# 基于分类别 PCA 散度的高光谱图像分类波段选择

黄 睿 何明一
(西北工业大学电子信息学院 陕西省信息获取与处理重点实验室 西安 710072)

**摘 要**：波段选择是去除高光谱图象段间冗余，实现降维的有效方法。该文提出了一种新的基于分类别主成分分析(PCA)散度的波段选择方法。即首先对训练集各类样本分别进行 PCA 变换去相关并计算散度，接着分析相应 PCA 变换系数获得对各类样本分类都重要的原始波段，在综合考虑波段的相关度，散度和子集规模的基础上获得最终选择波段。复杂度分析表明该方法较局部寻优的前向搜索计算量大为降低，提高了效率，并用高光谱遥感图象的分类实验进行了验证。

**关键词**：高光谱图像分类，波段选择，分类别 PCA，散度
**中图分类号**: TP751.1 **文献标识码**：A **文章编号**：1009-5896(2005)10-1588-05

# Band Selection Using Divergence of Class-within PCA in Hyperspectral Images Classification

Huang Rui He Ming-yi
(*Shanxi Key Lab of Information Acquisition and Processing, Northwestern Polytechnical University, Xi'an*710072, *China*)

**Abstract** Band selection from multispectral or hyperspectral image data is an effective method to remove redundancy among bands and thus reduce dimension. An efficient algorithm using divergence based class-within principal component analysis (PCA) and analysis of corresponding coefficients is proposed. At first, the covariance of each class is diagonalized through PCA transforms on class data respectively, and then the divergence only depends on the summation of individual feature separability of transformed bands. Secondly, after an analysis of corresponding PCA transform coefficients, the candidate bands, original bands essential to classification, are determined by majority vote. At last, the final band subset is obtained by analyzing the dependency and divergence of bands in every subset generated according to the correlations of original band in candidates. Compared with sequential forward selection, the proposed method reduces the computation complexity, and encouraging results have been shown by experiments with an Airborne Visible/InfraRed Imaging Spectrometer (AVIRIS) data set.

**Key words** Hyperspectral image classification, Band selection, Class-within principal component analysis, Divergence

## 1 引言

高光谱遥感技术的发展为人们提供了比多光谱传感器更为详尽精确的地物光谱信息，但如何对这成百波段的高维数据进行有效处理，克服“维数灾难”又成为难点。利用高光谱段间存在大量冗余的特点进行降维处理，针对应用目的(如分类、识别、压缩等)获取有效特征，降低计算量，是解决高维数据问题的重要方法之一。数据降维可通过特征提取和特征选择来实现，在高光谱数据处理中则对应波段提取和波段选择[1]。波段提取通过由高维向低维的投影变换实现降维，该投影应尽量保留有用信息；波段选择则是根据搜索策略在原始波段空间寻找满足某准则函数的波段子集来达到目的。严格说来，波段选择是波段提取的特例，但与波段提取不同的是它不进行投影变换，保持了原始波段的物理含义，展现了地物的光谱特性，而这些往往更令领域专家感兴趣。

总的说来，波段选择有两种思路。一是将其作为最优特征子集选择问题，根据选择准则进行全局或局部搜索[2,3]，其中，选择准则要尽可能地保留所需信息。对于分类问题，考虑类的可分离性，可以多种距离(如B距离、JM距离、M距离等)或信息论方法(如散度、互信息等)为衡量准则，其中，在假设数据服从高维正态分布时，散度计算更为准确方便[4]。对子集的搜索利用穷举法可得到最优解[5]，但其计算量常让

2004-07-07 收到， 2004-12-08 改回
航空科学基金(01F53028)资助课题

人难以承受，因此实际应用中经常采用次优的前向搜索法。与前者相比，虽然后者的计算量得到大幅降低，但随着选择波段数和类别数的增长快速增长。通过正交变换，如主成分分析(PCA)，噪声调节主成分分析(NAPCA)、Fisher判别分析(DA)等，利用特征分析(eigenanalysis)得到波段子集[1,6]则是另一种思路。

