science, 1998, Vol. 12 Col. 3, PP. 247-268

[32]. Elvidge, C.D., Chen, Z.K., and Groeneveld, D.P.（1993）, Detection of trace quantities of green vegetation 1990 AVIRIS data., Remote Sensing of Environ., 44（2-3）, PP. 271-279.

[33]. Emmett Ientilucci, Hyperspectral Image Classification Using Orthogonal Subspace Projections: Image Simulation and Noise Analysis. http://www.cis.rit.edu/~ejipci/Reports/osp paper.pdf

[34]. Ester M, Kriegel H P, Xu X. Knowledge Discovery in Large Spatial Databases: Focusing Techniques for Efficient Class Identification. In: Egenhofer M J, Herring J R, Portland M E, [e]ds. Proceedings of the 4th International Symposium on Spatial Databases（SSD'95）. Berlin: Springer-Verlag, 1995

[35]. Ester M. A Database Interface for Clustering in Large Spatial Databases. The 1st International Conference on Knowledge Discovery and Data Mining, Montreal, 1995

[36]. Ester M. A Density-based Algorithm for Discovering Cluters in Large Spatial D[at]abases with Noise. The 2nd International Conference on Knowledge Discovery and Data Mining, Portland, 1996

[37]. F. A. Kruse, J. W. Boardman, J. F. Huntington, Evaluation and Geologic Validation of EO-1 Hyperion, 2003

[38]. FABIO R, GIORGIO F. Support Vector Machines for Remote - Sensing Image Classification.

[39]. Foudan Salem, Menas Kafatos, Tarek El-Ghazawi, Richard Gomez, and Ruixin Yang,2002, HYPERSPECTRAL IMAGE ANALYSIS FOR OIL SPILL DETECTION, JPL Proceedings 2002

[40]. Franklin, SE, McDermid, GJ, 1993. Empirical relations between digital SPOT HRV and casi spectral response and lodgepole pine forest stand parameters. International Journal of Remote Sensing 14 （12）, 2331-2348

[41]. Frieman N．Learning Bayesian．Networks in the presence of missing values and hidden variables[M]．ICML，1997．

[42]. Fu Yingying, Liu Su-hong, Yu Sheng-quan, Tian Zhen-kun，Standard Format Design and Input Realization of Measured Spectral Data in The Spectral Knowledge Base, Geoscience and Remote Sensing Symposium, 2004. IGARSS '04. Proceedings. 2004 IEEE International， Volume 7, 2004 Page（s）:4444 - 4447 vol.7

[43]. Gillespie,A.,R. et al., 1990, Interpretation of residual images: Spectral mixture analysis of AVIRIS images., Owens Valley, California, in Proc. 2nd Airborne Visible/Infrared Imaging Spectrometer （AVIRIS） Workshop,（R.O.Green Ed.）,JPL Publ. 90-54,JPL Laboratory,Pasadena,CA,PP.243-270.

[44]. Goetz, A F H; Rowan, L C; Kingston, M J. Mineral identification from orbit - Initial results from the Shuttle multispectral infrared radiometer. Science. Vol. 218, PP. 1020-1024. 3 Dec. 1982

[45]. Goetz, A.F.H. & Srivastava,V.,1985, “Mineralogical Mapping in The Cuprite mining district, Nevada.”, In: G. Vane & A. F. H. Geotz（Eds.）, Proc. of the Airborne Imaging Spectrometer（AIS） Data analysis Worksop, Pasadena, USA,NASA-JPL Publ. 85-41:pp.22-31.

[46]. Goetz, Alexander F. H.; Rowan, Lawrence C. Geologic Remote Sensing. Science, Volume 211, Issue 4484, PP. 781-791, Feb.1981

[47]. Goetz, Alexander F. H.; Vane, Gregg; Solomon, Jerry E.; Rock, Barrett N. Imaging Spectrometry for Earth Remote Sensing. Science, Volume 228, Issue 4704, PP. 1147-1153. Jun.1985

[48]. Goodenough, D. G., A. Dyk, K. O. Niemann, J. Pearlman, H. Chen, T. Han, M. Murdoch, and C. West 2003, "Processing HYPERION and ALI for Forest Classification," IEEE Transactions on Geoscience and Remote Sensing, Vol. 41, No. 6, PP. 1321-1331.

[49]. Google Earth 软件终极教程之版本差异篇， 2005-9-21，http://www.godeyes.cn/news/2005/9/20/0929233853.htm

[50]. GUALTIERI J A, CROMP R F. Support vector machines for hyperspectral remote sensing classification. Proceedings of the SPIE, 27th AIPR Workshop, 1998,221- 232.

[51]. http://citeseer.nj.nec.com

[52]. http://worldwind.arc.nasa.gov

[53]. http://www.godeyes.cn/news/2005/11/3/1103153803.htm

[54]. Hughes, G. F.,1968, “On the mean accuracy of statistical pattern recognizers”, IEEE Trans.

[55]. Hunt G R．Electromagnetic radiation: the communication link in remote sensing [A]. In: B. Siegal and A.Gilleapie （Eds）, Remote Sensing in Geology[C], New York, Wiley, 1980.702.

