

一种基于改进子空间划分的波段选择方法

赵慧洁¹, 李明康¹, 李娜¹, 丁昊¹, 蔡辉²

(1. 北京航空航天大学 仪器科学与光电工程学院, 北京 100191;
2. 第二炮兵装备研究院, 北京 100085)

摘要: 高光谱图像具有光谱分辨率高、波段连续、数据量大、图谱合一等特点。然而较高的光谱分辨率会造成波段间相关性强, 信息冗余多。所以如何从数百个高光谱波段中选出有利于识别或分类的波段组合成为了高光谱应用需要解决的问题。文章针对相邻波段间相关性较大的特点, 提出一种改进的对波段相关矩阵进行全局搜索的子空间划分的波段选择方法。该方法克服了传统只利用相关向量对波段进行划分的缺陷, 利用整个相关矩阵进行全局搜索划分, 再在划分后的子空间内进行波段选择, 从而降低了波段之间的相关性。文章最后使用上述方法对 AVIRIS 数据进行波段选择, 并通过 SVM 方法对其进行地物分类, 结果表明该方法较不进行子空间划分的波段选择方法有较高的分类精度。

关键词: 波段选择; 高光谱图像; 子空间划分

中图分类号: TP75 **文献标志码:** A **文章编号:** 1007-2276(2015)10-3155-06

A band selection method based on improved subspace partition

Zhao Huijie¹, Li Mingkang¹, Li Na¹, Ding Hao¹, Cai Hui²

(1. School of Instrumentation Science & Optoelectronics Engineering, Beihang University, Beijing 100191, China;
2. Second Artillery Equipment Institute, Beijing 100085, China)

Abstract: Hyperspectral image has hundreds of successively narrow bands, which brings serious problems such as large correlation and redundant information. The selection of the optimal bands, which are suited for classification or recognition, has become a difficult work that needs to be overcome. In order to solve the problem of the large correlation among bands, a band selection method based on improved subspace partition through global search on correlation matrix was proposed. Through a global search, the band correlation matrix was divided into a series of subspace, from which the optimal bands were finally selected. The proposed method provides a band selection which has small correlation between each other. The result of an experiment which used Support Vector Machine (SVM) on an AVIRIS image shows that the proposed method is valid.

Key words: band selection; hyperspectral image; subspace partition

0 引言

高光谱图像较高的光谱分辨率造成了波段与波段之间的相关性强,信息冗余多的现象。为了解决高光谱遥感图像研究中的上述问题,波段选择方法的研究已经成为当今高光谱数据处理研究的一大课题^[1-2]。

常见的波段选择方法根据是否利用地物的类别信息从而分为基于信息量和基于地物类别两种波段选择方法。基于信息量的波段选择方法不需要利用地物的光谱信息,直接利用原图像计算每个波段所包含的信息总量和波段间的相关性来进行波段选择,从而剔除冗余的波段,达到波段选择的目的。在很多时候,需要利用基于信息量的波段选择方法对高光谱图像进行初步筛选,将初步选择后的波段再结合分类算法进一步进行波段选择,所以,基于信息量的波段选择方法也是一个研究热点。

常用的基于信息量的波段选择方法有基于熵与联合熵的方法,协方差矩阵行列式方法,最佳指数方法,自适应波段选择方法等^[3-6]。这些方法选择出的波段可能出现较为集中的现象,使得波段与波段之间仍有较强的相关性。

基于子空间划分的波段选择方法可以解决选择后波段较为集中的问题^[7-8]。该方法是利用波段与波段之间的相关性先将波段划分为几个连续的子空间。然后在划分出来的每个子空间里进行波段选择,从而克服了选择出的波段较为集中的缺点。

传统的子空间划分方法都是利用近邻相关系数向量 r 中的极小值点对高光谱数据进行划分^[7]或者人为设定阈值选择子空间^[8]。只利用近邻相关系数向量由于没有从全矩阵进行计算,从而会导致部分图像数据划分结果不准确,而人为设定阈值划分主观意愿较强,划分结果受人为选择阈值影响较大,无法实现自适应的划分。因此文中将提出一种利用相关矩阵进行全局搜索从而实现自适应子空间划分的方法。该方法通过对波段相关矩阵进行全局搜索,利用其成块的特点对其进行自适应子空间划分,再在划分后的子空间内利用自适应波段选择方法进行波段选择。这样子空间划分的结果更为准确,而选择后的波段则相关性较小。

