

文章编号: 1000-2022(2005)03-0289-07

城区高光谱遥感数据假彩色波段组合研究

杨金红¹, 尹 球², 顾松山¹,
周 宁², 刘晓明³

(1 南京信息工程大学 电子工程系, 江苏 南京 210044

2 中国科学院 上海技术物理研究所, 上海 200083; 3 中国运载火箭技术研究院, 北京 100076)

摘 要: 高光谱数据具有波段数目多、波段宽度窄、数据量庞大等特点, 如何根据具体的应用目的, 在众多的波段中选取最佳波段组合用于假彩色合成, 对于有效进行高光谱数据处理、分析及信息提取至关重要。以面阵推帚式机载超光谱成像仪 (PH I) 获得的上海市黄浦江附近复杂地表高光谱图像数据为例, 分析了图像所包含的信息量、各通道之间的相关性以及影像上各地物的光谱特征, 选出了那些包含信息量大、相关性小、光谱差异大的波段子集, 然后再结合协方差矩阵特征值法、最佳指数法和波段指数法波段组合方法选出了高光谱遥感图像的最佳波段组合。

关键词: 遥感; 高光谱遥感; 波段组合; 假彩色合成

中图分类号: TP75 **文献标识码:** A

20 世纪 80 年代遥感技术的最大成就之一是高光谱遥感技术的兴起, 它是一项在电磁波谱的紫外、可见光、近红外和中红外区域, 获取非常窄且光谱连续的图像数据技术^[1], 所以往往具有波段数目多、数据量庞大等特点。自世界上第一幅高光谱分辨率图像以全新的面貌呈现在科学界面前以来, 一些图像解译专家就不断利用新的处理手段和方法来增强图像信息, 但是高光谱所具有的波段数目多、数据量庞大等特点给图像解译人员带来诸多困难。如何从众多的波段中选出三波段用于合成假彩色图像, 以突出感兴趣区域, 显得尤为重要。以往波段选择方法都遵循信息量大和波段相关性小这两个因素 (如使用最多的最佳指数法), 但是对于高光谱数据来说, 遵循这两个原则已不能满足需要, 因为成像仪在对空间目标成像时, 还对空间每个像元在几十个乃至几百个窄波段上进行连续光谱覆盖, 即高光谱数据具有空间和光谱两种可视特性。从高光谱所具有的精细光谱特征出发来选择最佳波段, 至今研究的还不多。刘建平等^[2]分析了利用信息量方法来选择波段的局限性, 提出了基于类间可分性的高光谱数据最佳波段选择模型; 罗音等^[3]分析了遥感数据各通道的标准差、亮度差及各通道的相关性, 详细解释了信息熵、联合熵的概念, 并将其用于多光谱、高光谱最佳波段选择; 杜华强等^[4]分析了最大值、最小值、标准差等传统统计量作为高光谱遥感数据波段选择方法的优缺点后, 将分

收稿日期: 2005-01-07 改回日期: 2005-03-07

基金项目: 国家自然科学基金资助项目 (40271084)

作者简介: 杨金红 (1974-), 女, 山西稷山人, 硕士生, 研究方向: 高光谱遥感。E-mail yangjh0407@163.com.

形维数作为波段选择的一个指标。以上这些方法都没有从高光谱所具有的核心问题出发,即高光谱的重要性在于能够反映地物的精细光谱特征。因此在最佳波段选择方法(协方差特征值法、最佳指数法、波段指数法)基础上考虑影像中各地物的光谱特征应该是高光谱遥感数据波段选择方法的又一重要方面。