[56]. IBM, "Information Management System IMS/360, Application Description Manual" H20-0524-1. IBM Corp., White Plains, N.Y., July 1968

[57]. Information Theory, Vol. IT-14, 55-63, 1968.

[58]. Inmon, William H. Building the Data Warehouse （2nd Ed.）. Wiley. New York. 1996

[59]. J. Bowles, P. J. Palmadesso, J. A. Antoniades, M. M. Baumback, and L.J. Rickard, “Use of filter vectors in hyperspectral data analysis,” Proc.SPIE Infrared Spaceborne Remote Sensing III, pp. 148–157, 1995.

[60]. J. Theiler, D. Lavenier, N. Harvey, S. Perkins, and J. Szymanski, “Using blocks of skewers for faster computation of pixel purity index,” in SPIE Int. Conf. Optical Science and Technology, San Diego, CA, 2000.

[61]. Jagatheesan A, Moore R, Paton NW, Watson P. Grid data management systems & services. In: Freytag JC, Lockemann PC,Abiteboul S, Carey MJ, Selinger PG, Heuer A, eds. Proc. of the 29th Int’l Conf. on Very Large Data Bases （VLDB）. Berlin:Morgan Kaufmann, 2003. 1150.

[62]. Jelonek J, Krawiec K, Slowinski R. Rough set reduction of attributes and their domains for neural networks, International Journal of Computational Intelligence, 1995,11（2）, PP. 339-347

[63]. Jiawei Han, Micheline Kamber, Data Mining: Concepts and Techniques, Academic Press, Morgan Kaufmann Publishers, 2001

[64]. Jim Farley, Jeffrey Xie, Oracle Database 10g, Managing Geographic Raster Data Using GeoRaster, An Oracle Technical White Paper. 2003, Oracle Corporation

[65]. Jim Gray, The revolution in database architecture. In: Weikum G, König AC, Deßloch S, eds. Proc. of the ACM SIGMOD Int’l Conf.on Management of Data. ACM Press, 2004. pp1-4.

[66]. Jimenez, L.O., Morales-Morell, A., Creus, A. Classification of hyper-dimensional data based on feature and decision fusion approaches using projection pursuit, majority voting, and neural networks[J]. IEEE Trans. On Geosci. and remote sensing. 1999, 37（3）: 1360-1366.

[67]. Johnson, L.F., Billow C.R., （1996）,“Spectrometric estimation of total nitrogen concentration in Douglas-fir foliage”, Int. J. Remote Sensing,Vol.17, No.3, 489-500.

[68]. K. Staenz, T. Szeredi, and J. Schwarz, “ISDAS—A system for processing/analyzing hyperspectral data,” Can. J. Remote Sens., vol. 24, pp. 99–113, 1998.

[69]. Kariuki P C．Analysis of the effectiveness of spectrometry in detecting the swelling clay minerals in soils．M ．S．thesis，International Institute for Areospa ce Survey and Earth Sc iences，Enschede．The Netherlands．1999．44-67．

[70]. Knorr E M, Ng R T. Finding Aggregate Proximity Relationships and Commonalities in Spatial Data Mining. IEEE Transactions on Knowledge and Data Engineering, 1996, 8（6）, PP. 884-897

[71]. Ko S J, Lee Y H. Center weighted median filters and their applications to image enhancement. IEEE Trans, Circuits Syst, 1991, 38（9）: 984～993

[72]. Koperski K, Adhikary J, Han J. Spatial Data Mining: Process and Challenges Survey Paper. SIGMOD’96 Workshop on Research Issues on Data Mining and Knowledge Discovery （DMKD’96）, Montreal, Canadam, 1996

[73]. Koperski K, Han J. Discovery of Spatial Association Rules in Geographic Information Database. In: Egenhofer M J, Herring J R, Portland M E, eds. Proceedings of the 4th International Symposium on Spatial Database（SSD'95）. Berlin: Springer-Verlag, 1995.

[74]. Kruse, F. A., Careen-Young, K. S., and Boardman, 1990, J. W., Mineral mapping at Cuprite, Nevada with a 63 channel imaging spectrometer, Photogram. Eng. Remote Sens, 56, PP. 83-92,.

[75]. Leˆnio Soares Galva˜o, Antoˆnio Roberto Formaggio, Daniela Arnold Tisot.Discrimination of sugarcane varieties in Southeastern Brazil with EO-1 Hyperion data, Remote Sensing of Environment ,2005, 94: 523–534

[76]. Lee Y H, Kassam S A. Generalized median filtering and related nonlinear filtering techniques. IEEE Trans, Acoust, Speech, Signal Processing, 1985, ASSP-33: 672～683

[77]. Levene M, Vincent M W. Justification for Inclusion Dependency Normal Form. IEEE Transactions on Knowledge and Data Engineering, 2000, 12（2），PP. 281-291

[78]. Li D R, Chen T., KDG—Knowledge Discovery from GIS．The Canadian Conference on GIS, Ottwa, Canada, 1994.1001-1012

[79]. Li, X., Gao, F., Wang, J., Strahler, A., 2001. A priori knowledge accumulation and its application to linear BRDF model inversion. Journal of Geophysical Research 106 （11）, 11 925–11 935.