1 基于全局搜索的子空间划分方法

该方法先对波段相关矩阵进行二值化预处理,处理后再用全局搜索函数对二值化后的相关矩阵进行搜索,找到最佳划分位置。在搜索时,文中采用了记忆化搜索方法,有效的减少了搜索时间,提高了搜索效率。

1.1 高光谱数据波段相关矩阵

多个波段图像间的相关性表示了每个波段图像在相同空间位置上的像素的相关性,反映了不同波段图像之间的冗余,两波段之间的相关系数定义为:

$$R_{ij} = \frac{E\{(f_i - \bar{f}_i)(f_j - \bar{f}_j)\}}{\sqrt{E\{(f_i - \bar{f}_i)^2\}} \sqrt{E\{(f_j - \bar{f}_j)^2\}}} \quad (1)$$

式中: f_i 为第 i 个波段的像素值; \bar{f}_i 为第 i 个波段的平均像素值; E 为期望。相关系数代表了两个波段的相关程度。其值越大,代表两个波段间相关性越强。通过对任意两个波段的相关系数进行计算,可以得到波段间的相关矩阵。

$$R = \begin{matrix} R(f_1, f_1), R(f_1, f_2), \dots, R(f_1, f_N) \\ R(f_2, f_1), R(f_2, f_2), \dots, R(f_2, f_N) \\ \vdots \\ R(f_N, f_1), R(f_N, f_2), \dots, R(f_N, f_N) \end{matrix} \quad (2)$$

1.2 自适应子空间划分

1.2.1 二值化预处理及阈值判定

利用相关系数的大小差别先对相关矩阵进行阈值处理,将相关矩阵转化为同等大小的二值矩阵,其值分别为 0 和 1。由于子空间划分是对对角线上成块的区域进行划分,所以阈值的选取应该由分块的边界来确定。现使用以下的方法来确定阈值。对每一个对角线上的点做以下操作:找出该点左下方第一个灰度递减变化最大的点,其查找方向如图 1 所示,



图 1 查找阈值

Fig.1 Looking for threshold value

然后将对角线上所有的灰度递减变化最大点求和取平均值,即为阈值。然后再对整个灰度图像做阈值处理,大于阈值的取 1,小于阈值的取 0,这样就产生了一个关于原图像的二值图像。

1.2.2 全局搜索函数

设 $white[i,j]$ 代表相关矩阵上顶点为 (i,i) 与 (j,j) 的正方形内包含白点的个数, $black[i,j]$ 代表相关矩阵上顶点为 (i,i) 与 (j,j) 的正方形内包含黑点的个数。设函数 $f(i,j)$ 表示对顶点为 (i,i) 与 (j,j) 的正方形进行分割后白点个数与黑点个数的最大比值。其分割的主要目的使得该正方形内的白点最多,黑点最少。如图 2 所示划分时,可以使得白色点最多,而黑色点最少,从而将分块的区域划分出来。

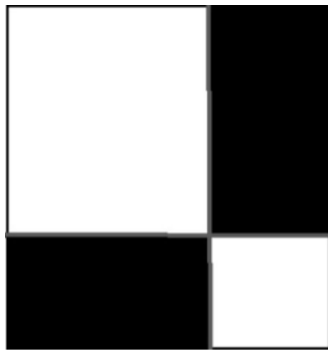


图 2 划分方法

Fig.2 Partition method

$f(i,j)$ 的递归方程如下:

$$f(i,j)=\max\{f(i,k)+f(k+1,j),f_{\text{first}}(i,j)\}$$

$$i < k < j - \text{minblock}$$

$$1 < i < N - \text{minblock}, i + \text{minblock} < j < N \quad (3)$$

式中: k 表示从第 k 个波段进行分割; minblock 代表最小的分块区间长度。 $f_{\text{first}}(i,j)$ 表示对第 i 个波段到第 j 个波段不进行分割的白点与黑点的原始比值。而 $f(i,j)$ 的直接计算公式如下:

$$f(i,j)=\max\left\{\frac{\text{white}[i,k]+\text{white}[k+1,j]}{\text{black}[i,k]+\text{black}[k+1,j]}, \frac{\text{white}[i,j]}{\text{black}[i,j]}\right\}$$

$$i < k < j - \text{minblock} \quad (4)$$

$$\text{white}[i,j]=\text{white}[i,k_{\text{max}}]+\text{white}[k_{\text{max}},j] \quad (5)$$

$$\text{black}[i,j]=\text{black}[i,k_{\text{max}}]+\text{black}[k_{\text{max}},j] \quad (6)$$