1 试验数据和方法

高光谱最佳波段选择试验采用 2003 年由面阵推帚式机载超光谱成像仪 (PH I) 获得的遥感图像数据。仪器光谱分辨率 5 nm; 地面分辨率 2.4 m × 2.4 m; 瞬时视场 10^{-3} rad; 总视场角 18° ; 飞行高度 2 000 m, 飞行区域位于上海市黄浦江卢浦大桥附近, 影像上主要地物是水体, 其次是建筑物、桥梁、公路、船只, 以及零星分布着的城市绿化树木; 数据波长范围: 紫色 1~5 波段 (410.95~428.55 nm), 蓝色 6~14 波段 (432.95~468.25 nm), 青色 15~21 波段 (472.65~499.05 nm), 绿色 22~33 波段 (503.65~554.65 nm), 橙色 34~47 波段 (559.25~620.00 nm), 红色 48~76 波段 (624.80~758.40 nm), 近红外 77~124 波段 (763.15~989.40 nm), 共 124 个波段。采用美国 RSI 公司开发的 ENV I 软件包, 先计算图像各波段的统计数据以及绘制影像上主要地物的光谱曲线, 然后用 IDL 语言对各种波段组合模型编程计算, 按从大到小列出每种组合波段的数值组合, 再结合高光谱数据所具有的精细光谱特性选出最佳组合, 分别将其赋予红绿蓝三颜色合成假彩色图像, 最后进行比较分析。

2 波段选择的数值评价

2.1 图像的标准差

根据遥感图像各波段包含的信息量进行数值评价来选择波段是进行波段组合的第一步^[5]。通过分析, 可以确定哪几部分或哪几个波段(即波段子集)包含信息量的多少。各波段的标准差反映了图像各像元灰度值与平均值总的离散度, 一定程度上反映了各波段的信息量, 其值越大, 所包含的信息量越大。图 1 给出了图像在 124 个波段的标准差。

从图 1 看, 波段 1~19 (410.95~490.25 nm)、78 (767.90 nm)、113~117 (936.40~955.70 nm) 标准差较小(基本都小于 2 000), 所以这些波段包含的信息量就少, 在选择波段时可以不予考虑; 而波段 20~72 (494.65~739.20 nm)、77 (763.15 nm)、79 (772.70 nm)、83~112 (791.80~931.50 nm)、118~124 (960.50~989.40 nm) 的标准差较大(介于 2 000 到 3 000 之间); 波段 73~76 (744.00~758.40 nm)、80~82 (777.50~787.10 nm) 的标准差很大(其值都在 3 000 以上), 其中波段 75 (753.60 nm) 的标准差最大, 所以这些波段包含的信息量就较多, 是理想的波段子集。

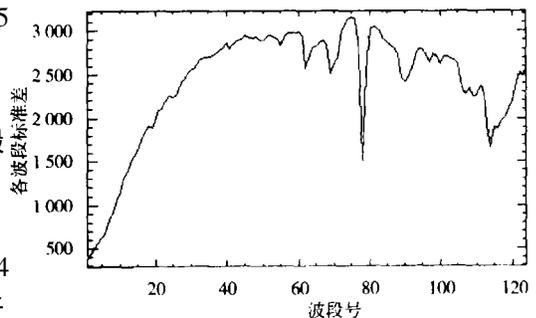


图 1 124 个波段的标准差曲线

Fig 1 Standard deviations of 124 bands

2.2 相关性

高光谱数据波段数目多、波段宽度窄的特点决定了高光谱数据波段间的相关性大、信息重叠度高^[6]。波段间的相关系数反映了两个波段间的信息重叠度, 如果两个波段间的相关系数大, 则说明它们的信息重叠度高, 因此可将相关性大的两个通道合二为一或者取其中一个。N

个波段间的相关系数矩阵形式为 (R_{ij} 为相关系数):

$$R = \begin{pmatrix} 1 & R_{12} & R_{13} & \cdots & R_{1N} \\ R_{21} & 1 & R_{23} & \cdots & R_{2N} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ R_{N1} & R_{N2} & R_{N3} & \cdots & 1 \end{pmatrix}$$

一般来说相邻波段间的相关系数较大,与相隔较远的波段间的相关系数较低(图 2 见图版,图像的明暗程度表示相关系数的大小)。

2.3 光谱特性

遥感数据是地物对电磁波反射信息及地物自身辐射信息的综合,本质上记载了地物的总辐射量即通常所说的 DN 值。各地物由于其结构、组成及物理化学性质不同,因而光谱特性在一般情况下也存在差异。基于此,从影像上选择了 9 种主要地物(水体、零星树木、公路、桥梁、船只、塑料顶棚 1、塑料顶棚 2、水泥平顶、沥青屋顶)绘制其在 124 个波段上的光谱(DN 值)曲线(图 3)。根据曲线的变化趋势来选择波段可以达到事半功倍的效果,一般来说 DN 值重叠较多的波段不宜选,差异较大的波段是理想的候选波段。

