

高光谱影像波段选择算法研究^{*}

苏红军¹, 杜培军², 盛业华¹

(1. 南京师范大学 虚拟地理环境教育部重点实验室, 南京 210046; 2. 中国矿业大学 地理信息与遥感科学系, 江苏 徐州 221008)

摘 要: 基于高光谱影像数据的特点, 分析了高光谱数据的降维方法。着重探讨了波段选择的若干算法: 熵及联合熵、最佳指数因子、自动子空间划分、自适应波段选择、波段指数和最优波段指数等算法。分析了各种算法的有效性、局限性和计算复杂度, 并针对波段指数的不足, 设计了最优波段指数(OBI) 波段选择新算法。最后通过具体的试验, 验证了各种算法的性能。

关键词: 高光谱影像; 波段选择; 特征提取与选择

中图分类号: TP751 文献标志码: A 文章编号: 1001-3695(2008) 04-1093-04

Study on band selection algorithms of hyperspectral image data

SU Hong-jun¹, DU Pei-jun², SHENG Ye-hua¹

(1. Key Laboratory of Virtual Geographic Environment, Nanjing Normal University, Nanjing 210046, China; 2. Dept. of Geographic Information & Remote Sensing Science, China University of Mining & Technology, Xuzhou Jiangsu 221008, China)

Abstract: Based on the characteristic of the HRS image data, the method of dimensionality reduction was discussed. And then some key algorithms such as joint entropy, optimal index factor, auto-subspace partition, adaptive band selection, band index and optimal band index were analyzed. In addition, the effectivity, shortage and complexity between different algorithms were discussed and compared. According to the shortage of band index, the new band selection method which named optimal band index (OBI) was proposed. At last, the feasibility of proposed methods was proved by experiment results.

Key words: hyperspectral image; band selection; feature extraction and selection

高光谱遥感是 20 世纪最后二十年, 人类在对地观测方面取得的重大技术突破之一, 也是当前及今后几十年内的遥感前沿技术。它利用成像光谱仪纳米级的光谱分辨率, 获取许多非常窄且光谱连续的图像数据, 实现地物空间、辐射、光谱信息的同步获取; 为每个像元提供数十至数百个窄波段的光谱信息, 并生成一条完整而连续的光谱曲线^[1~3]。

高光谱遥感大量的光谱波段数据为人们了解地物提供了丰富的信息, 这对于后续进行的地物分类和目标识别是十分有益的。然而波段的增多也必然导致信息的冗余和数据处理难度的增加。如何既能有效利用高光谱数据的最大信息, 又能较快地处理高光谱数据成为高光谱数据处理的研究热点和未来发展方向。其中有关光谱特征提取与选择的研究是一个重点。基于此, 本文在分析高光谱数据特征提取与选择方法的基础上, 着重探讨了高光谱影像波段选择的若干算法。

1 特征提取与特征选择

一般来说, 面对高光谱数十、数百的波段信息, 全部利用所有波段是不现实的, 所以此时高光谱降维操作就显得尤为重要。高光谱降维有两种方法, 即特征选择和特征提取。特征提取与选择的基本任务是如何从许多特征中找出那些最有效的

特征^[4]。在样本数不是很多的情况下, 用很多特征进行分类, 无论从计算的复杂度还是性能上来说均是不适宜的。因此如何将高维特征空间压缩到低维特征空间以便进行有效的分类成为一个重要的研究方向。

目前压缩波段有两种方法: 一是从众多波段中选择感兴趣的若干波段, 或选择信息量大、相关性小的若干波段; 二是利用所有波段, 通过数学变换来压缩波段, 如主成分分析法(PCA)、典型分析(CA) 等^[5]。本文主要介绍第一种降维方法, 即波段选择方法。波段选择是遥感图像识别与分类的重要环节之一, 对于高光谱遥感图像的特征选择和提取, 最初一般都集中在波段选择上^[6]。至今形成了比较成熟的波段选择算法, 主要分成两类: 一是基于信息量的最佳波段选择, 如熵与联合熵、组合波段的协方差矩阵行列式、最佳指数(OIF) ; 二是基于类间可分性的波段选择, 如均值间的标准距离、离散度、B 距离、类间平均可分性等。本文重点研究第一类算法。