本文提出一种基于分类别 PCA 变换散度及相应系数分析的波段选择方法。在训练集中对各类别数据分别作 PCA 变换去相关，得到变换后新波段组合的散度函数，其中各类别协方差阵对角化，因此散度值仅由各新波段散度的简单加和决定。分析相应的 PCA 变换系数，通过投票机制得到对各类数据分类都有重要作用的原始波段作为候选集。根据波段相关度可将得到的波段候选集分为若干个子集合，在综合考虑波段的相关度，散度和子集的规模基础上确定最终选择波段。本方法计算量主要集中在 PCA 变换，分析表明，当选择波段数不是过小时(选择波段数约小于波段总数和类别数之比的三次方根)，与前向搜索相比，在计算量上有显著优势。高光谱遥感图像的分类实验证明了方法的有效性。

在本文余下部分中，将逐一对新方法的 3 个步骤进行介绍，并进行计算复杂度分析，最后利用高光谱数据集与常用的等间隔选择和前向选择进行性能比较。

## 2　波段选择的新方法

### 2.1　分类别 PCA 变换后的散度准则

设样本集 $\boldsymbol{X}=[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_N]$，其中 $\boldsymbol{x}_j\in\mathbb{R}^L$，$L$ 为波段总数，$N$ 为样本数。样本集由 $m$ 类构成，即 $\boldsymbol{X}=\bigcup_{c=1}^{m}\boldsymbol{X}^{(c)}$，$\boldsymbol{X}^{(c)}=\left[\boldsymbol{x}_1^{(c)},\boldsymbol{x}_2^{(c)},\cdots,\boldsymbol{x}_{N_c}^{(c)}\right]$ 为各类数据集合，$N_c$ 为类 $c$ 数据个数，满足 $\sum_{c=1}^{m}N_c=N$。

散度是常用的衡量类别可分度的准则，$m$ 类数据平均散度为

$$D_{\mathrm{AVE}}=\frac{2}{m(m-1)}\sum_{i=1}^{m}\sum_{j=i+1}^{m}D_{ij} \tag{1}$$

其中 $D_{ij}$ 表示类 $i$ 和类 $j$ 的散度，一般假设数据服从正态分布，有

$$D_{ij}=\frac{1}{2}\mathrm{tr}\left(\left(\boldsymbol{\Sigma}_i-\boldsymbol{\Sigma}_j\right)\left(\boldsymbol{\Sigma}_j^{-1}-\boldsymbol{\Sigma}_i^{-1}\right)\right)+\frac{1}{2}\mathrm{tr}\left(\left(\boldsymbol{\Sigma}_i^{-1}+\boldsymbol{\Sigma}_i^{-1}\right)\left(\boldsymbol{u}_i-\boldsymbol{u}_j\right)\left(\boldsymbol{u}_i-\boldsymbol{u}_j\right)^{\mathrm{T}}\right) \tag{2}$$

其中 $\mathrm{tr}(\cdot)$ 表示矩阵的迹；$\boldsymbol{\Sigma}_i$，$\boldsymbol{u}_i$，$\boldsymbol{\Sigma}_j$，$\boldsymbol{u}_j$ 分别为类 $i$ 与类 $j$ 的协方差矩阵和均值。为了抑制式(1)中较大的 $D_{ij}$ 对平均散度的影响，常采用变换的散度(Transformed Divergence):

$$\mathrm{TD}_{ij}=1-\exp\left(-D_{ij}/8\right) \tag{3}$$

根据式(2)选择波段子集需要计算每种波段组合方式的散度值，计算量很大。通过对各类数据分别作 PCA 变换去相关，使各类的协方差阵对角化，从而使波段子集的类别可分性等同于子集内各波段类别可分性的加和。对各类作 PCA 变换，有

$$\boldsymbol{Y}^{(c)}=\boldsymbol{P}^{(c)^{\mathrm{T}}}\boldsymbol{X}^{(c)},\quad c=1,\cdots,m \tag{4}$$

