

基于类别可分性的高光谱图像波段选择

许明明^{1,2} 张良培¹ 杜博² 张乐飞²

(武汉大学测绘遥感信息工程国家重点实验室 武汉 430079)¹ (武汉大学计算机学院 武汉 430079)²

摘要 高光谱遥感数据具有丰富的光谱信息,应用十分广泛,但其冗余的光谱信息有时会限制高光谱图像的分类精度以及计算复杂度。为了提高解译效率,高光谱图像降维不可或缺,这也是高光谱图像处理的研究热点之一。提出了一种基于类别可分性的高光谱图像波段选择方法(Endmember Separability Based band Selection, ESBB),该方法通过 Mahalanobis 距离最大化图像中各类地物的可分性来确定最优的波段组合。相较于其他监督波段选择算法,该方法不需要大量训练样本,不用对每个组合做分类处理。对波段选择后的结果进行分类的实验结果证明,该方法是一个快速有效的波段选择方法,可以得到一个较好的分类精度。

关键词 监督波段选择,高光谱图像,类别可分

中图法分类号 TP751 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.4.056

Supervised Band Selection Based on Class Separability for Hyperspectral Image

XU Ming-ming^{1,2} ZHANG Liang-pei¹ DU Bo² ZHANG Le-fei²

(The State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China)¹
(School of Computer, Wuhan University, Wuhan 430079, China)²

Abstract Hyperspectral remote sensing data contain abundant spectral information, which are widely used, but sometimes the redundant spectral information limits the classification accuracy and computational complexity. In order to improve the interpretation efficiency, hyperspectral image dimension reduction is necessary, which is also one of the highlights in the hyperspectral image processing. This paper presented a hyperspectral image band selection method based on endmember separability (ESBS). This method maximizes class separability using Mahalanobis distance to determine the optimal band combination. Compared with other supervision of band selection algorithms, the proposed method does not need the training samples, and does not conduct classification during band selection. The experiment results show that the proposed method is effective and can get better classification accuracy.

Keywords Supervised band selection, Hyperspectral image, Class separability

1 引言

高光谱图像波段数多,导致数据存储所需空间大,图像处理时间长,精度下降。为了减少高光谱图像中的冗余信息,图像分析之前^[1,2],进行降维处理是有必要的。特征变换和波段选择是高光谱图像的两类降维方法。基于特征变换的降维方法(例如 PCA)改变了原始图像的物理意义,相比之下,波段选择是从高光谱图像所有波段中选择起主要作用的波段子集,既能大大降低高光谱图像的数据维数,又能比较完整地保留感兴趣的信息。本论文的研究重点在于波段选择。

高光谱图像的波段选择是一种组合优化问题,性能较好的波段组合具有信息量大、相关性小、类别可分性好的特点。有一类监督的波段选择算法的目标是最大化类间距离,类间距离的测度包括:散度、巴氏距离、JM 距离^[3-5]等。但是,这类算法需要充足的训练样本来正确估计各类的统计信息,很多情况下,我们并没有足够的训练样本,导致距离测度产生很大误差。还有一类算法是通过最大化分类精度来进行最优波段

的选择,这类算法有一个缺点,即每次组合都要进行分类,以得到分类精度进行比较,效率比较低。对于波段选择方法的搜索策略,许多文献都有介绍^[6-8]。本文为了避免穷尽搜索,简化计算,使用序贯前向选择法(SFS)进行波段的搜索。

针对上述波段选择算法存在的缺点,本文提出了一种基于类别可分性的高光谱图像波段选择,该方法通过 Mahalanobis 距离(马氏距离)最大化图像中各类地物的可分性来确定最优的波段组合。该方法与以往方法不同的地方在于:不需要大量训练样本,不需要对每个组合做分类处理,而且无需单独的初始波段选择,只需所有地物类别的端元光谱信息。端元光谱信息可以通过现有的自动端元选择算法^[9]提取或者人工在图像中选择,不会因为样本不足导致效果变差。最后的实验验证了所提的方法的有效性。

2 基于类别可分性的高光谱图像波段选择

2.1 Mahalanobis 距离(马氏距离)

马氏距离是由印度统计学家马哈拉诺比斯(P. C. Maha-

到稿日期:2014-04-27 返修日期:2014-07-01 本文受国家“973 计划”项目(2012CB719905),国家自然科学基金资助项目(61102128),中央高校基本科研业务费专项资金(211-274175)资助。

