

基于最佳波段指数和 J-M 距离可分性的高光谱数据最佳波段组合选取研究

——以环境小卫星高光谱数据在东莞市的应用为例

马娜^{1,2}, 胡云锋¹, 庄大方¹, 王新生²

(1. 中国科学院地理科学与资源研究所, 北京 100101; 2. 湖北大学资源环境学院, 湖北 武汉 430062)

摘要: 高光谱数据普遍存在波段相关性强、数据冗余严重的特点, 因此选择合适的波段组合, 是高效开展后续应用研究的基础。以东莞市为研究区, 应用环境与灾害监测预报小卫星(HJ-1A)高光谱数据, 在分析各波段信息含量和波段间相关性的基础上, 使用了3种经典的波段指数选择最佳波段组合; 针对经典模型应用中存在的问题, 对最佳指数模型进行了改进, 通过对波段均方差和相关性设置一定阈值, 筛选得到一个较为合理的波段组合; 最后, 针对草地、林地和耕地3种地物, 应用J-M距离模型对3种地物的可分性进行判别, 并指出: 50-80-108波段组合, 50-79-108波段组合以及50-80-111波段组合是分别用于草地—林地、草地—耕地、耕地—林地分类的最佳波段组合。

关键词: 环境与灾害监测预报小卫星; 高光谱; 最佳波段组合; 可分性

中图分类号: TP 75 **文献标志码:** A **文章编号:** 1004-0323(2010)03-0358-08

1 引言

高光谱遥感(Hyperspectral Remote Sensing)是以纳米尺度对目标进行连续光谱成像的新技术, 它使本来在宽波段遥感中不可探测的物质, 在高光谱遥感中能被探测^[1]。与传统多光谱遥感数据相比, 高光谱数据具有以下明显特点: 光谱分辨率高, 高光谱传感器可以获取从可见光到热红外的连续波段; 光谱范围广、波段窄、对地物可以连续成像; 信噪比普遍较低, 噪声增加, 增加了数据处理的难度^[2]。

应用传统遥感数据处理方法处理高光谱数据时, 经常遇到 Hughes 现象, 即“维数祸根”^[3-4]。高光谱数据中成百个波段构成的海量数据对后续数据处理造成了巨大压力, 主要表现在: ① 海量的高光谱数据需要更大的存储空间和更多的运算时间, 不仅要求有更高性能的计算机软、硬件配置, 并且常常导致操作难度增大^[5]。② 波段数成10倍的增加并不意味可利用信息亦等比例增加; 由于波段间相关

性高, 常常导致高光谱数据冗余严重, 信息密度降低。因此, 选择合适的波段组合, 是高效开展后续应用研究的基础和前提。

东莞市位于广东省中南部, 珠江口东岸, 东江下游的珠江三角洲。地处 $113^{\circ}39' \sim 113^{\circ}51' E$, $22^{\circ}55' \sim 23^{\circ}6' N$ (如图1)。处于亚热带, 受季风海洋性气候影响明显, 具有丰富的水热资源, 极有利于植物多样性的发展。主要植被类型为常绿针叶林、针阔叶混交林、典型常绿阔叶林、季风常绿阔叶林等。地带性土壤类型为赤红壤。四季常绿植被覆盖极有利于遥感解译工作。

2 基础数据和方法

环境与灾害监测预报小卫星星座 A、B 星(HJ-1A/B星)于2008年9月6日成功发射。HJ-1A星搭载了 CCD 相机和超光谱成像仪(HSI), HJ-1B星搭载了 CCD 相机和红外相机(IRS)。超光谱成像仪(HIS)所形成图像幅宽大于50km、地面像元分

收稿日期: 2009-12-16; 修订日期: 2010-05-18

基金项目: 科技部科技支撑计划(2006BAC08B05), 国家自然科学基金项目(40701150, 40971223), 瑞典科学基金亚洲联系项目, 中国科学院地理科学与资源研究所青年人才领域前沿项目(O7H70163AR)。

作者简介: 马娜(1984—), 女, 硕士研究生, 研究方向为自然资源与土壤遥感。E-mail: mana.hailang@163.com。

作者简介: 胡云锋(1974—), 副研究员, 主要从事资源环境遥感与地理信息技术应用研究。E-mail: huyf@lreis.ac.cn。

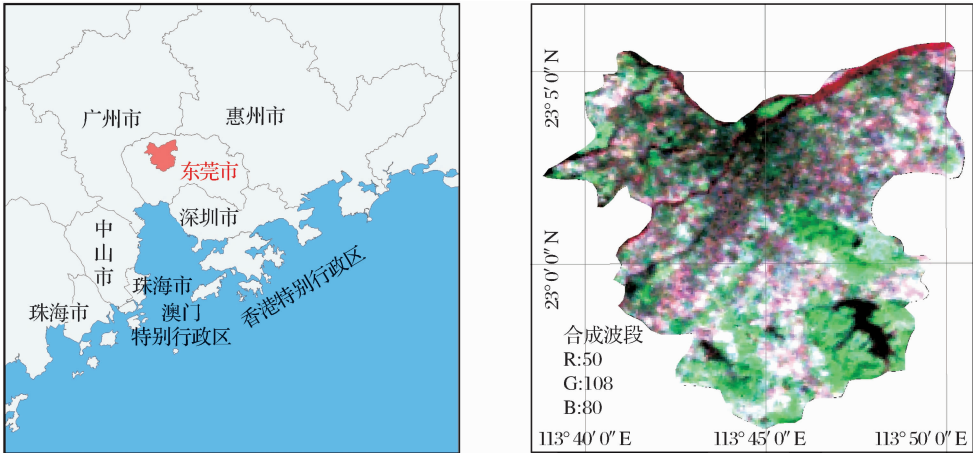


图 1 研究区位置以及高光谱合成影像图 (RGB=50,108,80)