[80]. Lin X, Zhou X, Liu C. Efficiently Matching Proximity Relationships in Spatial Databases. In: Guting R H, Papadias D, Lochovshy F, eds. Proceedings of the 6th International Symposium on Spatial Databases（SSD'99）. Berlin: Springer-Verlag, 1999

[81]. M. E. Winter, "N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," in Proc. SPIE Imaging Spectrometry V, 1999, pp. 266–275.

[82]. M. Lennon, G. Mercier, MC Mouchot, L. Hubert-Moy, "Spectral unmixing of hyperspectral images with the Independent Component Analysis and wavelet packets" IGARSS 2001 Conference, Sydney, Australia, 09-13 july 2001

[83]. Martha J. Fox, http://www.spacepda.net/abstracts/01/sbir_html/012158.html, 2001

[84]. Martin, M. E., S.D. Newman, J. D. Aber, and R. G. Congalton,1998,Determining forest species composition using high resolution spectral resolution remote sensing data, Rem. Sens. Environ. 65: 249-254.

[85]. Michael Stonebraker, Eugene Wong, Peter Kreps, Gerald Held, The Design and Implementation of INGRES, ACM Transactions on Database Systems, Vol.1, No.3, September 1976, PP. 189-222

[86]. Michio S., Tsuyoshi A., 1989. Seasonal visible, near-infrared and mid-infrared spectra of rice canopies in relation to LAI and above-ground dry phytomass. Remote Sens. Environ., 27: 119-127.

[87]. MM ASTRAHAN, MW BLASGEN, DD CHAMBERLIN, KP ESWARAN, JN GRAY, PP. GRIFFITHS, WF KING, and etc. System R: Relational

Approach to Database Management, ACM Transactions on Database Systems, Vol. 1, No.2, June 1976, PP. 97-137

[88]. Mouzon O D, Dubois D, Prade H. Using Consistency and Abduction Based Indices in Possibilistic Causal Diagnosis. IEEE, 2001, PP. 729-734

[89]. Murray A.T, Estivill castrov, Clustering Discovery Techniques for Exploratory Spatial Data Analysis . International Journal of Geographical Information Sciences,Vol.12, No.5, 1998, PP. 431-443.

[90]. Murry A.T., Estivill-castro V. Clustering Discovery Techniques for Exploratory Spatial Data Analysis. International Journal of Geographical Information Science, 1998,12（5）, PP. 431-443

[91]. Pawlak Z. Slowinski R. Rough set approach to multi-attribute decision analysis, European Journal of Operational Research . 1994.72（3）, PP. 443-459

[92]. R. N. Clark, G. A. Swayze, R. Wise, K. E. Livo, T. M. Hoefen, R. F. Kokaly, and S. J. Sutley, USGS Digital Spectral Library splib05a, USGS Open File Report 03-395, 2003.

[93]. Raymond F.K., Roger N.C., 1999. Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. Remote Sens. Environ., 67:267-287.

[94]. Reinartz T. Focusing Solutions for Data Mining: Analytical Studies and Experimental Results in Real World Domains. Berlin:Springer, 1999

[95]. Rissanen J. A Universal Prior for Integers and Estimation by Minimum Description Length. The Annuls of Statistics, 1983, 11（2）, PP. 416-431

[96]. Roberts D.A.,Yamaguchi.Y., and Lyon,R.J.P.,1985, Calibration of airborne imaging spectrometer data to percent Reflectance using field spectral measurements. Proceedings of nineteenth international symposium on remote sensing of environment.

[97]. SHAH C A, WATANACHATURAPORN P.VARSHNEY P K, et al. Some Recent Results on Hyperspectral Image Classification. http://citeseer.nj.nec.com.

[98]. Shimabukuro, Y.E. and J.A. Smith, 1991. The least-squares mixing models to generate fraction images derived from remote sensing multispectral data. IEEE Transaction on Geoscience and Remote Sensing, vol. 29（1）, pp. 16-20.

[99]. Smith,M.O., et al. （1987）, Calibrating AIS images using the surface as a reference.In Proc.3rd Airborne Imaging Spectrometer Data Analysis Workshop （G. Vane, Eds.）, JPL Publ. 87-30,JPL Laboratory, Pasadena,CA, PP. 63-69.

[100]. Smith,M.O., Ustin,S.L., Adams,J.B., and Gillespie,A.R.（1990）, Vegetation in deserts:1. Aregional measure of abundance from multispectral images., Remote Sensing Environment,31, PP. 1-26.

[101]. Swayze, G.A., and Clark, R.N., Spectral identification of minerals using imaging spectrometry data: evaluating the effects of signal to noise and spectral resolution using the Tricorder Algorithm: Summaries of the Fifth Annual JPL Airborne Earth Science Workshop, January 23- 26, R.O. Green, Ed., JPL Publication 95-1, p. 157-158, 1995.