从图 3a 看,桥梁在 124 个波段上的 DN 值大于其他任一地物,所以在任一波段上都能区分桥梁与其他地物;在 42~60(596.35~681.95 nm)波段、92~105(835.10~897.80 nm)波段,水体、零星树木、船只、公路的 DN 值差异较明显,是理想的波段子集;在图 3b 上,85~110(801.40~921.90 nm)波段各建筑物屋顶材料的 DN 值差异较其他波段显著,所以这几种建筑物屋顶使用的材料在这些波段子集上可分性要好一些,也是理想的波段子集,在进行选择时可予以考虑。具有重叠现象的波段并不是说不好,这要根据具体的研究目标而定。

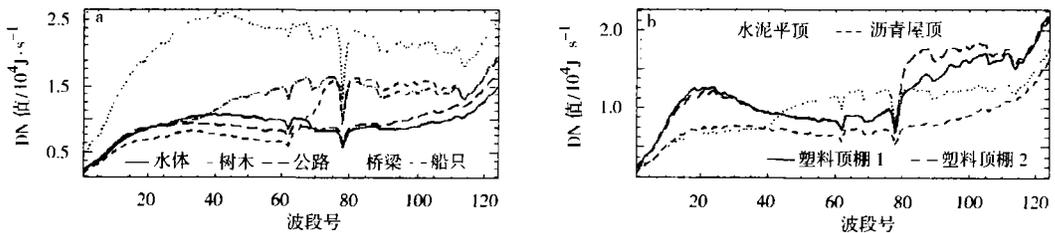


图 3 9 种地物的光谱特性

Fig 3 Spectral character of nine types of ground objects

3 协方差矩阵特征值法

在多波段遥感图像的波段组合选择中,章孝灿等、Charles^[7-8]提出了用计算 N 维数据熵值原理进行优选最佳波段组合的方法,最后选出了包含信息量丰富的波段组合。其公式为:

$$H = \frac{N}{2} + \ln(K_s) = \frac{N}{2} + \frac{N}{2} \ln(2\pi) + \frac{1}{2} \ln |M_s|$$

由于要选取的是三个波段,所以式中 $M_s = \begin{vmatrix} S_{11}^2 & C_{12} & C_{13} \\ C_{21} & S_{22}^2 & C_{23} \\ C_{31} & C_{32} & S_{33}^2 \end{vmatrix}$ 为三阶方阵。其中对角线上的元素 S_{ii}^2

分别为各波段的方差,非对角线上的元素为两个不同波段的协方差。因为图像的熵(信息量) H 随着变量协方差矩阵行列式的值变化,所以计算任意三波段组合的协方差矩阵行列式的值

并进行由大到小排序,就可选出包含信息量丰富的三波段组合。

为了验证上述方法在高光谱遥感数据波段组合选择中的有效性,计算 124 个波段中每三个波段组合的协方差矩阵行列式值并按从大到小排列(表 1,表中都考虑排在前 100 位的波段组合)。从排在前 100 位的波段组合来看,第 1 列的波段基本上集中在波段 36~46、47~48、51~52 上,第 2 列集中在波段 74~76、80、81 上,第 3 列集中在波段 122、123、124 上。这与前面分析的由标准差得出的包含信息量丰富的波段基本一致(其中波段 75 的标准差最大),而且选出的各组合波段相隔较远、相关性较小。所以协方差矩阵特征值法对于高光谱图像来说,也是一种具有一定精度且简便、经济的方法。根据以往的选择方法,协方差矩阵特征值最大的波段组合就是最优组合,但从合成的效果来看,影像中船只表面与零星树木基本是同一颜色,不能有效区分这两种地物。结合图 3a 各地物光谱的差异,发现它们在波段 51(639.10 nm)可分性大,所以选取波段 51、75、124(此波段组合排在第 55 位)做为组合波段,经过目视解译,将波段 75 赋予红色、波段 51 赋予绿色、波段 124 赋予蓝色。图 4(见图版)显示了波段 75、51、124 合成的假彩色图,可以看出图中各地物轮廓清晰可见,并且船只、零星树木呈不同颜色,可区别开来。但这种合成也有缺点:一是合成图像上某些建筑物屋顶使用的材料还不能区分开,这是因为它们在此三波段组合上的光谱差异不是十分明显;二是协方差矩阵特征值法是在服从正态分布的情况下得出的模型,而高光谱遥感图像数据并不完全服从正态分布,所以选出的波段组合并不是完全按照信息量的大小排列;三是从表中计算所得的结果来看,数据值比较大,在编程过程中采用了双精度计算,计算速度相对来说要慢得多。