其中，$\boldsymbol{P}^{(c)}=\left[\boldsymbol{p}_1^{(c)},\boldsymbol{p}_2^{(c)},\cdots,\boldsymbol{p}_L^{(c)}\right]$ 为 $\boldsymbol{X}$ 协方差的特征矢量阵。由此，式(2)可进一步写为

$$D_{ij}=\frac{1}{2}\sum_{l=1}^{L'}\frac{\left(\sigma_i^2(l)-\sigma_j^2(l)\right)^2}{\sigma_j^2(l)\sigma_i^2(l)}+\frac{1}{2}\sum_{l=1}^{L'}\frac{\sigma_i^2(l)+\sigma_j^2(l)}{\sigma_j^2(l)\sigma_i^2(l)}\left(u_i(l)-u_j(l)\right)^2 \tag{5}$$

其中 $\sigma_i^2(l)$，$u_i(l)$，$\sigma_j^2(l)$，$u_j(l)$ 分别表示变换后第 $l$ 波段类 $i$ 和类 $j$ 的方差和均值；$L'$ 为子集波段数，$1\le L'\le L$。得到 $m$ 类数据平均散度:

$$\begin{aligned}D_{\mathrm{AVE}}&=\sum_{l=1}^{L'}D_{\mathrm{ave}}(l)\\&=\frac{1}{m(m-1)}\sum_{l=1}^{L'}\sum_{i=1}^{m}\sum_{j=i+1}^{m}\left(\frac{\left(\sigma_i^2(l)-\sigma_j^2(l)\right)^2}{\sigma_j^2(l)\sigma_i^2(l)}+\frac{\sigma_i^2(l)+\sigma_j^2(l)}{\sigma_j^2(l)\sigma_i^2(l)}\left(u_i(l)-u_j(l)\right)^2\right)\end{aligned} \tag{6}$$

式(6)表明: (1) 波段子集的类别可分性完全由单个新波段的性能叠加决定，因此只需进行 $L$ 次单波段的散度计算即可获得任意波段子集的散度；(2) 散度越大的波段对子集的类别可分性贡献越大。

### 2.2　变换系数分析与投票

按散度大小降序排列经类内 PCA 变换得到的新波段，显然波段越靠前对类别可分性影响越大。考虑类 $c$ 的样本 $\boldsymbol{x}^{(c)}$，经 PCA 变换:

$$\begin{bmatrix}y_1^{(c)}\\y_2^{(c)}\\\vdots\\y_L^{(c)}\end{bmatrix}=\boldsymbol{P}^{(c)^{\mathrm{T}}}\begin{bmatrix}x_1^{(c)}\\x_2^{(c)}\\\vdots\\x_L^{(c)}\end{bmatrix}=\begin{bmatrix}p_{11}^{(c)}&p_{21}^{(c)}&\cdots&p_{L1}^{(c)}\\p_{12}^{(c)}&p_{22}^{(c)}&\cdots&p_{L2}^{(c)}\\\vdots&\vdots&\vdots&\vdots\\p_{1L}^{(c)}&p_{2L}^{(c)}&\cdots&p_{LL}^{(c)}\end{bmatrix}\begin{bmatrix}x_1^{(c)}\\x_2^{(c)}\\\vdots\\x_L^{(c)}\end{bmatrix},\ c=1,\cdots,m \tag{7}$$

上式表明，每个主成分(即新波段)都是 $\boldsymbol{x}^{(c)}$ 分量(即原始波段)的线性组合。$\boldsymbol{P}^{(c)}$ 的第 $i$ 行反映了分量 $x_i^{(c)}$ 对各主成分的作用，该行的元素越趋于零，该分量的作用越小；另外考虑到

各新波段对分类影响的不同，借鉴文献[6]中载荷因子(loading factor)定义，用

$$\overline{p}_i^{(c)} = \frac{1}{L}\sum_{j=1}^{L} a_j \left(p_{ij}^{(c)}\right)^2, \quad i = 1,\cdots,L \tag{8}$$