式中: k_{max} 为比值最大的划分波段。递归的边界条件为 $(j-i)=\text{minblock}$ 时,直接返回 $f_{\text{first}}(i,j)$ 的值。利用递归方式对相关矩阵进行全局搜索,得出所有的 k_{max} ,

将其记录下来,就能得到所有的划分区间。

1.2.3 记忆化搜索

在递归调用 f 函数进行运算时,会出现多次调用相同函数的情况。如在计算 $f(1,8)$ 的时候调用了 $f(1,5)$,而在计算 $f(1,9)$ 时候又再次调用了 $f(1,5)$ 。这种重复调用相同的函数会浪费大量时间。因此文中提出利用记忆化搜索方法,使用一个数组 $f[i,j]$ 记录调用函数 $f(i,j)$ 后计算出来的值。每当需要调用 $f(i,j)$ 的时候,预先判断 $f[i,j]$ 是否已经记录值,如果有记录则直接返回该值,若无,则再进行 $f(i,j)$ 的计算。这样就避免了重复运算,每个 i 和 j 对应的 f 函数只需计算一次,从而加快了整个全局搜索的速度。

2 自适应波段选择方法

(1) 子空间划分之后,在每个子空间里分别进行波段选择。由于自适应波段选择方法考虑了空间相关性以及谱间相关性,可以利用该方法进行波段选择^[9]。

根据自适应波段算法准则,该算法的数学模型为:

$$\text{Index}_i = \frac{\sigma_i}{(R_{i-1,i} + R_{i,i+1})/2} \quad (7)$$

$$\sigma_i = \left[\frac{1}{M \times N} \sum_{x=1}^M \sum_{y=1}^N (f_i(x,y) - \bar{f}_i)^2 \right]^{\frac{1}{2}} \quad (8)$$

式中: σ_i 为第 i 个波段的标准差; $R_{i-1,i}$ 为第 i 波段与前一个波段的相关系数。相关系数越小,两波段的独立性越大。 Index_i 是第 i 波段的指数,公式(1)和公式(8)给出了相关系数和标准差的求法。其中, M, N 为图片的行、列像素个数。

自适应波段选择算法充分考虑了信息量以及任意两波段之间的相关性。波段指数越大,相应图片的信息量越大。最终获取波段选择方法为选择波段指数最大的前 n 波段。

3 实验结果与讨论

3.1 子空间划分实验

实验数据采用 ENVI 软件自带的 AVIRIS “Effort” Corrected ATREM 局部图像,该数据空间维大小为 400×350 像素,共 50 个波段,波段范围从 $1.99 \sim 2.47 \mu\text{m}$ 。

图 3(a)为文中提出的方法的划分结果,而图 3(b)

为传统只利用相关向量 r 的划分结果。从图 3 中可以看出传统方法由于根据相关向量的极值判断,从而划分出的子空间小而多,与分块特性不相符。而文中方法充分考虑了整个相关矩阵的分块特性,自适应划分出的子空间则与分块特性较为相符,证明了文中方法具有更好的子空间划分能力。

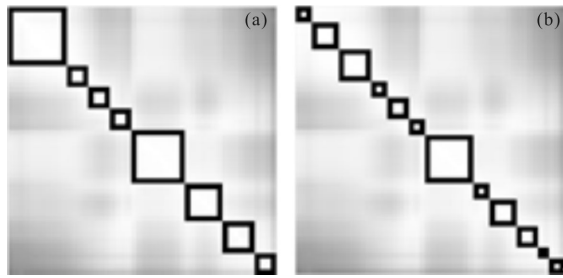


图 3 AVIRIS“Effort”子空间划分对比结果

Fig.3 Comparison result of the AVIRIS "Effort" subspace partition

3.2 分类验证实验

实验数据采用 AVIRIS 92AV3C 高光谱数据,此数据在研究波段选择和地物分类时被广泛采用^[10-11]。该数据空间维大小为 $145 \text{ pixel} \times 145 \text{ pixel}$,共有 220 个波段,波段范围从 $0.4 \sim 2.5 \mu\text{m}$ 。该数据还有一个参考地图(Reference Map),该地图上显示了部分的地物类别信息,可以判别某个像素隶属于哪一个类别。参考地图的值的范围为 $0 \sim 16$,0 为背景,1~16 为 16 个类别。这 16 个类别共有像素 10366 个,约占总像素的 49.7%。图 4 为该数据分布图和该数据的参考地图。

