

# 一种新的高光谱遥感图像降维方法

刘春红 赵春晖 张凌雁

(哈尔滨工程大学信息与通信工程学院, 哈尔滨 150001)

**摘要** 高光谱遥感图像的高数据维给图像进一步处理带来了困难,为了解决这一问题,提出了自适应波段选择(ABS)的降维方法。该方法充分考虑了高光谱图像的空间相关性和谱间相关性,通过计算各个波段的指数来选择信息量大并且与其他波段相关性小的波段。对各波段相应的指数重新排列之后,有两种方法来选择最终波段:一种是选择波段指数比设定指数大的波段,另一种方法是选择波段指数排在前 $n$ 个的所有波段。为了验证ABS方法的有效性,对降维后的高光谱图像进行了贝叶斯监督分类,分类结果表明自适应波段选择的方法能够选择出信息丰富的波段,分类精度与使用原始波段相比提高10.4%,计算复杂度大大降低。

**关键词** 高光谱遥感 自适应波段选择 贝叶斯监督分类 分类精度 Kappa系数

**中图分类号**: TN911.73 **文献标识码**: A **文章编号**: 1006-8961(2005)02-0218-05

## A New Method of Hyperspectral Remote Sensing Image Dimensional Reduction

LIU Chun-hong, ZHAO Chun-hui, ZHANG Ling-yan

(College of Information and Communication, Harbin Engineering University, Harbin 150001)

**Abstract** The high dimensions of hyperspectral remote sensing image have caused problems to further processing. In order to solve the above problems, this paper proposes an adaptive band selection (ABS) algorithm of dimensional reduction, which selects high informative and low correlative bands by calculating the index of each band. After calculating and rearranging the index of each band, there are two methods of selecting the final bands: one is to select the bands whose index is bigger than the specified index, another is to select the first  $n$  bands whose index is the first  $n$  bands. In order to testify the effect of ABS method, Bayesian supervised classification was implemented on the dimensionally reduced image. The results of the classification show that highly informative bands can be selected by ABS method, the classification accuracy of selected bands is 10.4% bigger than the original image, and the computing complication is decreased rapidly too.

**Keywords** hyperspectral remote sensing, adaptive band selection, Bayesian supervised classification, classification accuracy, Kappa coefficient

## 1 引言

高光谱遥感(hyperspectral remote sensing)又称为超谱遥感,是随着高光谱成像光谱仪(hyperspectral image spectrometer)的发展而使获取其图像成为可能。目前应用较多的两个高光谱成像仪器有美国NASA喷气推进实验室研制的224波段机载可见光成像光谱仪(AVIRIS)<sup>[1]</sup>;另一个为海军研究实验室(naval research laboratory's)的210波段

高光谱数字图像收集实验(HYDICE)传感器<sup>[2]</sup>。这些由高分辨率成像仪器获得的图像比多光谱传感器获得的图像有着更高的分辨率,人们可以提取从多光谱图像中难以获得的信息。这些发展给遥感应用带来了希望,同时也对处理技术提出了挑战。因此,必须研究更为适用的新处理技术。高光谱遥感图像大量的波段数目使得传统的遥感处理技术不再适用,这就需要降低高光谱图像的维数。

高光谱图像通常包括上百个连续分布的波段,整个数据可以看作一个图像立方体。高光谱数据两个

收稿日期:2004-04-22;改回日期:2004-07-06

第一作者简介:刘春红(1977~),女,2002年毕业于哈尔滨工程大学获通信与信息系统专业硕士学位,现为哈尔滨工程大学信息与通信工程学院博士研究生。所从事专业为信号与信息处理。E-mail:liuchunhong@hrbeu.edu.cn

相邻波段之间一般相隔仅有 10nm, 相邻波段的图像空间、谱间相关性都非常高。并不是所有的波段都有着同等的重要性, 通过使用特定的算法进行降维, 形成新的高光谱图像空间, 在不损失重要信息的条件下可以代表其他波段的信息。因此, 进行合理的降维是可行的。