表示原始波段 $i$ 对类 $c$ 的重要度，由此获得按照对类 $c$ 重要性降序排列的原始波段序列。其中权值:

$$a_j = D_{\text{ave}}(j) \Big/ \sum_{l=1}^{L} D_{\text{ave}}(l) \tag{9}$$

体现了各新波段对分类的影响。由式(8)，式(9)可得到相应于类别的 $m$ 个原始波段序列。

我们需要从这 $m$ 个原始波段序列中选出对所有类别都较为重要的波段，一个简单有效的方法就是投票。考虑到波段序列中位于前面的波段更重要，设计投票方法：定义 $k_c^{(i)}$ 为波段 $i$ ($1 \le i \le L$)在类 $c$ 序列($1 \le c \le m$)的位置，则该波段的重要度可表示为

$$I_{\text{tot}}(i) = \sum_{c=1}^{m}\left(L - k_c^{(i)} + 1\right) \Big/ L \tag{10}$$

并有各波段平均重要度为 $m(L+1)/(2L)$。选择不小于平均重要度的波段作为候选集。

### 2.3 波段子集的最终确定

候选集波段按相关度形成若干界限明确的子集，这一点我们可在后面的实验中看到。这些子集间相关度较低，信息差异大；而子集内部具有较强的相关性，因此所含信息量变化不大，少量的波段就可以代表子集。综合考虑子集的规模，波段的散度和相关度，设 $L^{(1)}$ 波段的候选集 $B$ 由 $q$ 个子集组成，有 $B = U_{i=1}^{q} B^{(i)}$， $B^{(i)} = \left\{b_j^{(i)}\right\}_{j=1}^{L_i}$， $\sum_{i=1}^{q} L_i = L^{(1)}$；最终波段集合 $S$ 从 $q$ 个子集中得到， $S = U_{i=1}^{q} S^{(i)}$。定义 $\max\text{corr}(b,S)$ 为波段 $b$ 与属于集合 $S$ 的波段的最大相关度，算法步骤如下：

(1) 各子集中波段按散度降序排列。

(2) $j=1$， $\text{flag}^{(i)}=1$ ( flag 为结束标志)。选择各子集的第一个波段， $S_1^{(i)} = \left\{b_1^{(i)}\right\}$。 $i = 1,\cdots,q$。

(3) $j = j+1$，若 $\text{flag}^{(i)} = 0$，或子集规模太小(如小于 10 个波段)，则不再考察该子集。若 $\max\text{corr}\left(b_k^{(i)}, S_{j-1}^{(i)}\right) \le T_{\text{corr}}^{(i)}$ ( $T_{\text{corr}}^{(i)}$ 为相关度阈值， $j \le k \le L_i$)，则令 $S_j^{(i)} = S_{j-1}^{(i)} U \left\{b_k^{(i)}\right\}$；否则 $\text{flag}^{(i)} = 0$。 $i = 1,\cdots,q$。

(4) 若 $\text{flag}^{(i)} = 0$ ( $i = 1,\cdots,q$)， $S = U_{i=1}^{q} S_{j-1}^{(i)}$，结束。

(5) 转至(3)步。

### 2.4 复杂度分析

与一般搜索方法相比，本方法首先对样本作了 PCA 处理以降低散度计算及搜索的负担。表 1 列出了本文方法与穷举法、前向搜索法主要计算量的比较情况，设选择 $L'$ 波段。可以看到，穷举法的计算量是难以负担的。若样本数 $N$ 较大( $N > mL'$)，前向搜索中，计算量为 $O\left(NLm^2L'^3\right)$，受样本数、波段总数、类别数及选择波段数影响；本文方法计算量 $O\left(NmL^2\right)$，与样本数、类别数及波段总数相关。大致当 $L' \ge \left\lceil \sqrt[3]{L/m} \right\rceil$ ($\lceil\cdot\rceil$ 表示上取整)时，计算负担低于前向搜索，而在实际应用中，$L'$ 一般都满足上述关系。以 AVIRIS 的 220 波段数据集为例，考虑两类问题( $m=2$)，显然当 $L' \ge 5$ 时计算量便低于前向搜索，且保持不变。因此，当 $L'$ 不是过小时，本文方法在计算量上有显著优势。