目前, 国内外对波段选择算法进行了很多研究。E. Arzua-ga-Cruz 等人^[7]设计了信息散度波段子集选择(information divergence band subset selection, IDSS) 算法, 并用模拟数据及遥感数据进行了验证, 效果较好。Paskaleva 等人^[8]基于数据主成分的分布, 提出了以波段不同主成分间最小距离、波段主成

收稿日期: 2007-02-27; 修回日期: 2007-04-10 基金项目: 国家自然科学基金资助项目(40401038) ; 地理空间信息工程国家测绘局重点实验室开放基金资助项目; 中国矿业大学科学基金资助项目(D200403)

作者简介: 苏红军(1985-), 男, 河南永城人, 硕士研究生, 主要研究方向为高光谱遥感数据处理、虚拟地理环境(hjsul@163.com); 杜培军(1975-), 男, 教授, 博导, 主要研究方向为高光谱遥感、遥感信息融合与数据挖掘、资源环境遥感; 盛业华(1965-), 男, 教授, 博导, 主要研究方向为 GIS 及虚拟现实技术、摄影测量与遥感。

分相同分布下的 Kullback-Liebler 散度变化和波段连续幅度间的经验相对变化等作为波段选择标准的新的波段选择策略,并应用于岩石及矿物的分类识别应用。Wang Su 等人^[9]提出了线性约束波段选择(linear constrained band selection, LCBS)的新的光谱影像波段选择算法,该算法在线性约束某波段影像的同时,最小化了该波段影像与其余波段影像的波段相关性和依赖性。P. Bajcsy 等人^[10]在系统总结前人研究成果的基础上,建立了基于监督和非监督算法的高光谱遥感波段选择算法方法论体系,该研究具有很高的指导价值。以上研究扩展和促进了高光谱波段选择算法的发展,进一步完善了高光谱遥感信息处理技术体系;然而高光谱遥感信息处理技术和算法仍需要不断完善。

2 波段选择主要算法

波段选择属于特征选择的一种,通过特征选择,可以强化那些最具有可分性的光谱波段。一般来说,选择最佳波段的原则有三点:一是所选择的波段的信息量应最大;二是所选择的波段数据间的相关性要小;三是研究区内欲识别地物的光谱响应特点能使某些类别地物之间最容易区分。那些信息含量多、相关性小、地物光谱差异大、可分性好的波段就是应该选择的最佳波段^[11~13]。

图 1 为针对高光谱影像的波段选择示意图。其中:影像预处理包括消除噪声、辐射校正、几何校正等;基本统计量包括均值、标准差、相关系数等,作为后续处理的依据。



图1 影像波段选择示意图

针对高光谱的不同应用目的,所选择的波段数目是不同的。一般来说,大多数所要求的波段是三个波段的组合,以便在处理影像时对三个波段分别赋予 R、G、B 颜色值,进行影像的后处理。但由于应用的目的不同,也有要求选择任意 N 个波段的。针对这种情况,本文研究的波段选择算法可以提供三波段组合也可以提供任意 N 个波段。

最终波段选择的方法有两种:一种是根据用户设定的阈值,指数大于阈值的波段就被选择出来;另一种方法是选择波段指数排列在前面的 N 个波段。

2.1 熵及联合熵(joint entropy, JE)

根据香农信息论理论,信息熵是在平均意义上来表征信息源总体特征的量;信息熵越大,图像所含信息越丰富。对高光谱数据而言,第 i 波段图像的熵值 H(i) 等于其图像辐射亮度值的概率与其概率对数的乘积。在此基础上,根据联合概率分布知识,可以得到三个波段图像的联合熵 H(i₁, i₂, i₃)。对所有可能的波段组合计算其联合熵,并按从大到小的顺序进行排序,联合熵最大的前 N 个波段组合就是最佳的。

实验中发现,信息熵算法虽然简单、易理解,但也有其缺

点。其计算时空复杂度相当高,对计算机性能要求也高。而且由于高光谱遥感数据的波段信息之间的强相关性,使得多种波段组合方式具有相同的联合熵,这也使算法的效果不太理想。

2.2 最佳指数因子(optimal index factor, OIF)

高光谱图像数据波段间标准差大,则信息量丰富;而波段间的相关系数小,则波段信息冗余度小。根据以上原理,Chavez 等人^[14]于 1982 年提出了一种组合波段的选优方法即最佳指数因子。该方法依据下式给出 N 个波段组合中最优的指数大小:

$$OIF = \sum_{i=1}^n S_i / \sum_{i=1}^n \sum_{j=i+1}^n |R_{i,j}|$$

(1)