[102]. The Committee for Advanced DBMS Function. Third-Generation Database System Manifesto, SIGMOD RECORD, 1990
[103]. Timothy A. Warner and Michael C. Shank, 1996, Spatial autocorrelation analysis of hyperspectral imagery for feature selection, Remote Sensing of Environment, Vol. 60, PP. 58-70.
[104]. Tung A K H, Hou J, Han J. Spatial Clustering in the Presences of Bostacles. IEEE Transactions on Data Engineering, 2001,11, PP. 359-369
[105]. Van G., Goetz, A.F.H., Terrestrial Imaging Spectroscopy, Remote Sensing of Environment, Vol. 37, 23-34, 1991
[106]. Vane Gregg, Geotz Alexander F H. Terresttrial imaging spectrometry: current status, future trends[J], Remote Sensing of Environment, 1993, 44:117-126
[107]. W.H.Inmon, Building the Data Warehouse, John Wiley & Sons. Inc. New York, 1993
[108]. 白继伟，基于高光谱数据库的光谱匹配分类技术研究，硕士论文，中国科学院遥感应用研究所，2002 年 6 月
[109]. 薄华、马缚龙、焦李成，图像数据挖掘的模型和技术，西安邮电学院学报，2004.7，vol.9 no.3
[110]. 布和敖斯尔，基于知识发现和决策规则的盐碱地遥感分类方法研究，中国图像图形学报，1999，4[Ａ]（11），PP.965～969
[111]. 蔡聪明，毕式定理的两个推广，http://episte.math.ntu.edu.tw/articles/sm/sm_25_12_1/.
[112]. 陈春香，数据发现在地球化学数据处理中的应用，桂林工业学报，1999，19（13），PP.230-240
[113]. 陈华，曹锦云，郑学峰，郑刚，腰杆图像数据库的ＷＥＢ访问实现，电脑开发与运用，第 17 卷，第 12 期，2004， p12-14
[114]. 陈述彭，童庆禧，郭华东. 遥感信息机理研究. 北京. 科学出版社. 1998.7. Page 139, iii
[115]. 程继华、施鹏飞，多层次关联规则的有效挖掘算法.软件学报，1998，第 12 卷第 9 期，PP. 937～941
[116]. 程继华、施鹏飞、魏暑生，基于概念的关联规则的挖掘，郑州大学学报（自然科学版），1998，30（20），PP. 27-30
[117]. 戴晓军、淦文燕、李德毅，基于数据长的图像数据挖掘研究，计算机工程与应用，2004，第 26 期，PP.41-44
[118]. 邸凯昌、李德仁、李德毅，Rough 集理论及其在 GIS 属性分析和知识发现中的应用，武汉测绘科技大学学报，1999，24（1）， PP. 6-10
[119]. 邸凯昌、李德仁、李德毅，基于空间数据发掘的遥感图像分类方法研究，武汉测绘科技大学学报，2000 年 2 月，第 25 卷第 1 期，PP.42-48
[120]. 丁祥武，挖掘时态关联规则，武汉交通科技大学学报，1999,23（4），PP. 365-367
[121]. 丁祥武，序列模式的一种模型及其挖掘，中南民族学院学报（自然科学版），1999，18（2），PP.44-483
[122]. 丁祥武.挖掘关联规则的一种预处理:合并交易.中南民族学院学报（自然科学版），1999，18（3），PP.21-25