现有的降维方法, 可以分为两类: 一类是基于变换的方法, 如主成分分析 (PCA)、正交子空间投影 (OSP)、正则分析 (CA)、离散小波变换 (DWT) 等。基于变换的降维方式, 其优点是可以经过若干变换直接将高维数据降低到几维甚至 1 维, 降维速度快; 缺点是对图像进行了变换, 改变了图像原有的特性。另一类是基于非变换的, 如波段选择、数据源划分<sup>[3]</sup>等。基于非变换的降维方式是在考察图像整体特点之后对图像进行选择 and 划分的, 它克服了变换法改变图像的缺点, 因此更有利于保持图像的原有特性。

目前的波段选择方法, 一般都是对于多光谱图像 (4~7 个波段) 进行组合选优的。最初的波段选择主要依赖于人工目视解译, 这个工作需要由有经验的专家来进行。但是随着遥感图像数字化的不断发展, 依赖于专家目视解译的方法越来越显示出它的局限性。例如, 有时候很难说明选择该波段的原因, 对一个新的图像数据来进行选择容易出现人为的判断失误<sup>[4]</sup>。对于数字化的高光谱图像进行准确的波段选择, 必须发展有效而简便的识别方法。

文献[5]利用 JM 距离和一般地表覆盖类型的光谱集合的专家系统来进行波段选择。然而通过进一步的研究发现该方法不仅需要在地表覆盖类型和成像时间的关系有着深入的了解, 而且要发展有效的 JM 距离选择策略。这些要求对于高光谱波段选择, 显然是不适用的。

文献[6]提出了基于类间可分性的最佳组合波段选择方法, 并对基于信息量和基于类间可分性的诸方法进行了仿真实验。但它所涉及的波段选择研究都是对于特定的波段组合进行的。在实际应用中, 由于高光谱图像维数过多, 如果从 220 个波段中选择出 100 个波段, 那么其组合数为  $C_{220}^{100}$ , 如果要想成功的辨识某些地物, 比如湖泊, 就要考虑到波段的排列问题, 那么可能的波段排列总数为  $P_{220}^{100}$ 。对于上述所有可能的波段组合, 再求其相应的 Bhattacharyya 距离、离散度、协方差矩阵等, 计算量将会十分巨大。

在充分考查高光谱图像的特性后, 提出了一种

自适应波段选择 (adaptive band selection, ABS) 降维的新方法, 应用此方法可以将信息含量丰富并且与其他波段相关性小的波段选择出来。

## 2 自适应波段选择

1982 年, 美国的 Chavez, Berlin 和 Sowers 等人提出了最优索引因子 (optimum index factor) 的组合波段选优方法<sup>[7]</sup>, 该方法依据下式给出  $n$  个波段组合中最优的指数大小

$$\text{OIF} = \frac{\sum_{i=1}^n \sigma_i}{\sum_{i=1}^n \sum_{j=i+1}^n |R_{i,j}|} \quad (1)$$

式中,  $\sigma_i$  为第  $i$  个波段的标准差,  $R_{i,j}$  表示第  $i$  个波段和第  $j$  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 3 个波段组合在一起。OIF 越大, 波段组合越优。按 OIF 从大到小的顺序进行排列, 选出最佳组合。

在实际应用中, 这种方法存在它的局限性, 它选择出来的最优波段组合未必是最优的。因为对一定位置的波段来说, 图像方差不变的, 不相邻的波段相关系数肯定要比相邻的小, 而且波段之间相隔越远相关系数就越小, 这样组合图像也许并不是图像分析者所期望能获得的最大信息量。另一个原因是 OIF 对于高光谱图像波段选择而言计算量过大。

对 OIF 方法和引言中所述的各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问相关性和谱间相关性, 并构造了相应的数学模型, 对求出的各个波段指数进行由大到小的排列, 系统根据设定的阈值自适应地选择需要的波段。ABS 方法依据的原则是:

- (1) 所选择的波段信息量要大;
- (2) 所选择的波段与其他波段的相关性要小。

根据上述原则构造的数学模型如下:

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

$$\sigma_i = \left[ \frac{1}{M \times N} \sum_{x=1}^M \sum_{y=1}^N (f_i(x,y) - \bar{f}_i)^2 \right]^{\frac{1}{2}} \quad (3)$$

$$R_{i,j}(x,y) = \frac{E\{(f_i(x,y) - \bar{f}_i)(f_j(x,y) - \bar{f}_j)\}}{\sqrt{E\{(f_i(x,y) - \bar{f}_i)\}^2} \sqrt{E\{(f_j(x,y) - \bar{f}_j)\}^2}} \quad (4)$$

其中,  $\sigma_i$  为第  $i$  个波段的标准差;  $E\{\}$  为数学期望,

$R_{i-1,i}$  和  $R_{i,i+1}$  是第  $i$  个波段与其前后两波段的相关系数, 相关系数越小, 两个波段数据之间的独立性越高;  $I_i$  是第  $i$  幅图像的指数大小;  $M$ 、 $N$  分别是图像的行、列像素数,  $f_i(x, y)$  是第  $i$  幅图像;  $\bar{f}_i$  是第  $i$  幅图像的像素平均值。

ABS 方法获得的指数充分考虑了每幅图像的信息富集程度与相邻波段的相关性,  $I_i$  越大, 相应图像的信息量就越大, 越具有代表性。最终波段选择的方法有两种: 一种是系统根据设定的阈值选择波段数目, 阈值的设定可以根据后续的处理需要自适应地选择适合的指数值, 确定了阈值之后, 指数大于该阈值的波段就被选择出来; 另一种方法就是选择波段指数排在前面的  $n$  个波段。

表 1 谱带 B28 ~ B39 的相关系数矩阵

Table 1 The correlation matrix of band B28 ~ B39

波段	B28	B29	B30	B31	B32	B33	B34	B35	B36	B37	B38	B39
B28	1.000 0											
B29	0.987 2	1.000 0										
B30	0.914 6	0.961 9	1.000 0									
B31	0.524 3	0.633 8	0.810 1	1.000 0								
B32	0.113 8	0.011 4	0.263 8	0.767 2	1.000 0							
B33	0.413 0	0.302 7	0.056 1	0.522 9	0.942 6	1.000 0						
B34	0.502 6	0.401 2	0.164 1	0.424 4	0.899 2	0.990 5	1.000 0					
B35	0.539 9	0.442 6	0.210 5	0.379 3	0.875 9	0.982 7	0.994 3	1.000 0				
B36	0.546 9	0.450 2	0.219 0	0.371 3	0.872 2	0.982 0	0.995 2	0.996 4	1.000 0			
B37	0.550 1	0.453 5	0.222 8	0.367 1	0.869 4	0.980 6	0.995 3	0.996 0	0.997 7	1.000 0		
B38	0.547 0	0.449 7	0.218 4	0.371 2	0.871 3	0.981 1	0.994 4	0.996 0	0.997 8	0.995 7	1.000 0	
B39	0.549 1	0.451 7	0.220 6	0.368 7	0.869 5	0.990 0	0.993 5	0.995 5	0.997 4	0.997 5	0.997 8	1.000 0

表 2 指数及其对应的波段号(前 18 个)

Table 2 Index and its corresponding band number (first 18th)

序号	指数	波段号	序号	指数	波段号	序号	指数	波段号
1	1012.5	23	7	909.66	36	13	860.12	25
2	996.04	22	8	903.21	26	14	855.93	37
3	932.56	21	9	886.95	35	15	840.25	16
4	932.55	20	10	875.92	18	16	839.87	38
5	911.09	19	11	874.33	27	17	813.24	33
6	910.47	24	12	865.39	17	18	798.35	15