其中: S_i 为第 i 个波段的标准差; R_{i,j} 表示第 i 个波段与第 j 个波段之间的相关系数。

选择的波段数目 n 一般取 3,即将所有可能的三个波段组合在一起。OIF 越大,则相应组合波段图像的信息量越大。在实际应用中,这种方法存在其局限性。首先,它选择出来的最优波段组合未必是最优的;其次, OIF 算法对于高光谱图像波段选择而言计算量仍然过大。比如计算波段间的相关系数,在实验中就要运行 40 min 以上。

2.3 自动子空间划分(auto-subspace partition, ASP)

高光谱影像的一个显著特点是,相邻谱带间存在较强的相关性,且这种谱间相关性比空间相关性要强得多^[15]。所以应用联合熵算法,所选出的波段基本上集中在一个连续的子空间;而连续的子空间往往相似性大,这就造成信息的重复使用,影响了后续的处理效果。波段数据的全局统计特性与局部特性存在差异,因而在全空间进行波段选择并不一定能选出最佳波段^[16]。

解决上述问题的思路就是将全部波段划分为若干个子空间,然后再进行波段选择。谷延锋等人^[15]提出一种自动子空间划分方法。该方法通过定义波段相关系数矩阵及其近邻可传递相关矢量,将高光谱数据空间划分为适合的数据子空间。这种划分方法有着充分的理论依据,反映了数据的局部特性。自动子空间划分就是根据相关系数矩阵灰度图的成块特点,依据高光谱影像相邻波段相关系数的大小,将波段划分为若干个子空间。本文提出了在波段选择前首先对所有波段进行分组,划分为若干个子空间,然后再进行波段选择的改进方案。

$$R_{i,j} = \frac{\sum_{k=1}^n (x_{ik} - \bar{x})(y_{jk} - \bar{y})}{\sqrt{\sum_{k=1}^n (x_{ik} - \bar{x})^2 \sum_{k=1}^n (y_{jk} - \bar{y})^2}}$$

(2)

其中: R_{i,j} 是波段 x_i 与 y_j 间的相关系数; x_{ik} 和 y_{jk} 分别是该波段影像内的第 k 个像素; n 代表一个波段图像内像素的总数目; ̄x 和 ̄y 分别是该波段 x_i 和 y_j 的均值。根据式(2),可以计算高光谱所有波段数据之间的相关矩阵。

2.4 自适应波段选择(adaptive band selection, ABS)

针对 OIF 方法在实际应用中存在的局限性,刘春红等人^[5]对 OIF 方法和其他方法进行充分研究之后,提出自适应波段选择的方法。该方法充分考虑了各波段的空间相关性和谱间相关性。其公式如下:

$$I_j = S_i / [(R_{i-1,i} + R_{i,j+1}) / 2]$$

(3)

其中: S_i 为第 i 个波段的标准差; R_{i-1,i} 和 R_{i,i+1} 是第 i 个波段与其前后两波段的相关系数或第 i 个波段与任意两个波段的相关系数; I_j 是第 i 幅图像的指数大小。

需要说明的是,当 $R_{i-1,i}$ 和 $R_{i,i+1}$ 取第 i 个波段与其前后两波段的相关系数时,得到的是三个波段组合的指数值;当 $R_{i-1,i}$ 和 $R_{i,i+1}$ 取第 i 个波段与任意两个波段的相关系数时,得到的是单个波段的指数值。所以该方法既可以得到三个波段的组合结果,也可以得到任意 N 个波段的结果。这可由用户根据不同的需要来选择。

这种方法充分考虑了每幅图像的信息富集程度与相邻波段的相似性。指数越大,相应波段组合图像的信息量就越大,越具有代表性。

2.5 波段指数(band index, BI)

由于基于全局的波段选择算法,选择的波段往往是连续地集中在某一个连续的子空间中。而连续的子空间往往相似性大,这就造成信息的重复使用,影响了后续的处理效果。姜小光等人^[12]提出了一种基于波段指数的选择方法,该算法先根据相关系数矩阵对所有波段进行分组,再分别计算每一波段的波段指数,最后进行波段选择。