表 1 协方差矩阵行列式值的三波段组合表

Table 1 Three bands combinations of covariance matrix

波段组合	协方差矩阵行列式值
band40 band75 band124	40 138 131 680 380 322.0
band40 band76 band124	39 732 201 131 367 858.0
band40 band75 band122	38 922 906 617 021 858.0
band39 band75 band124	38 880 704 121 009 570.0
band40 band75 band123	38 635 901 881 201 066.0
band40 band76 band122	38 497 310 609 405 641.0
band42 band75 band124	38 467 702 973 300 998.0
band39 band76 band124	38 448 068 720 008 364.0
band38 band75 band124	38 400 711 175 569 801.0
band40 band76 band123	38 241 783 921 624 482.0
⋮	⋮

4 指数法

4.1 最佳指数法

图像数据的标准差越大,所包含的信息量也越大;而波段间的相关系数越小,表明各波段图像数据的独立性越高、信息冗余度越小。故可采用美国查维茨提出的最佳指数波段选择法^[9],即

$$O_{\text{F}} = \sum_{i=1}^3 S_i / \sum_{j=1}^3 |R_{ij}|.$$

其中 O_{F} 是最佳指数因子, S_i 为 i 波段的标准差, R_{ij} 是 i 波段与 j 波段之间的相关系数。

在众多的波段组合中, O_{F} 越大意味着此三个波段间的相关性越小、包含信息量越丰富。因此,对所计算出的最佳指数因子 O_{F} 进行由大到小排序,最佳波段组合问题可以得到初步解决。表 2 是最佳指数计算结果。

从表 2 看, 虽然选出的排在前面的波段组合中都是包含信息量丰富的波段, 但是有较多的相邻波段, 如 75 与 76、74 与 75、44 与 45 等, 而相邻波段间的相关性较大, 其相关系数高达 0.956 2 以上, 这在图 2(见图版) 中可看出。很显然, 这在高光谱最佳波段选择中是不可取的。从排在前面 100 位组合中考虑, 第 1 列集中在波段 38~ 40、42~ 45, 第 2 列集中在波段 42~ 48、74~ 76, 第 3 列集中在波段 75~ 76、80~ 83、94~ 95、101~ 102, 再结合图 3b 两种不同的塑料顶棚在波段 83(791.80 nm)可分性好, 而波段 44 和 75 都是包含信息量较丰富的波段, 所以选取 44、75、83 波段组合(排在第 19 位), 再考虑颜色赋值, 将 44、83、75 做为最佳组合。图 5(见图版) 是合成的结果, 塑料顶棚与水泥平顶、沥青屋顶都区别开来了。

表 2 最佳指数值的三波段组合表

Table 2 Three bands combinations of optimal index

波段组合	最佳指数值
band40 band75 band76	1 207 095 077 74
band40 band45 band75	1 204 345 684 20
band40 band74 band75	1 203 777 123 09
band40 band45 band95	1 203 627 044 96
band40 band45 band76	1 203 576 568 09
⋮	⋮
band42 band75 band76	1 200 927 816 82
band44 band45 band75	1 200 802 548 14
band40 band42 band75	1 200 694 450 14
band45 band75 band76	1 200 618 861 67
band44 band45 band95	1 200 560 588 86
⋮	⋮