### 3 贝叶斯分类法

为了说明经波段选择后的图像更有利于分类, 采用贝叶斯监督分类方法对高光谱图像进行分类。首先参考土地资源利用图, 选定训练区域和训练样本。设类别  $i$  总数为  $L$ , 用训练区域内的点作为光谱模型, 计算每个类别的平均值向量和协方差矩阵, 作为分类的特征模型。设类别  $i$  的像素总数为  $N(i)$ , 则  $n$  维平均值向量  $M(i)$  和协方差矩阵  $\Sigma(i)$  分别由

为了验证自适应波段选择方法的有效性, 采用高光谱 AVIRIS 原始图像进行了实验。该图像取自 1992 年 6 月拍摄的美国印第安纳州西北部印第安遥感试验区的一部分, 去掉受噪声和水汽吸收较明显的谱段, 从原始的 220 个波段中选取了 180 个作为研究对象。它包含了农作物和森林植被的混合区。

180 个波段的部分相关系数矩阵如表 1 所示, 可以看出相邻波段的相关性一般比较大, 随着波段距离的增大, 相关系数相应地减小。用自适应波段选择方法得出的指数如表 2 所示, 其中, 指数  $I_i$  由大到小排列, 后面是对应的波段号(该波段号为在 180 个波段集中的序号)。本文根据波段对应的指数值选择了指数排在前面的 50 个波段。

下式确定<sup>[8]</sup>。

$$M(i) = \begin{bmatrix} m_1(i) \\ m_2(i) \\ \dots \\ m_n(i) \end{bmatrix} = E_i \{X\} = \frac{\sum_{j=1}^{N(i)} X_j(i)}{N(i)} = \frac{1}{N(i)} \begin{bmatrix} \sum_{j=1}^{N(i)} x_{1j}(i) \\ \sum_{j=1}^{N(i)} x_{2j}(i) \\ \dots \\ \sum_{j=1}^{N(i)} x_{nj}(i) \end{bmatrix} \quad (5)$$

$$\Sigma(i) = E_i \{ (X(i) - M(i))(X(i) - M(i))^T \} = \frac{\sum_{j=1}^{N(i)} X_j(i) X_j(i)^T}{N(i)} - M(i) M(i)^T \quad (6)$$

贝叶斯分类法的判决规则为

$$P(X/i)P(i) \geq P(X/j)P(j) \quad (j=1,2,\dots,l; j \neq i) \quad (7)$$

其中,  $P(i)$  为类别  $i$  的先验概率, 如果式(7)成立, 那么  $X$  属于类  $i$ 。

如果将高光谱遥感图像用“多变量正态分布”来描述, 那么有

$$P(X/i) = \frac{1}{(2\pi)^{n/2} |\Sigma(i)|^{1/2}} \times \exp\left[-\frac{\{X - M(i)\}^T \Sigma(i)^{-1} \{X - M(i)\}}{2}\right] \quad (8)$$

将式(8)代入式(7)中, 取自然对数并省去公共项, 就可以得到如下判别规则:

$$g_i(X) \geq g_j(X) \quad (j=1,2,\dots,l; j \neq i) \quad (9)$$

那么  $X$  属于类别  $i$ 。

这里

$$g_i(X) = -\ln |\Sigma(i)| - (X - M(i))^T \Sigma(i)^{-1} (X - M(i)) \quad (10)$$

### 4 实验结果与分析

原始 220 波段图像第 50、27、17 波段作为 RGB 通道的假彩色图像如图版 I 图 1, 图版 I 图 2 为实验参照的部分土地资源利用图的假彩色合成图像。为了比较分析, 实验中分别对 220 波段原始高光谱图像、去掉噪声波段的 180 波段图像和经波段选择的 50 波段图像经同样的贝叶斯分类, 分类图如图版 I 图 3 ~ 图 5 所示。所分类别数为 6, 所选类别和样本数如表 3 所示。

表 3 类别及样本的选取

Table 3 Selection of classes and samples

类别名称	玉米	牧场	草地	干草	大豆	林地
训练样本个数	527	282	227	247	658	274
检验样本个数	1376	479	660	499	2576	1229