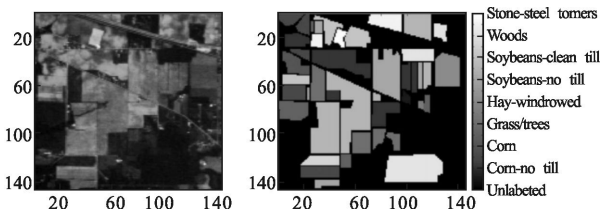


图 4 AVIRIS 92AV3C 图像

Fig.4 Image of AVIRIS 92AV3C

对该协方差矩阵利用文中提出的子空间划分方法进行分割后,得到如下 5 个子空间(1, 107), (108, 121), (122, 143), (144, 182), (183, 220)。

在选择出来的五个子空间里利用自适应波段选择方法计算每个波段的 Index 值,再在每个子空间内分别对计算后的 Index 值进行排序,然后再根据需求选择前 n 个波段进行实验。实验采用的分类算法为

支持向量机。支持向量机是一个监督分类算法,需要对其提供样本进行训练。该实验将该图像中的 16 类地物的样本数分别取其一半当作训练样本(选取的时候利用随机方法),剩下的样本则当作测试样本。由于训练样本是随机选取,文中对该实验进行两次,分别得到如下的分类结果。表 1 中实验一与实验三为两次随机选取训练样本和测试样本后利用文中提出的方法选取波段后实验得到的分类结果,而实验二和实验四则是分别对应于实验一和实验三同样的样本数据下不进行子空间划分而直接对波段进行选择后的分类结果。

表 1 分类实验结果

Tab.1 Result of the classification

Number of bands remained	220	70	41
Experiment 1	87.4%	85.6%	82.9%
Experiment 2	87.4%	82.35%	74.75%
Experiment 3	89.3%	84.9%	80.75%
Experiment 4	89.3%	80.6%	73.9%

从实验结果可以看出,在实验一与实验三中,当波段数为 41 的时候有 80% 左右的分类准确率,分类准确率下降了 7% 左右,由于实验是对 16 类进行分类,所以该值还是具有一定的适用性。通过对比四组实验,可以发现实验二和实验四中不进行子空间划分而直接进行波段选择后的分类结果较实验一和实验三差。实验一与实验二分类精度对比结果如图 5 所示,实验三与实验四的对比结果如图 6 所示。该实验可以说明文中提出的波段选择方法具有较好的分类效果。

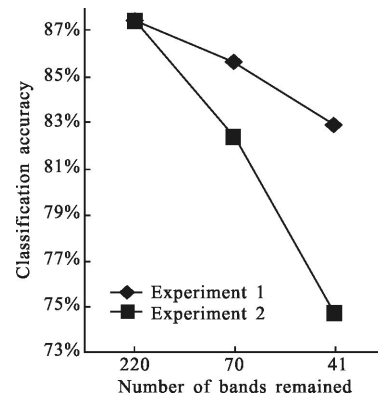


图 5 实验一与实验二分类精度对比结果

Fig.5 Classification comparison of experiment 1 and experiment 2

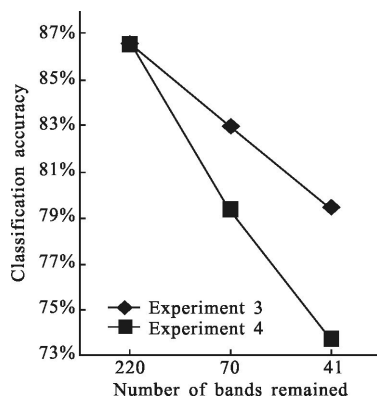


图 6 实验三与实验四分类精度对比结果

Fig.6 Classification comparison of experiment 3 and experiment 4

在上述实验的基础上,进一步分析选择后波段间相关性,实验结果如表 2 所示。

表 2 所保留波段的平均相关系数

Tab.2 Mean correlation coefficient of the selected bands

Number of bands remained	70	41
Experiment 1	0.814 5	0.815 3
Experiment 2	0.886 6	0.890 2

表 2 表示实验一实验二所保留波段的平均波段相关系数。可以看出文中所提方法选择出的平均波段相关系数小于不经过子空间划分的平均波段相关系数,从而说明利用文中方法选择出来的波段相关性较小,证明该了方法的有效性。