Fig. 1 The position and hyspectral image of the study area

分辨率为 100 m,具有±30°侧视能力和星上定标功能。工作谱段为 0.45~0.95 μm,谱段数为 110~128,平均光谱分辨率为 5 nm。

本实验所使用的高光谱数据于 2008 年 12 月 25 日获得,光谱景序列号 44656,产品序列号 48401,共 115 个谱段,各波段增益 0.7。由于 1 到 20 以及 113、114 和 115 波段是坏数据,所以实际参与分析的波段为 92 个。其中 21 到 87 波段为可见光部分,90 到 112 为近红外部分。数据首先经过对数残差大气纠正,形成用于后续研究的基础数据,其值域在 0~1 之间。

得到基础数据后,即可开展最佳波段组合研究。选择最佳波段组合,主要需要考虑 3 个因素:第一,波段信息量多少;第二,波段间相关性强弱;第三,研究地物光谱特点。波段自身信息量丰富、波段间相关程度小、波段组合对目标地物的光谱反映差异大、可分性好;这样的波段组合就是最佳波段组合^[6]。

为此,对波段信息量度量的指标有信息熵^[7-8]、均方差^[7,9]等指标,对波段相关性强弱度量的指标则为相关系数指标。在对上述 2 类指标进行分析的基础上,需要结合一定的模型和算法,得到信息量高、相干性低的波段组合。在这方面,经典方法很多,如波段指数^[6]、最佳指数(OIF; Optimum Index Factor)^[10-11]、主成分分析(PCA; Principal Component Analysis)^[12-13]、最小噪声分离(MNF; Minimum Noise Fraction)^[1,14]、自适应波段选择(ABS; Adaptive Band Selection)^[15]和联合熵^[7]判别。然而,使用上述方法得到的波段组合仅仅是总体上较好的波段组合;针对特定地物特征,还需要使用一定方法来判定波段组合对地物的区分能力。在这方面,经典的方法有如:J-M

距离^[16-17]、B 距离(Bhattacharyya Distance)^[18-19],离散度(Divergence)^[5,20]等。

3 最佳波段选取及可分性分析

3.1 波段信息量分析

从信息论角度来看,信息熵是评价系统信息有序程度和效用的指标,但是信息熵模型的运算通常具有很高的时空复杂度,数据量大或数据量化级别高都可导致其不可计算^[21]。相比之下,统计学中的均方差反映为图像各像元灰度值与平均值之间的离散度^[6],其算法相对简单,运算的时空效率较高;更重要的是,均方差方法与信息熵方法得到的波段信息量评价结果具有较好的一致性^[9]。方差大,则图像灰度级分布分散,指示图像的反差大,因此可以得到更多信息;方差小,则图像反差小、对比度不大,色调单一均匀,因此得不到太多信息。因此,本文采用了波段均方差指标衡量波段信息量大小。图 2 显示

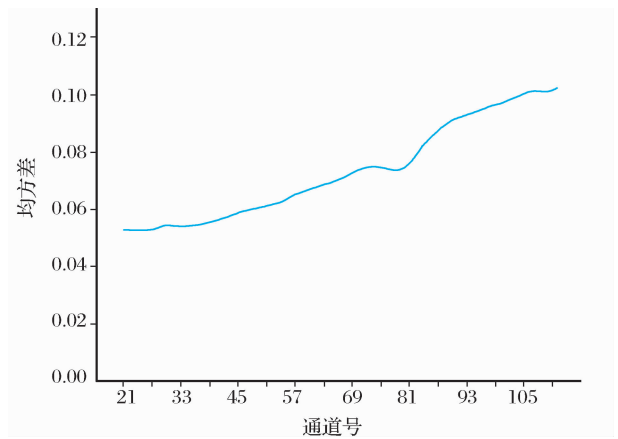


图 2 研究区 HJ-1A 高光谱数据波段均方差变化

Fig. 2 The mean square deviation variation trend of the HJ-1A hyperspectral data of the study area

了东莞市环境与灾害监测预报小卫星高光谱数据各波段的均方差变化趋势。

总体上,从可见光部分到近红外部分,各波段的均方差呈稳定上升趋势,仅在局部(波段 74~80)则出现了下降现象。深入的分析表明,均方差最大的 10 个波段依次为:112、111、107、108、110、109、106、105、104 和 103,均方差最小的 10 个波段依次为:

23、24、22、21、25、26、27、28、33 和 32。其中,波段 112 的均方差最大,指示波段 112 图像反差最大,所蕴含信息量可能较多;波段 23 的均方差最小,指示波段 23 图像反差最小,所蕴含信息量可能最少。对上述两个波段进行制图(图 3),也可以清楚发现波段 23 噪声影响严重,图像模糊,信息负载量低;波段 112 图像清晰,对比度好,信息负载量大。