4.2 波段指数法

波段指数^[6]的定义: 设 R_{ij} 为通道 i 与 j 之间的相关系数, 高光谱数据被分为 k 组, 每组的波段数分别为 n_1, n_2, \dots, n_k 。定义波段指数 $p_i = \frac{S_i}{R_i}$, 其中 $R_i = R_w + R_a$, $R_w = \frac{1}{n_k} \sum_{j=1}^{n_k} R_{ij} (i \neq j)$ 。式中 S_i 为第 i 波段的标准差, R_w 为第 i 波段与所在组内其他波段相关系数的绝对值之和的平均值, R_a 为第 i 波段与所在组外的其他波段之间的相关系数的绝对值之和。从公式可以看出, 它综合考虑了波段的信息量和相关性两个因素。标准差越大, 表明波段的离散程度越大, 所含的信息量越丰富, 而波段的总体相关系数的绝对值越小, 表明通道数据的独立性越强, 信息冗余度越小。

在利用高光谱数据进行假彩色合成时, 一般应从近红外、红光、蓝绿光波段中各选一个最优波段进行合成。为此, 可将 124 个波段根据波长大小分为 3 组, 即波段 1~ 47(除红光外的可见光组)、48~ 76(红光组)、77~ 124(近红外)。然后利用波段指数法得出各组的波段指数排列, 表 3 是对数据经过波段指数计算得出的各组排序结果。

从表 3 可以看出: 在可见光组内, 排在前面的 30~ 47 波段都是包含信息量丰富的波段(这可从图 1 看出), 它们之间的波段指数相差较红光组大, 且它们是相邻波段, 相关系数较大, 再结合图 3a 5 种地物在波段 45(610.50 nm)可分性较大, 因此可做为优选波段之一。在红光波段组内, 波段指数相差较小, 排在前面的波段 48~ 61、73~ 76 也都包含丰富的信息量, 但是波段 48~ 61 与波段 73~ 76 之间相关系数较小, 因此可取波段 48~ 61 或 73~ 76 之间任意一个, 考虑 75 波段是包含信息量丰富的波段, 所以取它做为另一优选波段。在近红外波段, 排在前面的 80~ 87、93~ 96、98~ 99、101~ 104 波段相关性较可见光组小, 较红光组大, 波段指数相差也较前两组大, 其中在波段 81(782.30 nm)上各建筑物屋顶可分性好一些, 而且波段指数值又大。所以选取波段 45、75、81 作为组合波段, 再考虑到颜色的赋值问题, 经目视解译, 将波段

81、45、75作为最佳组合,图 6(见图版)是合成的假彩色图。由图可见,影像中船只、零星树木、各建筑物屋顶以不同的颜色显示,判读效果好于前两种波段选择方法得出的合成图。所以,波段指数法应用于由面阵推帚式机载超光谱成像仪 (PH I)获得的遥感数据处理时是一种较理想的方法。

表 3 高光谱数据波段指数分组排序

Table 3 Grouping sort of hyperspectral data band index

除红光外可见光波段	波段指数	红光波段	波段指数	近红外波段	波段指数
40	51 386 650	48	39 413 494	81	50.192 272
45	51 024 932	60	39 159 578	82	50.031 125
44	50 913 501	51	39 115 268	80	49.782 743
43	50 900 192	52	39 109 839	83	49.449 929
42	50 765 936	75	39 088 003	95	48.402 606
39	50 493 094	57	39 085 545	94	48.135 411
46	50 263 396	58	38 970 284	84	47.956 609
38	50 167 046	59	38 968 425	98	47.646 279
41	50 051 401	53	38 987 197	96	47.558 034
47	49 973 937	49	38 738 594	85	47.323 944
37	49 489 984	76	38 714 089	99	47.226 445
36	49 365 702	56	38 683 279	101	47.165 579
35	48 784 529	50	38 645 553	102	47.143 792
34	48 565 335	54	38 632 962	86	46.881 706
33	48 145 550	74	38 622 047	103	46.702 907
32	47 325 733	61	38 527 224	87	46.515 880
31	46 272 628	73	37 730 929	104	46.316 683
30	45 947 459	55	37 665 909	93	45.799 584
⋮	⋮	⋮	⋮	⋮	⋮

5 结 论

(1)根据高光谱数据各波段包含的信息量、波段间的相关性以及影像中各地物 DN 值光谱曲线的差异性,再结合协方差矩阵特征值法、最佳指数法、波段指数法来进行最佳波段选择以达到遥感数据各地物的有效分类方法,弥补了以往波段选择的不足之处,取得了较好的效果。