设整个高光谱空间共有 N 个波段,被分为 k 组,每组的波段数目分别为 $n_1, n_2, n_3, \dots, n_k$; 定义的数学模型如下:

$$P_i = \sigma_i / R_i \tag{4}$$

$$R_i = R_w + R_o \tag{5}$$

其中: σ_i 为第 i 波段的均方差; R_w 为第 i 波段与所在子空间内其他波段相关系数的绝对值之和的平均值; R_o 为第 i 波段与所在子空间以外的其他波段之间的相关系数的绝对值之和。

由于同一组内各波段的相关性强而组间波段的相关性弱,一个波段的整体相关性强弱主要与其与组内各波段的相关性大小决定。各个组的大小即构成子空间的波段数不同,因此用组内一波段与其他组波段相关系数绝对值之和的平均值作为其波段指数分母的一项,能合理地反映该波段的整体优劣水平。所以波段指数能综合地反映通道信息含量和相关性两个因素,可作为选择波段的重要参数之一。

2.6 最优波段指数(optimal band index, OBI)

波段指数方法虽然综合考虑了波段的信息含量和相关性两个因素,但由于该指数基于成像光谱数据具有分组成块的结构特点而设计,它在分组时,每个组的维数是不同的,造成了每组的 R_o 由于所处组维数的不同而人为地影响了计算的效果。而且如果分组没有考虑影像的空间结构信息,之前并没有对噪声波段进行处理,可能会影响到最终的计算结果。

基于以上考虑,本文提出了一个用最优波段系数来衡量波段选择标准的新的改进算法。分形几何中的分形维数是表达空间结构、空间复杂度的一个工具和度量指标。因此,可以用遥感影像分形维数来定量表示其空间结构的变化。其基本思想是:首先用分形理论中的分维数对噪声波段进行删除,预选出影像空间结构、质量及光谱特征较好的波段;然后用图像的分形维数划分子空间。

分形维数计算方法采用双层地毯法(double blanket method, DBM)^[17]。其基本算法是:假定图像中 (X, Y) 像素点的灰度值构成一个表面,分别在该表面的上部和下部与该表面相距某一尺度处做两个“毯”,将该表面包起来。计算上下“毯”所包围的体积,然后获得“双毯”的表面积。步长与面积做对数回归得到回归直线斜率 K ; 该图像的分形维数为 $D = 2.0 -$

K , 即得该幅遥感图像的分形维数。然后运用最优波段系数进行处理。最优波段系数公式如下:

$$OBC_i = \sigma_i / R_{s_i} \tag{6}$$

其中: OBC_i 为第 i 波段的最优波段系数; σ_i 为第 i 波段的标准差; R_{s_i} 为第 i 波段与其余波段的相关系数的和。计算每个波段最优波段系数的值,并从大到小排序,根据用户的要求即可输出最优波段。

3 实验研究

3.1 实验数据

实验影像数据采用中国科学院上海技术物理研究所研制的实用型模块化成像光谱仪(operational modular imaging spectrometer, OMIS)所获取的 64 波段 536×512 数据,其光谱范围为 $460 \sim 1\,100\text{ nm}$, 波段宽为 10 nm 。研究区经纬度: $40.262\,377\text{ N}, 116.258\,722\text{ E}$ 。图 2 为研究区的高光谱影像数据立方体(由 12、24、20 波段假彩色合成)及研究区某地物光谱曲线图。

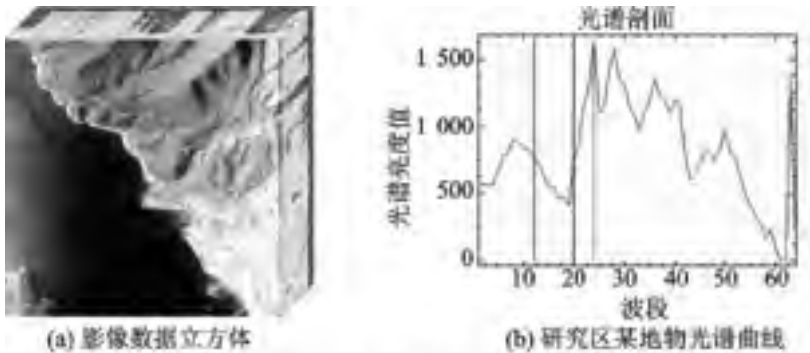


图2 研究区高光谱数据立方体(OMIS数据)