[123]. 杜培军、陈云浩，高光谱遥感信息智能处理的若干理论与技术问题，科技导报，第 24 卷总 211 期，2006， PP. 47-51
[124]. 杜云艳、苏奋振、杨晓梅、王敬贵、陈秀法, 中国海岸带及近海科学数据平台研究与开发，海洋学报，Vol.26, No.6, 2004.11
[125]. 方涛，龚健雅，李德仁，影像数据库建立中的若干关键技术，武汉测绘科技大学学报，第 22 卷第 3 期，1997 年 9 月， PP. 266-269
[126]. 甘甫平、刘圣伟、周强，德兴铜矿矿山污染高光谱遥感直接识别研究，地球科学——中国地质大学学报，Vol.29,No.1,2004
[127]. 甘甫平、王润生、郭小方、王青华，高光谱遥感信息提取与地质应用前景——以青藏高原为试验区，国土资源遥感，No.3,2000
[128]. 耿修瑞，2004，高光谱遥感图像目标探测与分类技术研究，中科院遥感所博士学位论文
[129]. 宫辉力、赵文吉、李京，多元遥感数据挖掘系统技术框架，中国图像图形学报，2005 年 5 月，第十卷第五期，PP.620-623
[130]. 宫鹏，浦瑞良,郁彬.1998.不同季相针叶树种高光谱数掘识别分析．遥感学报．2（ 3）：211-217
[131]. 纪钢、张小川，图像信息处理及图像数据库模型分析，计算机工程与应用，2003.8， PP. 64-65
[132]. 江寒， 陈露， Google 地图服务大事记， 2006-3-21， http://www.3snews.net/modules/article/view.article.php?37/c1
[133]. 冷秀华，张杰，马毅，宋平舰， 高光谱遥感数据管理系统原型设计，第十四届全国遥感技术学术交流会论文集，2003 年 10 月
[134]. 李德仁、王树良、李德毅、王新洲，论空间数据挖掘与知识发现的理论与方法，武汉大学学报 信息科学版，第 27 卷第 3 期，2002.6，PP.221-233
[135]. 李德仁、王树良、史文中、王新洲，论空间数据挖掘和知识发现，武汉大学学报 信息科学版，第 26 卷第 6 期，2001.12，PP.491-499
[136]. 李飞鹏，秦前清，李德仁，海量遥感影像数据库实时压缩系统的设计与实现，计算机工程与应用，第 26 期，2003 年， PP. 9-11
[137]. 李航、岳丽华，基于 COM 和 ArcSDE 的遥感图像数据库的开发，计算机应用，第 25 卷，第 5 期，2005 年 5 月， PP. 1212-1214
[138]. 李军，刘高焕，迟耀斌，朱重光， 大型遥感图像处理系统中集成数据库设计及应用，遥感学报，Vol.5,No.1, PP. 41-45, 2001.1
[139]. 李军、李琦、毛东军、郭玲玲，遥感影像数据库研究，计算机工程与应用, 第 27 期， PP. 32-35，2003.
[140]. 李宗华，彭明军，基于关系数据库技术的遥感影像数据建库研究，武汉大学学报信息科学版，第 30 卷第 2 期，2005 年 2 月， PP. 166-169
[141]. 刘大昕，张春林，聂亚杰，张子杨，数据仓库与 OLAP 技术，计算机仿真，第 20 卷第 5 期，2003 年 5 月， PP. 40-43
[142]. 刘建平、赵英时，孙淑玲，高光谱遥感数据最佳波段选择方法试验研究，遥感技术与应用，第 16 卷，第 1 期，2001， PP. 8-13
[143]. 刘明宇，王钰，郑崇勋，燕楠，应用非负矩阵分解方法提取注意力相关脑电特征，生物物理学报，第 22 卷，第 1 期，2006 年 2 月， PP. 67-72
[144]. 刘鹏，毕建涛，曹彦荣，何建邦，遥感影像数据库引擎设计与实现，地球信息科学，第七卷，第二期，2005.6， PP. 105-110

[145]. 刘伟东，高光谱遥感土壤信息提取与挖掘模型，博士论文，中国科学院遥感应用研究所，2002

[146]. 刘卫忠，徐重阳，蔷薇，多层客户机/服务器结构分析，网络世界，2000年1月，http://www.cnw.com.cn/cnw_old/2000/htm2000/B87DF4F0748543F0B3ADFFB794F32422.htm

[147]. 刘钊、蒋良孝，图像数据挖掘之研究，计算机工程与应用，2003 年，第 33 期，PP.202-204

[148]. 刘志刚, 秦前清, 李德仁等. 基于混合核函数的支撑向量机及其在遥感影像土地利用中的分类. 测绘信息与工程, 2003, 28（5）:PP. 1- 3.

[149]. 骆剑承, 周成虎, 梁怡, 等. 支撑向量机及其遥感影像空间特征提取和分类的应用研究,遥感学报, 2002, 6（1）: 50- 55.

[150]. 马超飞，刘建强，遥感图像多维量化关联规则挖掘，遥感技术与应用，第 18 卷第 4 期，2003 年 8 月， PP. 243-247

[151]. 马建文、马超飞，基于空间角度理论的卫星光学遥感数据认知与挖掘，中国图像图形学报，1999，4[Ａ]（11），PP.918-923

[152]. 孟小峰、周龙骧、王珊，数据库技术发展趋势，软件学报，Vol.15, No.12, 2004, PP. 1822-1836

[153]. 邵峰晶，于忠清，数据挖掘——原理与算法，中国水利水电出版社，北京，2003 年

[154]. 沈清，汤霖编著，模式识别导论，国防科技大学出版社，1991 年 5 月第一版

[155]. 帅艳民、朱启疆、王培娟、王锦地、刘素红、白香花，地物波谱数据仓库系统设计研究，计算机工程与应用，2003 年，第 33 期，PP.199-201

[156]. 谭倩、赵永超、童庆禧、郑兰芬，植被光谱为特征提取模型，遥感信息，2001.1，PP.14-18

[157]. 陶 卿.一种新的机器学习算法：Support Vector Machines，模式识别与人工智能，2000 年 9 月，Vol.13，No.3

[158]. 陶冶宇，马东洋，徐青，解志刚，基于 ORACLE 多分辨率遥感影像数据库的设计，测绘学院学报，第 22 卷第 1 期，2005 年 3 月， PP. 65-68