(2)利用协方差矩阵特征值法可得到包含信息量丰富的波段,但是它是在完全服从正态分布的情况下得出的模型,选出的波段组合并非完全按照信息量大小排列。另外这种模型的缺点是计算数据量大,处理速度慢。

(3)最佳指数法考虑了图像的标准差(信息量的大小)和波段间的相关性两个因素,但是这种方法得到的结果并不尽人意,往往出现相邻波段的组合情况,相邻波段的相关性一般都很大,显然这不符合要求。考虑到各地物的光谱特征后选出的波段组合对图象效果有所改善。

(4)波段指数是从高光谱数据具有的特点出发,综合考虑了图像的信息量和相关性两个因素,并且是在分组成块的情况下设计的模型,而且还发现选出的最佳波段组合与由最佳指数法(考虑光谱特性后)得出的基本一致(因为波段 44 与 45、81 与 83 之间的相关系数很大,可取其中之一代替另一波段)。因此,波段指数法对高光谱数据来说是一种较理想的最佳波段选择方法。

(5)明确区分一幅影像上各个地物的最佳波段组合可能并不只有一组,也许在此波段组合上这两种地物区分效果好,而另两种地物区分效果不好。所以要根据具体的研究对象和研究目标来选择那些包含信息量丰富、相关性小、且光谱差异又显著的波段。

参考文献:

- [1] 浦瑞良,宫 鹏. 高光谱遥感及其应用[M]. 北京: 高等教育出版社, 2000: 3-4

- [2] 刘建平, 赵英时, 孙淑玲, 等. 高光谱遥感数据最佳波段选择方法试验研究 [J]. 遥感技术与应用, 2001, 16(1): 7-13
- [3] 罗 音, 舒 宁. 基于信息量确定遥感图像主要波段的方法 [J]. 城市勘测, 2002(4): 28-32
- [4] 杜华强, 赵宪文, 范文义. 分形维数作为高光谱遥感数据波段选择的一个指标 [J]. 遥感技术与应用, 2004, 19(1): 5-9.
- [5] 崔林丽. 高光谱遥感数据在荒漠化监测中的定量反演研究 [D]. 哈尔滨: 东北林业大学森林经济学系, 2002 12-13
- [6] 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—以北京顺义地区为例 [J]. 遥感技术与应用, 2002, 17(2): 59-65
- [7] 章孝灿, 黄智才, 赵元洪. 遥感数字图像处理 [M]. 杭州: 浙江大学出版社, 2003: 174-175.
- [8] Charles Selecting band combination from multispectral data[J]. Photogrammetric Engineering and Remote Sensing 1985 51 (6): 681-687
- [9] 刘建平. 高光谱遥感数据处理分析软件系统设计与实现 [D]. 合肥: 中国科学技术大学计算机系, 2001: 22-23

Band Combination Research of City Region Hyperspectral Remote Sensing Data in False Color Synthesis

YANG Jin-hong¹, YIN Qiu², GU Song-shan¹,
ZHOU Ning², LU Xiaoming³

(1 Department of Electronic Engineering NU IST, Nanjing 210044 China

2 Shanghai Institute of Technical Physics Chinese Academy of Sciences Shanghai 200083 China

3 Chinese Academy of Launch Vehicle Technology, Beijing 100076 China)

Abstract In comparison with multispectral remote sensing the characteristics of hyperspectral remote sensing data are more channels narrower bandwidth and larger amount of data. It is very important to make optimal band combination in numerous bands for false color synthesis according to specific application purpose for further management analysis and information extraction of hyperspectral remote sensing data. Taking the complex ground surface hyperspectral image near the Huangpujiang in Shanghai from an airborne Pushbroom Hyperspectral Imager (PHI) as an example, the information content, the correlation of different channels and spectral character of ground objects are analysed to extract the band subsets with vast scale information content, smaller correlation, larger spectral difference, and then the optimal band combination of hyperspectral image is determined by using methods of covariance matrix, optimal index and band index.

Key words remote sensing hyperspectral remote sensing band combination false color synthesis