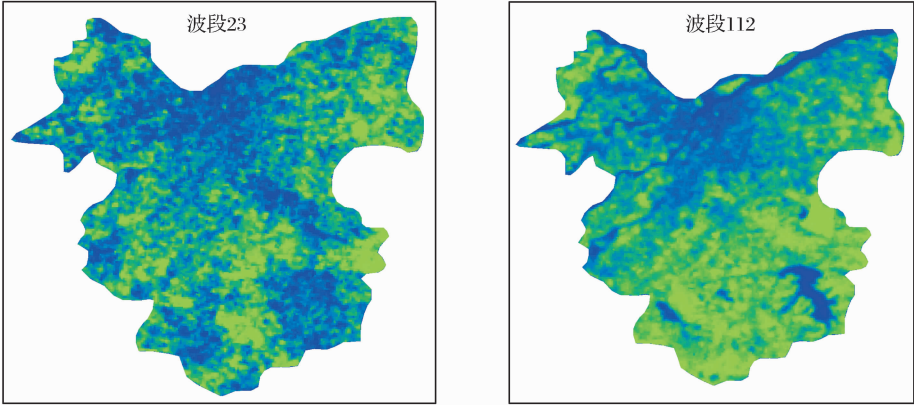


图 3 波段 23 和波段 113 的彩色分割图

Fig. 3 Color segmentation map of band 23 and band 113

3.2 波段相关性分析

由图 1 可见,107、111 和 112 等 3 个波段信息量最大。但是,由于这 3 个波段邻近,波段间的相关性极强,因此这 3 个波段组合后所包含的信息量并不一定最大。因此,在选择最佳波段组合时,必须进一步分析波段相关性的影响。为此,可以利用相关系数分析波段间信息的重叠程度。相关系数越大,表明数据相似,信息重叠度高;反之,相关系数越小,表明数据冗余度低,信息重叠程度越小。图 3 即显示了高光谱数据波段间相关性的变化。

图 4 显示:① 波段 21 到 57 区段(即蓝绿光波段)、波段 58 到 87 区段(即红光波段)、以及波段 88 到 112 区段(即近红外波段)等 3 个区段内部,波段之间的相关性极高。各波段之间的相关系数都在 0.91 以上,尤其是近红外区段内部,各波段之间的相关性最高,相关性均值达到 0.98;部分相邻波段相关系数高达 0.99(如波段 89 与波段 90,波段 89 与波段 88);② 上述 3 个区段之间,波段之间的相关系数明显低于区段内部各波段之间的相关系数。蓝绿波段与红光波段和近红外波段的相关系数均值分别为 0.86 和 0.59,红光与近红外的相关系数均值为 0.71。其中,波段 107 和 23 的相关性最低,为 0.529;93 和 92 的相关系数最高,为 0.997。

依据图3的上述分析表明,HJ-1小卫星高光谱

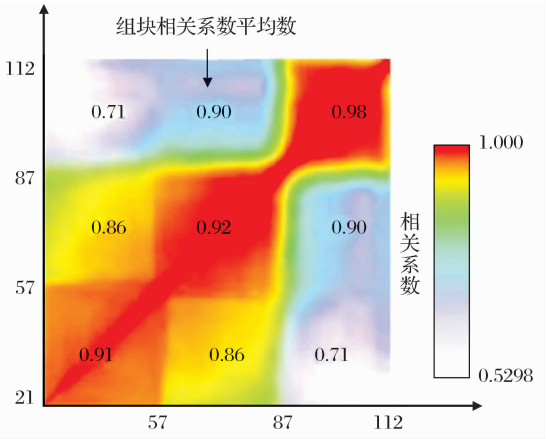


图 4 波段间相关系数矩阵图

Fig. 4 Band correlation coefficient matrix

数据存在明显的分组成块现象。根据相关系数,波段 21~57 构成蓝绿波段组块,波段 58~87 构成红光波段组块,波段 88~112 构成近红外波段组块。各组块内部相关程度高,信息重叠程度大,数据冗余程度高;组块之间相关程度低,信息重叠程度低,数据冗余程度低。由波段间相关性所指示出的数据分组成块现象也表明:在后续的研究中,只有从 3 个区段各选取适当波段进行组合,才能保证组合后的波段彼此之间保持为最独立。

3.3 最佳波段指数分析

由于反映数据离散程度的均方差可以指示波段

信息量,而反映波段间相关程度的相关系数可以指示波段的独立性。因此,按照一定的方法将两者综合起来,就形成了不同的波段指数方法。由此得到的最优波段可以兼顾波段的信息量和独立性,最终得到一个信息量大、独立性好的波段。

文献[15]提出了一种基于自适应波段选择 (ABS: Adaptive Band Selection) 的最佳波段选择方法。该方法计算本波段均方差除以本波段与前后相邻两个波段的相关系数的均值的商。公式如下:

$$I_i = \frac{\sigma_i}{(R_{i-1,i} + R_{i,i+1})/2}$$

(1)

其中: I_i 是波段 i 的 ABS 指数大小, σ_i 表示波段 i 的均方差, $R_{i-1,i}$ 和 $R_{i,i+1}$ 分别表示波段 i 与其前后两个波段的相关系数。 I_i 越大,说明波段的信息量越大,独立性越强,波段就越具有代表性。最终确定最佳波段选择的方法有两种,一种是根据确定的阈值,当 I_i 大于这个阈值时,即为被选入的波段;另一种是取排名在前面的 n 个波段。