为了与传统的波段选择方法(利用信息统计量均值、方差等信息)对比,实验中计算了各个波段亮度的均值和标准差。均值反映了波段图像的亮度信息;标准差则在一定程度上反映了相应波段所含信息量的大小,如图 3 所示。其中:曲线①为光谱亮度值均值;曲线②为光谱亮度值均方差。

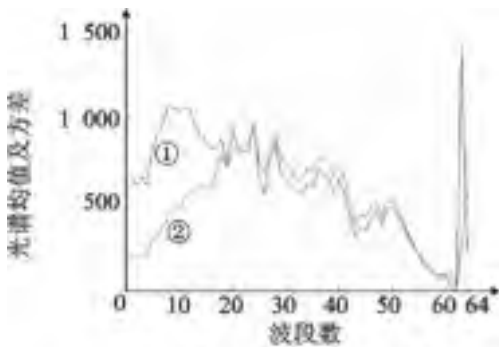


图3 波段亮度的均值和标准差

3.2 三波段组合结果分析

需要说明的是,以上算法中,熵及联合熵算法、最佳指数因子、自动子空间划分均为三波段组合选择算法;另外,自适应波段选择既可以选择三波段组合也可以选择任意 N 个波段。

以下是以各种算法选择的三波段组合。其中只选取相关指数值排在前 10 位的结果。

由表 1 可以得出以下结果。联合熵波段选择的结果,使波段容易聚集在某一连续的波段间,如明显集中在 $35 \sim 45$ 。由于相邻波段其相关性大,造成冗余信息多,效果不理想。

最佳指数因子波段选择的结果是,所选波段间的距离仍然较近,波段间的相关性较强,组成图像的视觉效果比用其他方

法所求的组合波段的合成图像的视觉效果差。

表 1 各算法三波段组合选择结果(部分)

排序	JE	OIF	ASP	ABS
1	45, 42, 43	18, 24, 28	20, 24, 12	28, 1, 3
2	44, 42, 43	18, 24, 27	24, 63, 12	28, 1, 2
3	45, 43, 44	18, 27, 28	20, 63, 12	28, 2, 3
4	46, 42, 43	18, 23, 24	24, 63, 20	28, 1, 4
5	43, 39, 42	17, 24, 28	—	28, 3, 4
6	46, 43, 45	18, 24, 29	—	28, 2, 4
7	43, 41, 42	18, 24, 37	—	24, 1, 3
8	45, 42, 44	6, 24, 28	—	24, 1, 2
9	45, 39, 43	9, 24, 28	—	24, 2, 3
10	43, 26, 42	18, 23, 28	—	24, 1, 4

采用子空间划分方法,使选择的波段覆盖的波段范围大为扩展。例如本次实验结果基本上覆盖了 12 ~63 的所有波段范围。可以看出,对波段进行子空间划分后,由于冗余波段减少,选出的波段即是那些信息量大、相关性小,具有明显区分度的最佳波段组合。

自适应波段选择方法充分考虑了各波段的空间相关性和谱间相关性及每幅图像的信息富集程度与相邻波段的相似性。实验结果表明,自适应波段选择方法能够选择出所需的波段,该方法有助于减少后续处理的计算量并能提高分类效果。

3.3 N 波段结果分析

有时由于应用的目的不同,需要选择高光谱的前 N 个波段。本文据此研究了相关的算法。自适应波段选择、波段系数和最优波段指数三种算法就是选择任意 N 个波段的算法。

以下是以各个算法选择的前 N 个波段的结果(由于篇幅原因,此处选择 $N=10$)。表 2 列出的分别是自适应波段选择、波段指数和最优波段指数选择出来的前 10 个波段的排序结果。

表 2 各算法前 N 波段选择结果(部分)

排序	ABS	BI	OBI	排序	ABS	BI	OBI
1	24	21	24	6	20	15	22
2	28	20	28	7	21	14	20
3	23	18	27	8	18	16	21
4	27	19	23	9	29	13	29
5	22	17	18	10	37	24	37