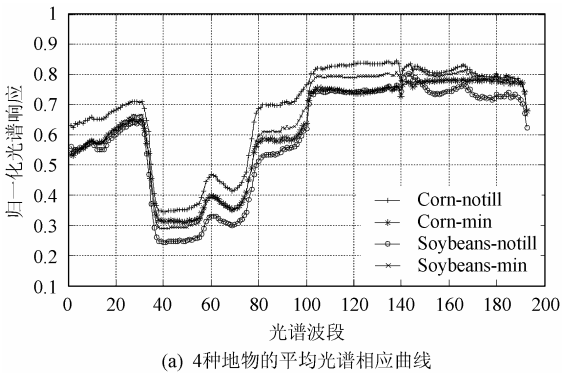
## 3 分类实验

实验数据来自美国 AVIRIS 多光谱扫描仪 1992 年在印第安纳获取的220波段高光谱数据，从中选出质量较好的190个波段进行实验。我们分别选取 4 种地物和 3 种地物进行分类，地物类别和样本数如表 2，它们的平均光谱相应曲线如图 1。可以看到第 1 个实验中 4 种地物光谱响应比较类似，分类难度较大。

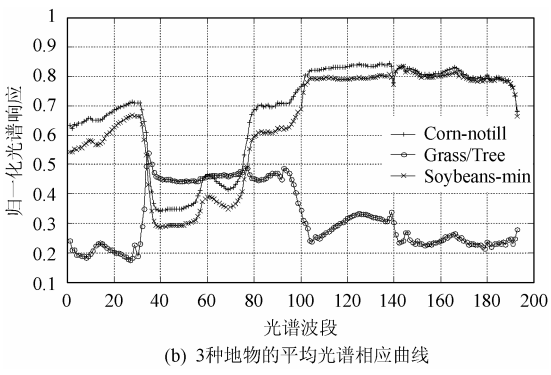
实验 1，实验 2 分别采用 3 种方法进行波段选择：等间隔波段选择①，前向搜索波段选择②及本文的方法，在得到的波段子集上用最大似然分类器进行分类，3 种方法的比较见表 3。其中类对的散度按式(3)计算。实验 1 中，本文方法得到的候选波段集相关度矩阵可图形化表示为图 2(a)，图中标出了对角线块的平均相关度，可见它们明显分为 4 个子集合；实验 2 的候选波段集相关度矩阵如图 2(b)所示，可划分为 3 个子集。两个实验中，相关度阈值分别取为各子集平均相关度和各子集相关度极值的平均，本文方法最终选出 13 和 6 个波段，为方便比较，方法①、方法②也选择相同的波段数。

**表 1 3 种方法计算量的分配情况**

| | 分类别 PCA | | | 散度 | |
|---|---|---|---|---|---|
| | 协方差、均值 | Householder-QR 算法 | 矩阵乘 | 计算量 | 比较 |
| 穷举法 | — | — | — | $O\left(\max\left(L'^3m^3, NL'^2m^2\right)\right)\cdot O\left(L'^{L'}\right)$ | $O\left(L'^{L'}\right)$ |
| 前向搜索 | — | — | — | $O\left(\max\left(Lm^3L'^4, NLm^2L'^3\right)\right)$ | $O(LL')$ |
| 本文方法 | $O\left(NL^2\right)$ | $O\left(mL^3\right)$ | $O\left(NmL^2\right)$ | $O(L)$ | $O(L\log L)$ |



(a) 4种地物的平均光谱相应曲线



(b) 3种地物的平均光谱相应曲线

图 1

表 2(a)　4 种地物类别及样本分配

| 地物类别 | 训练样本 | 测试样本 |
|---|---|---|
| Corn-notill | 171 | 342 |
| Corn-min | 120 | 240 |
| Soybeans-notill | 136 | 272 |
| Soybeans-min | 285 | 570 |
| 总计 | 712 | 1424 |