[159]. 童庆禧，高光谱遥感的现状与未来，遥感学报，第七卷，增刊， PP. 1-12，2003 年。

[160]. 童庆禧，遥感信息传输及其成像机理研究，中国科学院院刊，第 1 期，PP. 31-33，2002 年。

[161]. 童庆禧，遥感信息获取技术的研究与发展，遥感应用的实践与创新，测绘出版社， PP. 50-55。1990

[162]. 童庆禧，郑兰芬，高光谱遥感发展现状，遥感知识创新文集，中国科学技术出版社， PP. 13-25, 1999

[163]. 童庆禧，郑兰芬，王晋年等，1997。湿地植被成象光谱遥感研究，遥感学报，1（1）：50－57。

[164]. 涂星原， 基于数值属性的关联规则的挖掘，郑州工业大学学报，1998，19（3），PP.72-75

[165]. 汪鹏, 非负矩阵分解：数学的奇妙力量，计算机教育，第 10 期，2004，PP. 38-40

[166]. 王晋年，张兵等，以地物识别和分类为目标的高光谱数据挖掘，中国图象图形学报，1999 年 11 月，第 4 卷第 11 期， PP. 957-964.
[167]. 王凯峰，秦前清，基于单类 SVM 的遥感图像目标检测，计算机工程与应用，第 32 期，2005， PP. 63-64
[168]. 王密，龚健雅，李德仁，大型遥感影像数据库的空间无缝数据组织，武汉大学学报信息科学版，第 26 卷第 5 期，2001 年 10 月， PP. 419-424
[169]. 王双成、林士敏、陆玉昌，贝叶斯网络结构学习分析，计算机科学，第 27 卷第 10 期，2000， PP. 77-79
[170]. 王宇飞，基于网络的遥感影像服务系统及技术研究，博士论文，中科院遥感所，2001
[171]. 吴信才、郭玲玲、李军，RDBMS 和 COM 的海量遥感影像数据的管理与 WEB 发布，中国图像图形学报，Vol.7A, No.4, 2002.4
[172]. 肖利，挖掘序列形式的模型研究，计算机科学，1998（专刊），PP135-136
[173]. 肖利、金远平、徐宏炳，一个新的挖掘广义关联规则算法，东南大学学报，1997，27（6）， PP.76-81
[174]. 许龙飞、杨晓昀， KDD 中广义关联规则发现技术研究，计算机工程与应用，1998 年 9 月，PP. 32-35
[175]. 阎平凡，最小描述长度与多层前馈网络设计中的一些问题，模式识别与人工智能，第 6 卷第 2 期，1993 年， PP. 143-148
[176]. 燕守勋、张兵、赵永超、郑兰芬、童庆禧、杨凯，矿物与岩石的可见-近红外光谱特性综述，遥感技术与应用，第 18 卷，第 4 期， PP. 191-201，2003
[177]. 燕守勋、李兴、张兵，蒙皂石与膨胀土光谱吸收参量相关关系研究，遥感学报，第九卷第三期，2005 年 5 月，PP.328-337
[178]. 杨勤，基于 COM 的三层客户/服务器模型，计算机应用研究，第二期，2001 年， PP. 109-111
[179]. 杨学兵、刘胜军、蔡庆生，一种实时过程控制中的数据挖掘算法研究，计算机应用，1999，19（9）， PP.8-10
[180]. 杨宇艇，潘云鹤，庄越挺. 图像数据库的研究现状与发展. 计算机科学，Vol.23,No.24, 1996, PP. 8-15
[181]. 叶静、蔡之华，遥感图像中的数据挖掘应用概述，计算机与现代化，2003 年第 10 期，PP.36-38
[182]. 曾澜. 21 世纪初我国国家空间信息基础设施发展总体思路.中国遥感奋进创新二十年学术论文集. 气象出版社. 2001.
[183]. 张兵，2002,时空信息辅助下的高光谱数据挖掘。中国科学院遥感应用研究所博士论文。
[184]. 张芬，高炎，多分辨率无缝数据库在影像数据库系统中的应用，测绘通报，第 4 期，2005 年， PP. 40-42
[185]. 张霞，光谱指数时间谱特性研究及其在种植模式信息提取中的应用，博士论文，中国科学院遥感应用研究所，2006 年 1 月
[186]. 张雄飞，网络环境下高光谱数据库构建及其应用实践，硕士论文，中国科学院遥感应用研究所，2003 年 6 月
[187]. 张雄飞，张兵，张霞，郑兰芬，童庆禧， 高光谱数据在数据库中的高效存储技术研究，遥感学报，Vol.8 No.5， PP. 404-408，2004

[188]. 张宇、王希勤、彭应宁，自适应中心加权的改进均值滤波算法，清华大学学报自然科学版，1999 年 第 39 卷 第 9 期
[189]. 张志，应用服务器的发展趋势，北京城市学院学报，总第 71 期，2005 年 9 月， PP. 81-83
[190]. 赵艳玲，何贤强，王迪峰，潘德炉，基于 web 海洋卫星遥感产品的查询系统，东海海洋，第 23 卷第 1 期，2005 年 3 月， PP. 32-39
[191]. 郑庆良，张翔，杨莹，网络服务器模型分析与实现，杭州电子工业学院学报，第 24 卷第 4 期，2004 年 8 月， PP. 95-98
[192]. 周成虎、张健挺，基于信息熵的地学空间数据挖掘模型，中国图像图形学报，1999，4[A]（11），PP.943-951
[193]. 左万利，含有类别属性数据库中联系性规则的挖掘，吉林大学自然科学学报，1999（1），PP.33-37

# 博士期间发表文章

1） Xing LI, Bing ZHANG, Qingxi TONG, Wenjuan ZHANG. Demand-oriented Hyperspectral Database and Its Applications. International Geo-science and Remote Sensing Symposium, 2005, 5: 3227-3230

2） Xing Li, Xingtang Hu, Bing Zhang, Xia Zhang, Junsheng Li, Xiaoying Li, A hyperspectral-environmental database in China. SPIE proceedings on Remote Sensing for Environmental Monitoring, GIS Applications, and Geology. 2005, Vol. 5983:374-382.