许明明(1990-),女,博士生,主要研究方向为高光谱混合像元分解, E-mail: xumingming900405@126.com; 张良培(1962-),男,博士,教授,主要研究方向为测绘、遥感图像处理、人工智能、模式识别; 杜博(1983-),男,博士,副教授,主要研究方向为模式识别、计算机视觉和图像处理; 张乐飞(1986-),男,博士,讲师,主要研究方向为模式识别、图像处理。

lanobis)提出的^[10],表示数据的协方差距离。它是一种有效的计算两个未知样本集的相似度的方法。与欧氏距离不同的是,它考虑到总体的变异(方差)对距离远近的影响,并且是尺度无关的。

马氏距离也可以定义为两个服从同一分布并且其协方差矩阵为 Σ 的随机变量的差异程度^[11]:

$$d(x_i, x_j) = \sqrt{(x_i - x_j)^T \Sigma^{-1} (x_i - x_j)}, i \neq j \quad (1)$$

式中, x 是通过自动端元选择算法提取或者人工在图像中选择出来的不同类别的端元光谱曲线, Σ 为整个数据集的协方差矩阵。为了减少计算量,也可使用部分有代表性的数据集计算协方差矩阵。

2.2 算法描述

基于类别可分性的高光谱图像波段选择的算法流程如下(使用序贯前向选择法(SFS)进行波段搜索):

(1)数据归一化;

(2)对每一个波段分别计算每两个类别之间的马氏距离,将各类别间的马氏距离取均值。得到每个波段上各类别间的马氏距离均值,能够使各类别间的马氏距离均值最大的波段即为第一个被选择出来的波段 $B1$,此时已选择的波段集合记做 $\Phi^S = \{B1\}$ 。

(3)将剩下的波段依次并入到已选择的波段集合中,计算当前波段集合各类别间的马氏距离均值。同样地,能够使各类别间的马氏距离均值最大的波段即为第二个被选择出来的波段 $B2$,此时已选择的波段集合记做 $\Phi^S = \Phi^S \cup \{B2\}$ 。

(4)继续重复步骤(3),直到已选择的波段的数目达到要求为止。

对于具有 L 个波段 p 类地物的高光谱影像,为了选择第一个波段,我们需要计算 $L * p * (p-1)/2$ 次距离(见式(1)),此时端元光谱曲线只有一个数值,协方差矩阵退化为方差。当选择第二个波段时,我们需要计算 $(L-1) * p * (p-1)/2$ 次距离,此时的端元光谱曲线为 $2 * 1$ 的向量,协方差矩阵大小为 $2 * 2$ 。依次类推。假设选择 m 个波段,那么,算法中基本语句的语句频度为 $T = p * (p-1) * L * ((m(m+1))^2)/8$ 。

从上述算法流程可以看出,该方法不需要大量训练样本,不需要对每个组合做分类处理,而且不需要单独的初始波段选择,只需要所有地物类别的端元光谱信息。端元光谱信息可以通过现有的自动端元选择算法提取或者人工在图像中选择,不会因为样本不足导致效果变差。由于提出的算法主要是基于端元的可分性,因此提出的算法命名为ESBB。

2.3 算法性能评价

为了评价所选择的波段集合所含有的类别信息以及类别的可分性信息,对最优波段组合分类处理,并与使用全部波段的分类结果进行比较。本文采用的是最大似然分类算法。此外,为了证明所提算法的有效性,本文选取两种常用的监督波段算法进行对比。

(1)基于JM距离的波段选择算法

Jeffries-Matusita距离同时兼顾一次统计变量和二次统计变量,在测度高光谱多维空间中两类统计距离时具有较好的效果。因此,采用各地物间平均JM距离作为目标函数可以获得波段选择的结果。

两类地物间的 x_i 与 x_j 间的JM距离定义为:

$$J_{i,j} = \int \{ \sqrt{p(r|x_i)} - \sqrt{p(r|x_j)} \}^2 dx \quad (2)$$

式中, $p(r|x_i)$, $p(r|x_j)$ 为两个类别的条件概率分布。

当每个类别的分布符合高斯分布时,JM距离可以简化为:

$$J_{i,j} = 2(1 - e^{-B_{i,j}}) \quad (3)$$

式中:

$$B_{i,j} = \frac{1}{8} (\mu_i - \mu_j)^T \left(\frac{\Sigma_i + \Sigma_j}{2} \right)^{-1} (\mu_i - \mu_j) + \frac{1}{2} \ln \left(\frac{|\Sigma_i + \Sigma_j|/2}{|\Sigma_i|^{1/2} |\Sigma_j|^{1/2}} \right) \quad (4)$$

上式也称为巴氏距离, μ_i, μ_j 是每类地物的样本均值, Σ_i, Σ_j 是每类地物的样本方差。

从上式也可以看出,对于高光谱影像,需要足够的样本才能对每一类的均值和方差进行准确的估计。所以该方法受训练样本的限制,训练样本不足会导致统计信息估计不准确,波段选择的效果也会受到影响。