4 结 论

文中针对相邻波段间相关系数较大的情况,通过对波段间相关矩阵进行分析,提出了一种基于全局搜索的子空间划分的波段选择方法。由于高光谱拥有大量的波段,从而使得相邻波段间具有较强的相关性,进而对波段选择造成了一定的难度。而波段相关矩阵具有成块的特点,通过对相关矩阵进行分析,对波段进行子空间划分,能有效的消除波段选择后波段之间的相关性。传统的子空间划分方法只利用了相邻波段间的相关系数,或者划分方法较为主观。文中提出的子空间划分方法通过对相关矩阵进行全局搜索,从而进行子空间划分。子空间划分后,再在每个子空间内进行波段选择,这样选择后的波

段之间的相关性较小,所包含的信息量也越大。从实验结果上来看,可以看出文中提出的方法在波段选择上具有较好的效果。

由于文中在进行波段选择时没有考虑物体的类别信息,在波段选择后续进行分类或识别的时候,可能会限制分类或识别的精度,这在以后的工作中将是另外一个值得研究的方向。

参考文献:

- [1] Wang Jingli. Maximum band screening and its application to hyperspectral target detection [J]. *Infrared and Laser Engineering*, 2012, 41(6): 1514–1519. (in Chinese)
王静荔. 波段最大筛选法及其在高光谱目标探测中的应用 [J]. *红外与激光工程*, 2012, 41(6): 1514–1519.
- [2] Wang Yiting, Huang Shiqi, Liu Daizhi, et al. Novel band selection method based on target detection [J]. *Infrared and Laser Engineering*, 2013, 42(8): 2294–2298. (in Chinese)
王艺婷, 黄世奇, 刘代志, 等. 一种新的基于目标检测的波段选择方法 [J]. *红外与激光工程*, 2013, 42(8): 2294–2298.
- [3] Bajcsy P, Groves P. Methodology for hyper spectral band selection photogram metric engineering and remote sensing [J]. *Photogrammetric Engineering and Remote Sensing Journal*, 2004, 70(7): 793–802.
- [4] Liu Jianping, Zhao Yingshi, Sun Shuling. Experimental studies about methods on optimal bands selection of hyperspectral remote sensing datasets [J]. *Remote Sensing Technology and Application*, 2001, 16(1): 7–13. (in Chinese)
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 [J]. *遥感技术与应用*, 2001, 16(1): 7–13.
- [5] Yang Jinhong. Optimal band selection methods of hypersepctral remote sensing data [D]. Nanjing: Nanjing University of Information Science&Technology, 2005.(in Chinese)
杨金红. 高光谱遥感数据最佳波段选择方法研究 [D]. 南京: 南京信息工程大学, 2005.
- [6] 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.(in Chinese)
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.
- [7] Gu Yanfeng, Zhang Ye. Feature extraction based on automatic subspace partition for hypersepctral images [J]. *Remote Sensing Technology and Application*, 2003, 18(6):

- 32-35. (in Chinese)
- 谷延锋, 张晔. 基于自动子空间划分的高光谱数据特征提取[J]. 遥感技术与应用, 2003, 18(6): 32-35.
- [8] Dong Yanhua. Research on key technology of hyperspectral remote sensing image processing [D]. Harbin: Harbin University of Science and Technology, 2006. (in Chinese)
- 董延华. 超光谱遥感图像处理关键技术研究 [D]. 哈尔滨: 哈尔滨理工大学, 2006.
- [9] Han Ruimei, Yang Minhua. Study on an improved method of band selection of hypersepctral remote sensing data [J]. *Geomatics & Spatial Information Technology*, 2010, 33(3): 137-139. (in Chinese)
- 韩瑞梅, 杨敏华. 一种改进的高光谱遥感数据波段选择方法的研究[J]. 测绘与空间地理信息, 2010, 33(3): 137-139.
- [10] Guo Baofeng, Gunn Steve R, Damper R I. Band selection for hyperspectral image classification using mutual information [J]. *IEEE Geoscience and Remote Sensing Letters*, 2006, 3(4): 522-526.
- [11] Wang Ligu, Wei Fangjie. Band selection for hypersepctral imagery based on combination of genetic algorithm and ant colony algorithm[J]. *Journal of Image and Graphics*, 2013, 18(2): 235-242. (in Chinese)
- 王立国, 魏芳洁. 结合遗传算法和蚁群算法的高光谱图像波段选择[J]. 中国图象图形学报, 2013, 18(2): 235-242.