3） Xing Li, Xia Zhang, Bing Zhang, Qingxi Tong, A Web-based Spectral Database for Precision Agriculture and its applications in China, The 8th International Conference on Precision Agriculture: *Oral Abstract Accepted*

4） 李兴、张兵、张霞、李俊生，高光谱数据仓库模型设计，遥感学报，2003 增刊，7： 61－68。

5） 李兴，张兵，胡兴堂，燕守勋。一种基于线性混合光谱理论的岩石光谱模拟模型。遥感学报，2004 增刊第 8 卷，41-48。

6） 燕守勋、李兴、张兵，蒙皂石与膨胀土光谱吸收参量相关关系研究，遥感学报，第九卷第三期，2005 年 5 月，PP.328-337

7） 张 兵，李兴，燕守勋，甘甫平，秦 善，易维宁。我国典型岩矿波谱数据及其应用，遥感学报 2004 增刊第 8 卷，36-41。

8） 张霞，李兴，李俊生，卫征，利用光谱指数监测作物长势变化研究，遥感学报，2003 增刊，7：120－124。

9） Bing Zhang, Xing Li, Xia Zhang, et al., Image classification supported by digital geomorphology model, Proceedings of SPIE, 2003 , Vol. 5286:700-703.

10） Xiongfei Zhang, Xing Li, Xia Zhang, et al., Preliminary research in using the technology of data mining to analyze remote sensing data, Proceedings of SPIE, 2003，Vol. 5286, P. 980-985.

11） Jiwei Bai, Xing Li, Xingtang Hu, Xiongfei Zhang, Yongchao Zhao, Bing Zhang, Qingxi Tong, and Lanfen Zheng. Classification methods of the hyperspectral image based on the continuum-removed. Proc. SPIE 2003, Vol. 4897: Multispectral and Hyperspectral Remote Sensing Instruments and Applications, p325-329.

12） 胡兴堂，张兵，李兴，高连如，面向专业应用的高光谱图像处理系统体系结构设计，遥感学报，2003 增刊，7： 54－60。

13） Wenjuan ZHANG, Bing Zhang, Xing Li, Xia Zhang. Online Analysis and

Management of Spectral Data in Spectral Database. International Geo-science and Remote Sensing Symposium, 2005, 5: 3212-3214. (IGARSS 05).

14） WEI Zheng, ZHANG Xia, ZHANG Bing, LI Xing. General and Specific Methods Studying on Bands Selection of Hyperspectral Remote Sensing Data. International Geo-science and Remote Sensing Symposium, 2005, 5: 3223-3226.

# 博士期间参与项目

2002/10-2005/06：国家典型地物标准岩矿波谱数据库。负责三个子课题的进度控制、质量控制；岩矿模拟与识别模型研究；光谱数据后期处理；新疆地区影像预处理与镶嵌；汇编报告与项目验收。财务决算。

2004/10-2005/03：面向精准农业的光谱数据库系统。数据库设计、项目计划、进度控制、质量控制；系统设计、数据库开发；财务预算与决算。

2003/10-2004/06：环境遥感监测系统。负责系统中数据库部分设计，包括遥感图像数据库、环境监测数据库、环境背景数据库等。

2003/04-2004/04：多维遥感信息处理系统。主要负责其中的高光谱影像子系统部分的系统设计、接口，以及数据库设计，同时全面负责本子系统的项目管理工作。

2002/10-2003/10：高光谱图像处理与分析系统。负责界面设计和系统总体框架设计、高光谱图像分类算法分析与研究、软件测试。

2002/10-2002/12：植被生化参量反演。采用决策树方法对名古屋地区高光谱进行精细分类。

2001/03-2001/05：顺义地区航空高光谱数据获取。地面光谱数据采集、光谱数据预处理、遥感影像辐射校正、几何校正等预处理工作。

2000/09-2000/12：HIPAS 制图与统计系统。利用 ArcView 进行二次开发。

# 致 谢

经过数个月的夜以继日，终于完成本文。回想起进入遥感所的五年，不禁感慨万千。从二十二岁到二十七岁，这是人生最宝贵的五年，我因在中科院遥感所高光谱室度过这五年而感到幸运。孔子曾曰：吾十有五而志于学，三十而立。十五岁进入高中学习，而如今已近而立，我也更加明确我的人生目标和发展方向。