(2)估计组分方差最小方法(MEAC)

不同于式(1)与式(2),MEAC方法的目标函数为^[12]:

$$\arg \min \{ \text{trace} [(\hat{S}^T \hat{S})^{-1}] \} \quad (5)$$

式中, \hat{S} 为端元光谱集合。

MEAC方法每次迭代不是计算类间距离,而是计算估计组分方差,选择估计组分方差最小的波段组合。该方法同样不需要训练样本,不需要对每个组合做分类处理,但是受初始波段选择的影响,不同的初始波段最终的结果不同。

3 实验结果与分析

实验所采用的RODIS数据为意大利北部Pavia大学校园的高光谱图像,空间分辨率为1.3m,图像的大小为 610×340 像素。在原始图像中,去除水汽吸收和低信噪比的波段,参与处理的图像波段数为103,实验所用图像及真实地面分布如图1所示。

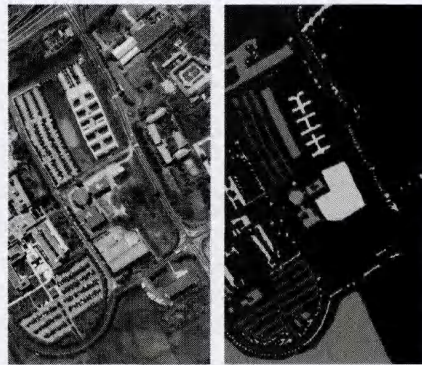


图1 原始影像及真实地面分布

对于ESBB方法以及MEAC方法,只需要从训练样本的每类地物中随机选取1条;对于JM距离方法,为了保证统计的准确性,得到较好的波段选择效果,从每类地物随机选取300条光谱曲线。使用最大似然分类法(每种地物的训练样本为300,其他的作为测试样本)对输出的最优波段组合图像进行分类,求出其总体分类精度(Overall Accuracy, OA)(计算公式如式(6)所示)来表征波段组合的效果,分类精度越高,说明波段组合的效果越好。同时,将选择的波段数目设置为10:5:50,用以讨论波段数目的变化对分类结果的影响。对各方法波段选择的结果进行分类的结果如图2所示(多次试验取均值),其中最下面的直线表示使用所有波段进行分类的

(下转第296页)

- [8] 黄远程. 高光谱影像混合像元分解的若干关键技术研究[D]. 武汉: 武汉大学, 2010
- [9] Manolakis D, Shaw G. Detection algorithms for hyperspectral imaging applications [J]. Signal Processing Magazine, IEEE, 2002, 19(1): 29-43
- [10] Harsanyi J C, Farrand W H, Chang C I. Detection of subpixel signatures in hyperspectral image sequences[C]// Proceedings of the American Society for Photogrammetry and Remote Sensing, 1994: 236-247
- [11] Scharf L L, McWhorter L T. Adaptive matched subspace detectors and adaptive coherence estimators[C]// Conference Record of the Thirtieth Asilomar Conference on Signals, Systems and Computers, 1996. IEEE, 1996: 1114-1117
- [12] Park B, Windham W R, Lawrence K C, et al. Contaminant classification of poultry hyperspectral imagery using a spectral angle mapper algorithm [J]. Biosystems Engineering, 2007, 96 (3): 323-333
- [13] Chang C I. Orthogonal subspace projection (OSP) revisited: a comprehensive study and analysis[J]. IEEE Transactions on Geoscience and Remote Sensing, 2005, 43(3): 502-518
- [14] 刘凯, 张立福, 杨杭, 等. 面向对象分析的非结构化背景目标高光谱探测方法研究[J]. 光谱学与光谱分析, 2013, 33(6)
- [15] Fuhrmann D R, Kelly E J, Nitzberg R. A CFAR Adaptive-Matched Filter Detector [J]. IEEE Transaction on Aerospace and Electronic Systems, 1992, 28(1): 208-216
- [16] Matteoli S, Acito N, Diani M, et al. An automatic approach to adaptive local background estimation and suppression in hyperspectral target detection [J]. IEEE Transactions on Geoscience and Remote Sensing, 2011, 49(2): 790-800
- [17] Bioucas-Dias J M, Nascimento J M P. Hyperspectral subspace identification [J]. IEEE Transactions on Geoscience and Remote Sensing, 2008, 46(8): 2435-2445
- [18] 王千, 王成, 冯振元, 等. K-means 聚类算法研究综述[J]. 电子设计工程, 2012, 20(7): 21-24
- [19] DeLong E R, DeLong D M, Clarke-Pearson D L. Comparing the areas under two or more correlated receiver operating characteristic curves; a nonparametric approach [J]. Biometrics, 1988, 44 (3): 837-845