按照第 3 节的方法构造贝叶斯监督分类器, 将式(10)作为分类时的判别函数, 将分类精度和 Kappa 系数作为评价分类效果的标准。实验得到的分类混淆矩阵如表 4、表 5 和表 6 所示, 混淆矩阵中对角线上的元素值越大, 表示分类结果越可靠。

220 波段图像总体分类精度为

$$(5261/6819) = 77.2\%$$

Kappa 系数 ( $\times 100$ ) 为 68.1% ;

Kappa 误差为 0.000049。

180 波段图像总体分类精度为

表 4 220 波段分类混淆矩阵

Table 4 220 bands classification mixture matrix

	玉米	牧场	草地	干草	大豆	林地	总计
玉米	774	0	0	0	602	0	1376
牧场	0	418	0	0	60	1	479
草地	0	1	227	0	425	7	660
干草	0	0	0	259	240	0	499
大豆	85	0	0	0	2491	0	2576
林地	0	16	0	0	121	1092	1229
总计	859	435	227	259	3939	1100	6819

表 5 180 波段分类混淆矩阵

Table 5 180 bands classification mixture matrix

	玉米	牧场	草地	干草	大豆	林地	总计
玉米	921	0	0	0	455	0	1376
牧场	0	430	0	0	46	3	479
草地	0	20	236	0	390	14	660
干草	0	0	0	474	25	0	499
大豆	120	0	0	0	2456	0	2576
林地	0	12	0	0	17	1200	1229
总计	1041	462	236	474	3389	1217	6819

表 6 50 波段分类混淆矩阵

Table 6 50 bands classification mixture matrix

	玉米	牧场	草地	干草	大豆	林地	总计
玉米	1094	2	1	0	279	0	1376
牧场	10	435	5	0	18	11	479
草地	1	39	605	0	7	8	660
干草	0	4	0	495	0	0	499
大豆	423	13	3	0	2137	0	2576
林地	0	27	3	0	1	1198	1229
总计	1528	520	617	495	2442	1217	6819

$$(5717/6819) = 83.8\%$$

Kappa 系数 ( $\times 100$ ) 为 78.0% ;

Kappa 误差为 0.000037。

50 波段图像总体分类精度为

$$(5964/6819) = 87.5\%$$

Kappa 系数 ( $\times 100$ ) 为 83.7% ;

Kappa 误差为 0.000028。

错分的原因是:

(1) 原始图像精度不高。由于实验使用的 AVIRIS 图像是从高空拍摄, 其分辨率仅为  $20m \times 20m$ , 像元混合的概率很大, 如果将像元归为某一类, 势必会带来分类误差, 导致分类精度下降, 不能反映地物的真实情况。

(2) 分类精度的大小容易受到人的主观因素的影响, 比如不同的人会选取不同的训练样本, 样本分

布的均匀程度也会影响分类效果,这些都会导致不同的均值向量和协方差矩阵,但是可以通过反复实验得到最佳分类结果。

(3) 3 幅分类图像中对玉米的分类精度都比较低,其原因主要是由于该图像的摄取时间为 1992 年的 6 月,此时玉米正处于生长的早期,其光谱特征与其他有叶植物(如大豆)非常相近,所以玉米有很大一部分被错分成了大豆。

通过分析 3 幅分类图像可以发现:220 波段分类图像由于没有去掉受噪声和水汽吸收影响的波段,具有噪声性质的像元也被统计进分类器,导致分类精度过低。180 波段分类图像与 50 波段分类图像相比,50 波段分类图像的分类精度要比 180 波段高出 3.7%,其原因在于通过 ABS 所选择的 50 个波段信息量更为集中,而 180 波段中有一些波段是具有干扰性质的波段,把这些波段都考虑进去,计算均值向量和协方差矩阵时就会出现统计错误,因而对分类结果有一定的影响。就计算量而言,从 180 个波段中选择出同样的 50 个最优波段,ABS 方法的计算量是 OIF 方法的  $1/C_{180}^{50}$ 。从这两方面来看,自适应波段选择的方法都具有不可比拟的优越性。