从图 5 中可以清楚看出,各波段 ABS 指数变化趋势与它们的均方差变化趋势非常相似。总体上,从可见光部分到近红外部分,ABS 指数呈稳定上升趋势,仅在局部(波段 74~80)出现了下降现象。按照 ABS 的选择方法,取排名前 3 位是 110、111 和 112。但事实上,这 3 个波段为相邻波段,相关性极高,因此由此组成的波段组合显然不合理。究其原因,是因为 HJ-1A 高光谱数据波段相关系数总体变化不大,ABS 数模型中的 $(R_{i-1,i} + R_{i,i+1})/2$ 基本稳定不变,因此 ABS 指数主要受波段均方差所控制。由此可见,对于 HJ-1A 高光谱数据,ABS 指数模型方法并不适用。

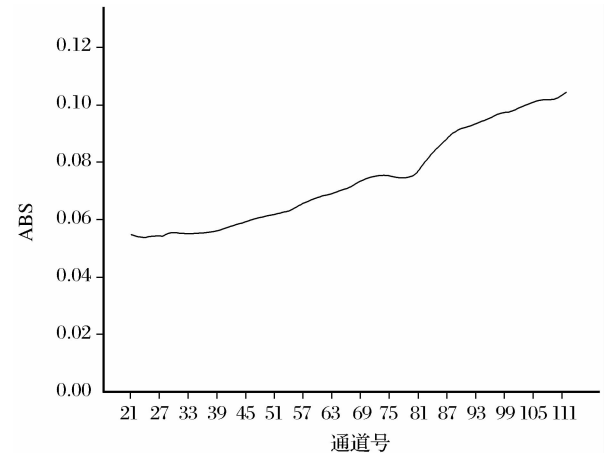


图 5 研究区 HJ-1A 高光谱数据自适应波段指数变化
Fig. 5 The adaptive band selection variation trend of the HJ-1A hyperspectral data of the study area

与自适应波段指数模型相类似的模型还有波段指数模型^[6]。该模型考虑了波段可能集群成块的特点。波段指数模型如下:

$$P_i = \frac{\delta_i}{R_i}, (R_i = R_w + R_a)$$

(2)

其中: δ_i 为波段 i 的均方差, R_w 为波段 i 与所在组内其它波段相关系数的绝对值之和的平均值, R_a 为第 i 波段与所在组以外的其它波段之间的相关系数的绝对值之和。最终选择各个区段内波段指数最高的波段,由此形成最佳波段组合。在本研究中,根据上一节中有关波段间相关系数计算成果,将 HJ-1A 高光谱数据分成 3 个组,即蓝绿光波段(波段 21 ~ 57)、红光波段(波段 58 ~ 87) 以及近红外波段(波段 88 ~ 112),并计算其波段指数。

图 6 显示了研究区高光谱数据波段指数的变化。与 ABS 指数相比,波段指数在不同光谱区段显示了较好的分离性。从可见光部分到近红外部分,波段指数总体呈上升趋势,但在局部则出现陡降现象,这种陡降主要发生在不同波段群组(蓝绿光、红光、近红外)分界处。3 个光谱区段中,波段指数最高的分别是波段 57、波段 87 和波段 108。然而,对 3 个波段之间相关性的分析表明,波段 87 与波段 108 相关系数依然极高(0.97)。分析表明,这是因为 HJ-1A 高光谱数据波段相关系数总体变化不大,波段指数模型中的 R_w 基本相同,波段指数主要受波段均方差所控制。因此,对于 HJ-1A 高光谱数据,波段指数模型方法并不适用。

以上两种方法仅仅是分析每一个波段的信息量及其独立性的关系,由此计算得到各波段的“波段指数”,最终选择具有最大“波段指数”的几个波段组合

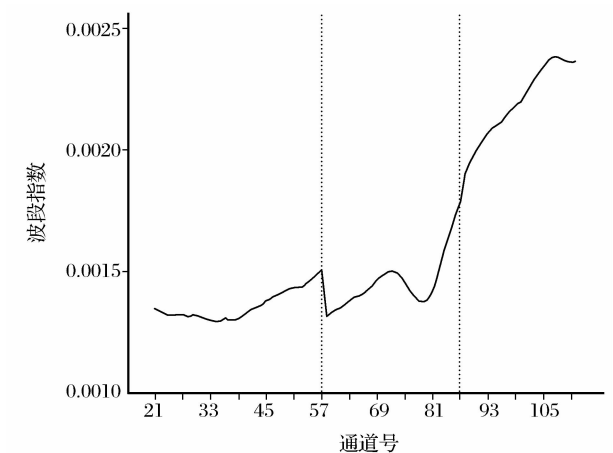


图 6 研究区 HJ-1A 高光谱数据波段指数变化
Fig. 6 The band index variation trend of the HJ-1A hyperspectral data of the study area

形成“最佳”波段。但是,这些方法在逻辑上是存在疑问的。即“信息量—独立性”最好的几个波段机械组合,并不一定代表波段组合整体的“信息量—独立性”最好。对此,Chavez^[11]提出了针对波段组合的最佳指数因子(OIF:Optimum Index Factor)方法,公式如下:

$$OIF = \sum_{i=1}^3 S_i / \sum_{j=1}^3 |r_{ij}| \tag{3}$$

其中: S_i 为波段 i 的标准差, r_{ij} 为波段 i 与波段 j 之间的相关系数。

OIF 最大的波段组合,就是整体最优的波段组合。对高光谱数据来说,其总波段数达到 100 余个,本研究中研究总波段数为 92 个;因此通过简单的排列组合可知,共需要计算 $92 \times 91 \times 90$ 次。由于数据量巨大,因此我们采用基于 C++ 编程技术,求出所有组合的 OIF,表 1 列出了 OIF 排名前 20 的波段组合。