在进入中国科学院直接攻读博士学位的这五年里，研究生院一年课程学习，拓展了知识层次与宽度；之后在遥感所跟随导师从事科研项目研究，三年来工作能力和研究能力不断提高；最后一年集中进行论文写作，遨游在高光谱遥感与数据库的知识海洋中，每当发现一两朵美丽而充满希望的浪花，欣喜之情便溢于言表。

父亲说：博士论文是大事，是寒窗廿年的结晶。在此，我要感谢这五年里我一路走来，对我关爱、关心、帮助的许许多多老师和同学们，没有你们，也就没有这篇博士论文。

首先衷心感谢我的导师童庆禧院士对我研究工作的精心指导，您不断地从思想上、方法上引导我走上一条通往科学殿堂的路。您渊博的知识、敏锐的洞察力、高屋建瓴的概括能力和思维方式、以及旺盛的工作精力永远是我学习的榜样！这将一直激励着我在将来的工作和学习中不断进取。

衷心感谢我的导师郑兰芬研究员。您对我的日常生活与工作给予了细致入微的关怀，曾无数次给我耐心的指导，使我能够很快地掌握学科知识，并融会到科研工作中去。您经常与我分享您的人生感悟，教导我做人做事的原则，让我能够在顺境和逆境中保持一颗平常心。

衷心感谢我的导师张兵研究员。几年来，您一直是我的良师益友，在研究工作和生活方面给予我诸多帮助，给我创造了一个十分舒适完备的研究环境。您对高光谱遥感透彻的理解和众多的创意让我能够拓宽研究思路、触类旁通。您是我们研究室的主任，然而您对待学生如同对待一个真诚的朋友，让我时时刻刻体会到您无微不至的关心。您对实验室的管理也非常的人性化和科学化，我们实验室和谐团结、积极向上的环境与您卓越的管理是分不开的。

感谢燕守勋研究员对我研究工作的帮助。您淡泊明志，对待科研与生活的态度值得我学习。记得与您同坐在遥感所门前的阶梯上畅聊锡林格勒大草原，当时，我感受到科研生活的另一种美。

感谢北京市农林科学研究院农业信息技术研究中心刘良云研究员、国土资源部航空物探中心甘甫平博士、中国科学院安徽光机所易维宁研究员、北京大学地质系秦缮教授、北京师范大学资源与环境学院刘素红教授对我研究工作的众多建议和支持。

感谢人教处余琦主任、吴晓青老师、刘戈平老师对我完成学业的支持。你们热情的工作、对学生认真负责的态度、从学生角度换位思考问题的方式，让我们感觉到：遥感所是我们的家。

感谢我们课题组的全体成员对我学习和生活上的帮助。这个集体包括：张霞、周莉萍、刘建贵、谭倩、赵永超、刘团结、刘良云、吴传庆、白继伟、刘伟东、张雄飞、陈正超、耿修瑞、方俊永、卫征、胡兴堂、胡方超、高连如、李俊生、张文娟、刘学、焦全军、申茜、罗文斐、刘翔、张靓、张浩、李儒等。感谢你们与我一路同行、伴随我走过这不一般的五年！我永远是你们最亲密的兄弟！在此，不禁回想起以前和你们在一起的一幕一幕：2000 年 10 月在实验室与清华大学的韦麟等熬夜共同完成 HIPAS 二期项目；2000 年 12 在昌平一小镇利用 ARCVIEW 连夜赶制亚运村地图；2001 年 3 月在九华山庄旁边的北京精准农业基地测量光谱；2001 年 5 月在北京五环北大未名湖测量水体光谱；2002 年 7 月在温州雁荡山聆听全国成像光谱领域专家报告； 2002 年圣诞前夕在遥感所前面的草坪上打雪仗；2003 年 4 月在天津蓟县测量光谱及清东陵半日游；2003 年 6 月在周口店测量光谱及百花洞半日游；2003 年 10 月第一次进入人民大会堂参加第三世界科学院院士大会，聆听国家主席报告；2003 年 11 月前往安徽合肥送岩矿样本并长途奔袭武汉索取软件资料；2004 年 2 月只身独往合肥察看项目进度；2004 年 4 月转战某一军事院校测量光谱，儿时驾乘坦克的愿望得以实现；2004 年 4 月平谷两日游，其乐融融；2004 年 5 月我成功联系第八届科博会，让大家都去人民大会堂听领导的报告；2004 年 6 月，谭倩师姐回娘家，聊天吃饭；2004 年 10 月雁栖湖秋游；2004 年 11 月在国贸参加国际对地观测技术与应用博览会；2004 年 12 月好伦哥聚餐；2005 年 2 月在赛迪大酒店闭关撰写《高光谱遥感科学》一书；2005 年 6 月手头上负责的两个项目验收；2006 年 5 月再次平谷二日游……真挚友谊、拳拳在心。

最后我要感谢父母对我的抚育之恩，是你们给了我健康的体魄，让我知道如何为人处事；感谢我的妹妹，你们无私的关爱让我体会到亲情的可贵；感谢我的夫人，你的爱，是我生命的动力。谨以此文献给你们！

Life is not a race, but a journey to be savored each step of the way!

李 兴

2006 年 5 月 26 日