由结果可以得出以下结论: ABS 和 OBI 两种算法所得波段基本相同,如前 10 个波段完全相同,前 20 个波段有 19 个是相同的,只是排序有稍微的不同;而 BI 与 ABS、OBC 相比,前 10 个波段中也有 4 个波段是相同的,而前 20 个波段相同的个数也分别达到了 15、16 个。说明了这些算法的效果还是比较一致的。同时,根据波段亮度的均值和标准差图像可知,在前 5 个波段中 ABS 算法所选波段的标准差最大为 947.066、最小为 843.136; OBI 算法所选波段的标准差最大为 947.066、最小为 794.514; 而 BI 算法所选波段的标准差最大只有 812.619、最小仅为 703.881。这也说明该算法在选择高光谱波段方面是可行的;同时也说明以上算法在选择高光谱波段方面是可行的。

4 结束语

高光谱遥感经过近几年的发展,出现了多种针对高光谱特

点的数据处理算法和技术。本文对这些算法进行了归纳分析,并用实验分析了各种算法的性能和优缺点。分析了各种算法的有效性、局限性和计算复杂度,并针对波段指数的不足,设计了最优波段指数波段选择新算法。最后通过具体的实验,验证了各种算法的性能。

总之,虽然波段选择算法众多,但面向应用时,波段选择的效果要以地物识别分类达到最佳效果为目标。因此,笔者下一步的工作将重点研究有针对性、目的性和面向应用目标的波段选择算法。

参考文献:

[1] 张良培,张立福.高光谱遥感[M].武汉:武汉大学出版社,2005.

[2] 赵英时.遥感应用分析原理与方法[M].北京:科学出版社,2003.

[3] 蒲瑞良,宫鹏.高光谱遥感及其应用[M].北京:高等教育出版社,2000.

[4] 边肇祺,张学工.模式识别[M].北京:清华大学出版社,2000.

[5] 刘春红,赵春晖,张凌雁.一种新的高光谱遥感图像降维方法[J].中国图象图形学报,2005,10(2):218-222.

[6] 陈桂红,唐伶俐,姜小光.高光谱遥感图像特征选择和提取方法的比较[J].干旱区地理,2006,29(1):143-149.

[7] ARZUAGA-CRUZ E, JIMENEZ-RODRIGUEZ L O, V LEZ-REYES M. Unsupervised feature extraction and band subset selection techniques based on relative entropy criteria for hyperspectral data analysis [C] //Proc of SPIE Algorithm and Technologies for Multispectral, Hyperspectral, and Ultrospectral Imagery. 2003:462-473.

[8] PASKALEVA B, HAYAT M. Optimized algorithm for spectral band selection for rock-type classification[C] //Proc of SPIE Pefense and Security Conference, Algorithm and Technologies for Multispectral, Hgperspectral, and Ultrospectral Imagery. Orlando: [s. n.], 2005:131-138.

[9] WANG Su, CHANG C I. Linearly constrained band selection for hyperspectral imagery[C] //Proc of SPIE Algorithm and Technologies for Multispectral, Hyperspectral, and Ultrospectral Imagery. 2006.

[10] BAJCSY P, GROVES P. Methodology for hyperspectral band selection[J]. Photogrammetric Engineering and Remote Sensing Journal, 2004, 70(7):793-802.

[11] 赵春晖,刘春红.超谱遥感图像降维方法研究现状与分析[J].中国空间科学技术,2004,5:28-36.

[12] 姜小光,王长耀,王成.成像光谱数据的光谱信息特点及最佳波段选择——以北京顺义区为例[J].干旱区地理,2000,23(3):214-220.

[13] 陈蜜,易尧华,刘志刚,等.基于分块特性的高光谱影像波段选取方法的研究[J].测绘通报,2006,3:10-13,21.

[14] CHACVEZ P S, BERLIN G L, SOWERS L B. Statistical method for selecting landsat MSS ratios [J]. Journal of Applied Photographic Engineering, 1982, 1(8):23-30.

[15] GU Yan-feng, ZHANG Ye. Unsupervised subspace linear spectral mixture analysis for hyperspectral images[C] //Proc of International Conference on Image Processing. 2003:801-804.

[16] 王立国,谷延锋,张晔.基于支持向量机和子空间划分的波段选择方法[J].系统工程与电子技术,2005,27(6):974-977.

[17] 张华国,黄韦良,周长宝,等.关于 IKONOS 卫星遥感图像的分形特征研究[J].测绘通报,2005,5:15-18.