## 5 结 论

高光谱图像各波段间存在着高相关性和高冗余度,通过选择所需的波段来降维不会对结果有太大的影响,但会大大地降低计算量,从信息处理实效性的角度来看,通过选择波段来降低数据维数有很大的现实意义。实验结果证明,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

理的计算量并有助于提高分类效果。

## 参考文献 (Reference)

- 1 Green Robert O, Pavri Betina E, Chrien Thomas G. On-orbit radiometric and spectral calibration characteristics of EO-1 hyperion derived with an underflight of AVIRIS and in situ measurements at Salar de Arizaro, Argentina[J]. IEEE Transactions on Geoscience and Remote Sensing, 2003,41(6): 1194 ~ 1203.
- 2 Resmini Ronald G. The categorization of hyperspectral information (HSI) based on the distribution of spectra in hyperspace[A]. In: Proceedings of SPIE-The International Society for Optical Engineering [C], San Diego, California, USA, 2003,5093:581 ~ 590.
- 3 Zhang Jun-ping, Zhang Ye, Zou Bin, et al. Fusion classification of hyperspectral image based on adaptive subspace decomposition[A]. In: International Conference on Image Processing [C], Vancouver, BC, Canada, 2000,3: 472 ~ 475.
- 4 Petrie G M, Heasler P G, Warner T. Optimal band selection strategies for hyperspectral data sets [A]. In: International Geoscience and Remote Sensing Symposium [C]. Seattle, USA, 1998,3:1582 ~ 1584.
- 5 Millette T L. An expert system approach to spectral band selection for remote sensing analysis [A]. In: International Geoscience and Remote Sensing Symposium [C], Maryland, USA, 1990: 1285 ~ 1288.
- 6 Liu Jian-ping, Zhao Ying-shi. Methods on optimal bands selection in hyperspectral remote sensing data interpretation [J]. Journal of Graduation School, Academia Sinica, 1999,16(2):152 ~ 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):152 ~ 161.]
- 7 Chavez 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.
- 8 Bian Zhao-qi, Zhang Xue-gong. Pattern recognition [M]. Tsinghua University Press, 2000:9 ~ 30. [边肇祺,张学工. 模式识别[M]. 北京:清华大学出版社, 2000:9 ~ 30.]



图1 普通CCD彩色摄像机拍摄的夜间交通图像

Fig.1 Night color traffic scene



(a) 原始图像(晴天)



(b) 原始图像(雨天)

图2 车尾灯原始图像

Fig.2 Image of vehicle tail light

刘春红等：一种新的高光谱遥感图像降维方法



图1 波段50、27、17的假彩色图像

Fig.1 False color image of band 50、27、17

- 玉米
- 牧场
- 草地
- 干草
- 大豆
- 林地



图2 土地资源利用图

Fig.2 Field resource utilized map

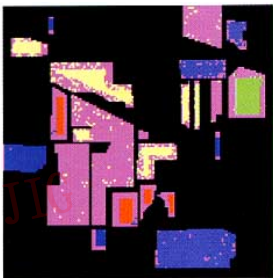


图3 220波段分类图

Fig.3 220 bands classification figure

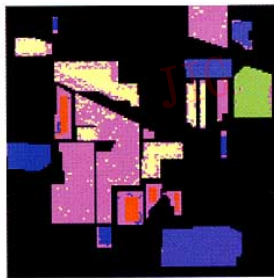


图4 180波段分类图

Fig.4 180 bands classification figure

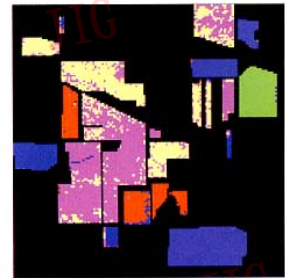


图5 50波段分类图

Fig.5 50 bands classification figure