表 2(b)　3 种地物类别及样本分配

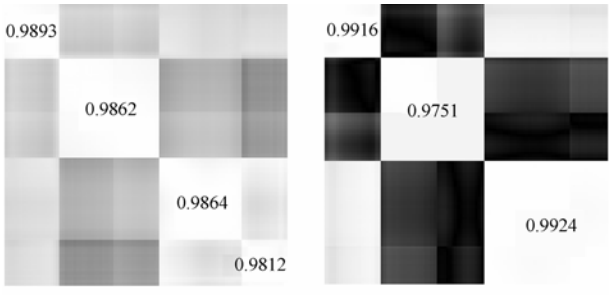
| 地物类别 | 训练样本 | 测试样本 |
|---|---|---|
| Corn-notill | 171 | 342 |
| Grass/Tree | 113 | 226 |
| Soybeans-min | 285 | 570 |
| 总计 | 569 | 1138 |

表 3(a)　实验 1 中 3 种波段选择方法得到的波段子集及分类精度

| | 方法① | 方法② | 本文方法 |
|---|---|---|---|
| 波段子集 | 14，28，42，56，70，84，98，112，126，140，154，168，182 | 57，142，122，155，5，29，68，14，40，35，95，42，117 | 29，52，122，155，28，73，135，166，15，36，112，146，107 |
| 分类精度(%) | 76.40 | 80.09 | 80.97 |

表 3(b)　实验 2 中 3 种波段选择方法得到的波段子集及分类精度

| | 方法① | 方法② | 本文方法 |
|---|---|---|---|
| 波段子集 | 15，46，77，108，139，170 | 15，29，164，12，156，61 | 15，39，151，29，72，115 |
| 分类精度(%) | 85.24 | 91.39 | 90.69 |



(a) 实验1候选波段集相关度矩阵　(b) 实验2候选波段集相关度矩阵

图 2

从表中可以看到，由于实验 1 的地物光谱特征较为接近，分类难度大，尽管选择了较多的波段，其分类精度仍低于实验 2。等间隔波段选择是最简单、最快的选择方法，但没有对数据进行分析，分类精度最低；前向搜索方法是一种常用的局部最优搜索，与它相比本文方法在精度上差别不大，但在计算量上下降到它的 $L/\left(L'^{3}m\right)$，并且随着选择波段和类别的增多，优势更趋显著。

## 4　结束语

本文提出了一种新的波段选择方法，基于分类别 PCA 变换散度及相应系数分析的波段选择。该方法由分类别 PCA 变换后散度计算、候选集获取和最终子集确定 3 步构成。通过对样本各类数据分别作 PCA 变换去相关，将散度函数中协方差阵对角化，使得变换后的新波段组合散度值仅由单个新波段的散度加和决定，从而避免了计算波段各种可能组合散度值的沉重负担。复杂度分析表明，本文方法计算量由样本数、波段总数和类别数决定，在一般情况下计算量远低于前向搜索。在高光谱遥感图象的分类实验中，本文方法也取得了好的结果，是一种快速高效的波段选择方法。

## 参 考 文 献

[1] Velez-Reyes M, Linares D M. Comparison of principal-component-based band selection methods for hyperspectral imagery. Image and Signal Processing for Remote Sensing VII, Proc. SPIE, 2002, 4541: 361 − 369.

[2] Withagen Paul J, Breejen Eric den, *et al*.. Band selection from a hyperspectral data-cube for a real-time multispectral 3CCD camera. Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VII, Proc. SPIE AeroSense, 2001, 4381: 84 − 93.Sheffer D, Ultchin Y. Comparison of band selection results using.

[3] different class separation measures in various day and night conditions. Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery IX, Proc. SPIE, 2003, 5093: 452 – 461.

[4] Swain P H, King R C. Two effective feature selection criteria for multispectral remote sensing. First International Joint Conference on Pattern Recognition, Washington, DC, 1973: 536 – 540.

[5] J. P. Marques de sá著，吴逸飞译. 模式识别——原理、方法及应用. 北京: 清华大学出版社, 2002 : 116 – 118.

[6] Chang Chein-I, Du Qian, *et al*.. A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification. *IEEE Trans. on Geoscience and Remote Sensing*, 1999, 37(6): 2631 – 2641.

黄 睿： 女，1976 年生，博士生，研究方向为高光谱数据分析、模式识别和人工智能.

何明一： 男，1958 年生，教授，博士生导师，陕西省信息获取与处理重点实验室主任，主要研究方向为信息获取与智能处理、图象与图形处理和三维测量技术等.