表 1 最佳波段组合 OIF

Table 1 The OIF of the best band combination

| 序号 | 波段组合 | OIF | 序号 | 波段组合 | OIF |
|----|------------|----------|----|------------|----------|
| 1 | 71-108-112 | 0.131721 | 11 | 69-109-112 | 0.131314 |
| 2 | 71-107-112 | 0.13168 | 12 | 70-109-112 | 0.131304 |
| 3 | 69-108-112 | 0.131668 | 13 | 69-106-112 | 0.131296 |
| 4 | 70-108-112 | 0.131631 | 14 | 70-106-112 | 0.131267 |
| 5 | 69-107-112 | 0.131621 | 15 | 72-109-112 | 0.131235 |
| 6 | 70-107-112 | 0.131594 | 16 | 72-106-112 | 0.131181 |
| 7 | 72-108-112 | 0.131538 | 17 | 68-108-112 | 0.131101 |
| 8 | 72-107-112 | 0.131707 | 18 | 68-107-112 | 0.13108 |
| 9 | 71-109-112 | 0.131409 | 19 | 71-107-91 | 0.131006 |
| 10 | 71-106-112 | 0.131353 | 20 | 71-110-112 | 0.130986 |

然而进一步的分析表明,在表 1 中排名第一的 71-108-112 组合中,波段 108 与波段 112 的相关性很高,达到 0.969916。排名靠后的其它波段组合(如 71-110-112)也有类似问题。因此,由 OIF 方法得到的结果并不完全可用。由此可见,简单的 OIF 方法虽然开始从整体上考虑信息量与独立性之间的关系,但是没有对波段相关系数以及信息量做出必要的阈值限制,只是简单的进行加和统计。这种简单的加和统计会产生拉平效应,将一些信息量低、相关性高的波段纳入到最佳波段组合中来;因此,这种做法最终并不能得到整体信息含量高、内部各波段独立性强的波段组合。

为此,我们对 Chavez 提出最佳指数算法进行了改进,对各波段的信息量参数以及相关系数设置必

要的阈值。公式如下:

$$OIF = \sum_{i=1}^3 S_i / \sum_{j=1}^3 |r_{ij}|$$
$$S_i \geq S_0$$
$$|r_{ij}| < r_0$$

(4)

其中: S_i 为波段 i 的标准差, r_{ij} 为波段 i 与波段 j 之间的相关系数, S_0 是预设的方差阈值, r_0 是预设的相关系数阈值。

在本研究中,经反复试验,可设定如下阈值: $S_0 = 0.0542$, $r_0 = 0.87$ 。由此,筛选得到如表 2 的最佳波段指数排名。这些组合不但信息负载量大,而且相关性相对较小,所以可视为最佳波段组合。

3.4 地物的可分性分析

最佳波段组合不是绝对的。对于不同地表特征,最佳波段组合会发生相应变化。对于一些地物来说最好的波段组合,对于别的地物则可能是最糟糕波段组合。因此,在实践中,应当针对特定样本地

表 2 改进算法的最佳波组合 OIF

Table 2 Improved algorithm for the OIF of the best band combination

| 序号 | 波段组合 | OIF | 序号 | 波段组合 | OIF |
|----|-----------|----------|----|-----------|----------|
| 1 | 50-80-108 | 0.103834 | 11 | 49-80-112 | 0.103326 |
| 2 | 50-80-107 | 0.103803 | 12 | 49-80-110 | 0.103271 |
| 3 | 50-80-109 | 0.103648 | 13 | 48-80-108 | 0.103212 |
| 4 | 49-80-108 | 0.103609 | 14 | 49-80-106 | 0.103205 |
| 5 | 50-80-112 | 0.103608 | 15 | 49-80-101 | 0.103196 |
| 6 | 49-80-107 | 0.103572 | 16 | 48-80-107 | 0.10315 |
| 7 | 50-80-110 | 0.103536 | 17 | 50-79-108 | 0.10308 |
| 8 | 50-80-111 | 0.103487 | 18 | 50-79-107 | 0.103052 |
| 9 | 50-80-106 | 0.103443 | 19 | 48-80-109 | 0.10301 |
| 10 | 49-80-109 | 0.103412 | 20 | 47-80-108 | 0.103005 |

物,来研究选择出最佳波段组合。本文中以耕地、林地和草地为例,筛选出对 3 种地物具有最好分离性质的波段组合。

地表特征可分性的判定方法很多,如 J-M 距离(Jeffreys-Matusita Distance)、B 距离(Bhattacharyya Distance)、离散度、样本间平均距离、类别间相对距离等。相对于其它指标,J-M 距离被认为更适合于表达类别可分性^[17]。J-M 距离是基于条件概率理论的光谱可分性指标,公式如下^[1]:

$$J_{ij} = \left\{ \int_x [\sqrt{p(X/\omega_i)} - \sqrt{p(X/\omega_j)}]^2 dX \right\}^{1/2} \tag{5}$$