(上接第 275 页)

分类精度, 该分类精度可以作为一个标尺, 判断选择的波段子集能否在一定程度上消除全波段的干扰。

$$OA = \frac{\sum_{k=1}^n p_{kk}}{p} \quad (6)$$

其中, n 为图像中包含的类别的数目, p_{kk} 为第 k 类被正确分类的像元数目, p 为像元总数。

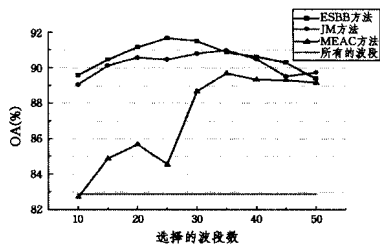


图2 不同波段选择方法不同波段数目的分类精度

从上述的实验结果可以看出: (1) 分类精度随着波段数目的增加, 总体呈现先上升后下降的趋势; (2) 选择的波段子集能够得到比全部波段要高的分类精度; (3) ESBB 方法在 25 个波段时达到最高分类精度 91.67%, 而 JM 算法在 35 个波段得到 90.97% 的分类精度, 本文方法利用较少的波段数目达到了最高的分类精度。

从上述的实验现象可以发现: 并不是波段越多, 分类精度越高, 过高的波段数目反而会增加类内的变化, 从而增加分类的难度, 因此, 快速有效的波段选择算法十分重要。本文提出的基于马氏距离的波段选择算法可以满足应用需求。

结束语 高光谱传感器设计之初是为了保证各类地物的准确的表达, 因为对于不熟悉的地物, 我们并不清楚利用哪些波长可以完整地描述该地物的特征光谱, 所以高光谱传感器将整个波长范围近乎均匀划分, 得到信息丰富的高光谱曲线。但是, 当我们确定图像中的地物类别以后, 并不需要所有的波段来表征地物特征, 只需要某些特定波长的反射率即可将图像中的各类地物完美区分, 此时, 过多的波段反而制约了图像的解译, 因此高光谱图像的波段选择十分重要。

基于此, 本文提出的基于类别可分性的高光谱图像波段选择算法可以快速有效地进行波段的选择, 与其他的监督方

法相比, 该方法不需要大量训练样本, 不需要对每个组合做分类处理, 而且不需要单独的初始波段选择, 鲁棒性强。通过实验对比, 本文提出的算法可以达到较高的精度, 是一种有效的波段选择算法。

参考文献

- [1] 杜博, 张良培, 李平湘, 等. 一种基于随机场模型的高光谱影像目标探测算法 [J]. 计算机科学, 2010, 37(6): 289-292
- [2] 李洪利, 沈焕峰, 杜博, 等. 一种高保真同态滤波遥感影像薄云去除方法 [J]. 遥感信息, 2011(1): 41-44
- [3] Harraguerri A. Visual method for spectral band selection [J]. IEEE Geoscience and Remote Sensing Letters, 2004, 1(2): 101-106
- [4] Huang Rui, He Ming-yi. Band selection based on feature weighting for classification of hyperspectral imagery [J]. IEEE Geoscience and Remote Sensing Letters, 2005, 2(2): 156-159
- [5] Backer S D, 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
- [6] 吴昊, 李士进, 林林, 等. 多策略结合的高光谱图像波段选择新方法 [J]. 计算机科学与探索, 2010, 4(3): 465-468
- [7] 苏红军, 杜培军, 盛业华. 高光谱影像波段选择算法研究 [J]. 计算机应用研究, 2008, 25(4): 1093-1096
- [8] 葛亮, 王斌, 张立明. 基于偏最小二乘法的高光谱图像波段选择 [J]. 计算机辅助设计与图形学学报, 2011, 23(11): 1844-1852
- [9] Xu Ming-ming, Du Bo, Zhang Liang-pei. Spatial-Spectral information based abundance-constrained endmember extraction methods [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2014, 7(8): 2004-2015
- [10] Mahalanobis P C. On the generalised distance in statistics [C]// Proceedings of the National Institute of Science of India, 1936, 12: 49-55
- [11] De Maesschalck R, Jouan-Rimbaud D, Massart D L. The mahalanobis distance [J]. Chemometrics and intelligent laboratory systems, 2000, 50(1): 1-18
- [12] Yang He, Du Qian, Su Hong-jun, et al. An efficient method for supervised hyperspectral band selection [J]. IEEE Geoscience and Remote Sensing Letters, 2011, 8(1): 138-142