其中: $p(X/\omega_i)$ 为条件概率密度,即第 i 个像元属于第 ω_i 个类别的几率。 J_{ij} 的值在 $0 \sim 2$ 之间,其大小代表样本间可分离程度的高度。 J_{ij} 位于 $0.0 \sim 1.0$ 之

间时,样本间不具有光谱可分性; J_{ij} 位于 1.0 ~ 1.9 之间时,样本间具有一定的光谱可分性,但存在较大程度的重叠; J_{ij} 位于 1.9 ~ 2.0 之间时,样本间具有很好的光谱可分性^[22]。

表 3 列出了各组合 J-M 距离排序。可以看出,所有波段组合对于林地和耕地的可分性都比较好,J-M

距离都在 1.99 以上;相对地,各波段组合对草地和林地的可分性较差一些,但也保持在 1.88 以上,即具有一定的光谱可分性,但存在一定程度的重叠。总的来看,50-80-108 波段组合,50-79-108 波段组合以及 50-80-111 波段组合是分别用于草地—林地、草地—耕地、耕地—林地分类的最佳波段组合。

表 3 不同波段组合对耕地、林地和草地的 Jeffreys-Matusita 距离
Table 3 Diferent band combination's Jeffreys-Matusita distance of farmland, woodland and grassland

| 序号 | 波段组合 | 草地—林地 | 序号 | 波段组合 | 草地—耕地 | 序号 | 波段组合 | 林地—耕地 |
|----|-----------|----------|----|-----------|----------|----|-----------|----------|
| 1 | 50-80-108 | 1.903240 | 1 | 50-79-108 | 1.978425 | 1 | 50-80-111 | 1.998974 |
| 2 | 50-79-108 | 1.903129 | 2 | 50-80-108 | 1.978033 | 2 | 50-80-106 | 1.998623 |
| 3 | 50-80-112 | 1.901026 | 3 | 50-80-107 | 1.975981 | 3 | 49-80-106 | 1.998434 |
| 4 | 50-80-111 | 1.899607 | 4 | 50-79-107 | 1.975924 | 4 | 50-80-107 | 1.997541 |
| 5 | 50-80-107 | 1.899464 | 5 | 50-80-106 | 1.970940 | 5 | 50-80-112 | 1.997505 |
| 6 | 50-80-110 | 1.899243 | 6 | 49-80-108 | 1.969346 | 6 | 50-80-109 | 1.997304 |
| 7 | 50-80-106 | 1.898559 | 7 | 49-80-107 | 1.966650 | 7 | 49-80-107 | 1.997173 |
| 8 | 50-79-107 | 1.898430 | 8 | 47-80-108 | 1.966063 | 8 | 49-80-109 | 1.997034 |
| 9 | 50-80-109 | 1.893820 | 9 | 48-80-108 | 1.965834 | 9 | 49-80-112 | 1.997022 |
| 10 | 49-80-108 | 1.888022 | 10 | 50-80-109 | 1.964668 | 10 | 49-80-101 | 1.996952 |
| 11 | 47-80-108 | 1.887979 | 11 | 48-80-107 | 1.962494 | 11 | 50-80-108 | 1.996496 |
| 12 | 48-80-108 | 1.886170 | 12 | 50-80-111 | 1.961706 | 12 | 50-79-107 | 1.996364 |
| 13 | 49-80-110 | 1.884571 | 13 | 49-80-106 | 1.959208 | 13 | 50-80-110 | 1.996351 |
| 14 | 49-80-112 | 1.884077 | 14 | 50-80-110 | 1.955338 | 14 | 49-80-110 | 1.996127 |
| 15 | 49-80-107 | 1.883141 | 15 | 49-80-109 | 1.953790 | 15 | 49-80-108 | 1.996038 |
| 16 | 49-80-106 | 1.881768 | 16 | 48-80-109 | 1.949561 | 16 | 48-80-107 | 1.995863 |
| 17 | 48-80-107 | 1.881089 | 17 | 50-80-112 | 1.944848 | 17 | 50-79-108 | 1.995753 |
| 18 | 49-80-109 | 1.876472 | 18 | 49-80-110 | 1.943469 | 18 | 48-80-109 | 1.995719 |
| 19 | 48-80-109 | 1.874094 | 19 | 49-80-112 | 1.922755 | 19 | 48-80-108 | 1.994111 |
| 20 | 49-80-101 | 1.867075 | 20 | 49-80-101 | 1.876746 | 20 | 47-80-108 | 1.993404 |

将表 2 和表 3 对比可以发现,OIF 指数大的波段组合与可分性最好的波段组合之间并没有严格的对应关系:OIF 指数大的波段组合并不表示可分性最好,可分性最好的波段组合也不一定就是 OIF 指数大的波段组合。50-80-108 波段组合的 OIF 指数最大,但该组合只是针对林地—草地的区分表现最好,对于草地—耕地、耕地—林地的区分则稍差,其 J-M 距离排名中分别为第 2 位和第 11 位。反过来,对草地—耕地、耕地—林地的区分最好的波段组合(即 50-79-108 波段组合以及 50-80-111 波段组合),它们在 OIF 中的排名则靠后,分别为第 17 位和第 8 位。

4 结 语

本文以东莞市为研究区,应用环境与灾害监测预报小卫星高光谱数据,先后使用了方差分析、相关系数分析、多种波段指数分析以及可分性指数分析

等手段,针对研究区的耕地、林地和草地特征识别进行了研究。研究可得到如下结论:

(1)对于 HJ-1A 高光谱数据,从可见光部分到近红外部分,各波段的均方差呈稳定上升趋势,仅在局部(波段 74~80)出现了下降现象。这也即表明,从可见光部分到近红外部分,各波段的信息量呈稳定上升趋势。波段 112 的均方差最大,指示波段 112 图像反差最大,所蕴含信息量可能较多;波段 23 的均方差最小,指示波段 23 图像反差最小,所蕴含信息量可能最少。

(2)HJ-1 小卫星高光谱数据存在明显的分组成块现象。波段 21~57 构成蓝绿波段组块,波段 58~87 构成红光波段组块,波段 88~112 构成近红外波段组块。各组块内部相关程度高,信息重叠程度大,数据冗余程度高;组块之间相关程度低,信息重叠程度低,数据冗余程度低。HJ-1A 高光谱数据分

组成块的现象表明,在后续的研究中,只有从3个区段各选取适当波段进行组合,才能保证组合后的波段彼此之间保持为最独立。

(3)通过使用3种经典的波段指数选择最佳波段组合的研究指示:简单地将信息量与独立性指标(即均方差和相关系数)进行复合计算,并不能得到符合期望的波段组合。为此,作者针对经典模型应用中存在的问题,基于最佳指数模型,通过对模型中的波段均方差和相关性设置一定阈值,可以筛选得到一个较为合理的波段组合。

(4)作者以耕地、林地和草地特征区分为例,运用J-M距离判别方法,研究了对草地—林地、草地—耕地、林地—耕地地区分性能最好的波段组合。研究指出:50-80-108波段组合、50-79-108波段组合以及50-80-111波段组合是分别用于草地—林地、草地—耕地、耕地—林地分类的最佳波段组合。但是,通过对比J-M距离判别法和改进的OIF指数方法可以发现,两者得到的最佳波段组合并没有严格的对应关系。OIF指数大的波段组合并不表示可分性最好,可分性最好的波段组合也不一定就是OIF指数大的波段组合。

本文主要基于最佳波段指数模型和J-M距离模型,对HJ-1A高光谱数据最佳波段组合进行了初步的研究。虽然得到了一些初步的结论,但是仍有众多需要在未来研究中改进的地方,主要有:波段信息量和独立性的度量指标,最佳波段组合模型的改进,波段变换模型的引入,以及应用其它自动或人工方法对特定地物进行分类,并基于分类结果对上述方法进行验证等等。

参考文献:

- [1] Tong Qingxi, Zhang Bing, Zheng Lanfen. Hyperspectral Remote Sensing[M]. Beijing: Higher Education Press, 2006. [童庆禧, 张兵, 郑兰芬. 高光谱遥感——原理、技术与应用[M]. 北京: 高等教育出版社, 2006.]
- [2] Su Hongjun, Du Peijun. Study on Feature Selection and Extraction of Hyperspectral Data[J]. Remote Sensing Technology and Application, 2006, 21(4): 288-293. [苏红军, 杜培军. 高光谱数据特征选择与特征提取研究[J]. 遥感技术与应用, 2006, 21(4): 288-293.]
- [3] Hughes G F. On the Mean Accuracy of Statistical Pattern Recognizers[J]. IEEE Transactions on Information Theory, 1968, 14(1): 55-63.
- [4] Li Xing. Study on Band Selection for Hyperspectral Image of Vegetable[D]. Shandong University of Science and University, 2006. [李行. 植被高光谱遥感影像特征波段的选择方法研究[D]. 山东科技大学, 2006.]
- [5] Chen Guihong, Tang Liling, Jiang Xiaoguang. Feature Selection and Extraction of Hyperspectral Data——Based on HyMap Data of Barrax in Spanish[J]. Arid Land Geography, 2006, 29(1): 143-149. [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——基于试验区 Barrax 的 HyMap 数据[J]. 干旱区地理, 2006, 29(1): 143-149.]
- [6] Jiang Xiaoguang, Tang Liling, Wang Changyao, et al. Spectral Characteristics and Feature Selection of Hyperspectral Remote Sensing Data——Taking Shunyi Region of Beijing as a Study Area[J]. Remote Sensing Technology and Application, 2002, 17(2): 59-65. [姜小光, 唐伶俐, 王长耀, 等. 高光谱数据的光谱信息特点及面向对象的特征参数选择——以北京顺义区为例[J]. 遥感技术与应用, 2002, 17(2): 59-65.]
- [7] Xu Han, Yan Qin, Xu Panlin, et al. The Best Band Selection and Quality Assessment of Multi-source Remote Sensing Image Fusion[J]. Science of Surveying and Mapping, 2007, 32(3): 72-74. [许茜, 燕琴, 徐泮林, 等. 多源遥感影像融合最佳波段选择及质量评价研究[J]. 测绘科学, 2007, 32(3): 72-74.]
- [8] Shannon C E. A Mathematical Theory of Communication[J]. Bell System Technical Journal, 1948, 27(3): 379-423.
- [9] Li Xiehui, Zhen Yi. Study on the Extraction of Remote Sensing Information of Land Use/Cover Based on Optimum Band Combinatin[J]. Journal of Anhui Agriculture Science, 2009, 37(14): 6696-6699. [李谢辉, 郑奕. 基于最优波段组合的土地利用/覆盖遥感信息提取研究[J]. 安徽农业科学, 2009, 37(14): 6696-6699.]
- [10] Yang Jinhong, Yi Qiu, Gu Songshan, et al. Band Combination Research of City Region Hyperspectral Remote Sensing Data in False Color Synthesis[J]. Journal of Nanjing Institute of Meteorology, 2005, 28(3): 289-295. [杨金红, 尹球, 顾松山, 等. 城区高光谱遥感数据假彩色波段组合研究[J]. 南京气象学院学报, 2005, 28(3): 289-295.]
- [11] Chavez P S, Berlin G L, Sowers L B. Statistical-method for Selecting Landsat MSS Ratios[J]. Journal of Applied Photographic Engineering, 1982, 8(1): 23-30.
- [12] Manolakis D, Marde D. Dimensionality Reduction Perspectral Imaging Data Using Local Principal Components Transforms [C]//Proceedings of SPIE, 2004, 5425: 393-401.
- [13] Yang Zhusheng, Guo Lei, Luo Xin. A PCA Based Band Selection Algorithm of Hyperspectral Image[J]. Microelectronics & Computer, 2006, 23(12): 72-74. [杨诸胜, 郭雷, 罗欣, 等. 一种基于主成分分析的高光谱图像波段选择算法[J]. 微电子学与计算机, 2006, 23(12): 72-74.]
- [14] Green A A, Berman M, Switzer P, et al. A Transformation for Ordering Multispectral Data in Terms of Image Quality with Implications for Noise Removal[J]. IEEE Transactions on Geoscience and Remote Sensing, 1988, 26(1): 65-74.
- [15] Liu Chunhong, Zhao Chunhui, Zhang Lingyan. A New Method of Hyperspectral Remote Sensing Image Dimensional Reduction[J]. Journal of Image and Graphics, 2005, 10(2): 218-222. [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.]
- [16] Bruzzone L, Roli F, Serpico S B. An Extension of the Jeffreys-Matusita Distance to Multiclass Cases for Feature Selection [J]. IEEE Transactions on Geoscience and Remote Sensing, 1995, 33(6): 1318-1321.
- [17] Wang Changyao, Liu Zhengjun, Yan Chunyan. A Experimental Study on Imaging Spectrometer Data Feature Selection and Wheat Type Identification[J]. Journal of Remote Sensing, 2006, 10(2): 249-254. [王长耀, 刘正军, 颜春燕. 成像光谱数据特征选择及小麦品种识别实验研究[J]. 遥感学报, 2006, 10(2): 249-254.]
- [18] De Backer S, Kempeneers P, Debruyne W, *et al.* A Band Selection Technique for Spectral Classification[J]. IEEE Geoscience and Remote Sensing Letters, 2005, 2(3): 319-323.
- [19] Kailath T. Divergence and Bhattacharyya Distance Measures in Signal Selection[J]. IEEE Transactions on Communication Technology, 1967, 15(1): 52-60.
- [20] Ball J E, West T, Prasad S, *et al.* Level Set Hyperspectral Image Segmentation Using Spectral Information Divergence-based Best Band Selection[C]//2007 IEEE International Geoscience and Remote Sensing Symposium, Vols 1-12-Sensing and Understanding Our Planet, 2007, IEEE: New York.
- [21] Liu Jianping, Zhao Yingshi. Methods on Optimal Bands Selection in Hyperspectral Remote Sensing Data Interpretation [J]. Journal of Graduate School, Academia Sinica, 1999, 16(2): 153-161. [刘建平, 赵英时. 高光谱遥感数据解译的最佳波段选择方法研究[J]. 中国科学院研究生院学报, 1999, 16(2): 153-161.]
- [22] Qi La, Liu Liangyun, Zhao Chunjiang, *et al.* Selection of Optimum Periods for Extracting Winter Wheat Based on Multi-temporal Remote Sensing Images[J]. Remote Sensing Technology and Application, 2008, 23(2): 154-160. [齐腊, 刘良云, 赵春江, 等. 基于遥感影像时间序列的冬小麦种植监测最佳时相选择研究[J]. 遥感技术与应用, 2008, 23(2): 154-160.]

Determination on the Optimum Band Combination of HJ-1A Hyperspectral Data in the Case Region of Dongguan Based on Optimum Index Factor and J-M Distance

MA Na^{1,2}, Hu Yun-feng¹, ZHUANG Da-fang¹, WANG Xin-sheng²

(1. Institute of Geographic Sciences and Natural Resources Research, Beijing 100101, China;

2. School of Resources and Environment, Hubei University, Wuhan 430062, China)

Abstract: Comparing with traditional remote sensing data, strong band relevant and data redundancy are prevalent in the hyperspectral remote sensing data, so band selection is necessary for the efficient research. This paper taking Dongguan as a study area, using Environment and Disaster Monitoring and Forecasting of small satellite hyperspectral data, comprehensively analyzed the information content and correlation of the bands, then using three kinds of classical band index to choose the best band combinations. Through contrastive analysis of the three models, we improved the Optimum Index Factor by giving thresholds to coefficient of correlation and mean square deviation to select the optimum band combinations on the problems in the classical models. In the end, taking the farmland, woodland and grassland for example, apply Jeffreys-Matusita Distance model to calculate their separability and pointed out that the band combination of 50-80-108, 50-79-108 and 50-80-111 are the best combinations for distinguishing grassland-woodland, grassland-farmland and woodland-farmland separately.

Key words: Environment and disaster monitoring and forecasting of small satellite; Hyperspectral; Optimum band combination; Separability