

## 基于聚类的高光谱图像无损压缩

粘永健<sup>①</sup> 苏令华<sup>②</sup> 孙蕾<sup>③</sup> 万建伟<sup>①</sup>

<sup>①</sup>(国防科技大学电子科学与工程学院 长沙 410073)

<sup>②</sup>(空军大连通信士官学校 大连 116600)

<sup>③</sup>(国防科技大学理学院 长沙 410073)

**摘要:** 高光谱海量数据的有效压缩成为遥感技术发展中需要迫切解决的问题。该文提出了一种基于聚类的高光谱图像无损压缩算法。针对高光谱图像不同频谱波段间相关性不同的特点, 根据相邻波段相关性大小进行波段分组。由于高光谱图像波段数量较多, 采用自适应波段选择算法对高光谱图像进行降维, 以获取信息量较大的部分波段, 利用  $k$  均值算法对降维后的波段谱矢量进行聚类。采用多波段预测的方案对各组中的波段进行预测, 对于各个分类中的每个像素, 分别选取与其空间相邻的已编码的部分同类点进行训练, 从而获得当前像素的谱间最优预测系数。对 AVIRIS 型高光谱图像的实验结果表明, 该算法可显著降低压缩后的平均比特率。

**关键词:** 高光谱图像; 无损压缩; 波段分组; 谱向聚类

中图分类号: TP751.1

文献标识码: A

文章编号: 1009-5896(2009)06-1271-04

## Lossless Coding for Hyperspectral Images Based on Spectral Cluster

Nian Yong-jian<sup>①</sup> Su Ling-hua<sup>②</sup> Sun Lei<sup>③</sup> Wan Jian-wei<sup>①</sup>

<sup>①</sup>(College of Electronic Science and Engineering, National Univ. of Defense Technology, Changsha 410073, China)

<sup>②</sup>(Dalian Communication Sergeant School of Air Force, Dalian 116600, China)

<sup>③</sup>(College of Science, National Univ. of Defense Technology, Changsha 410073, China)

**Abstract:** The request for efficient compression of hyperspectral images becomes pressing. A cluster-based lossless compression algorithm for hyperspectral images is presented. Because the spectral correlation differs in different bands, spectral band grouping algorithm is introduced to divide hyperspectral images into groups according to the correlation between each adjacent bands. The important bands which contain large useful information can be determined by using the adaptive band selection algorithm, on which  $k$ -means clustering is carried out according to the spectral vectors. The current band is predicted by using several preceding bands. For each pixel which belongs to a certain cluster, some causal neighboring pixels which have been coded are trained to get the optimal predictive coefficients. The reference bands are compressed by JPEG-LS standard while the final predictive errors are coded by Golomb-Rice. Experimental results show that the proposed methods produce competitive results when compared with other state-of-the-art algorithms.

**Key words:** Hyperspectral image; Lossless compression; Band grouping; Spectral cluster

### 1 引言

高光谱遥感是 20 世纪 80 年代兴起的新型对地观测技术, 已经在军事侦察和国民经济各领域得到了广泛应用。成像光谱仪作为一种新型的航天遥感器, 同时具有成像仪和光谱仪的功能, 使用成像光谱仪获取的遥感图像称为高光谱图像。随着谱间和空间分辨率的不断增长, 高光谱数据的波段宽度达到纳米量级 ( $10^{-2}\lambda$ ,  $\lambda$  为波长), 从可见光到近红外光谱区间波段数可达几十甚至上百个, 数据量非常庞大, 给

数据的存储和传输带了困难, 必须采用高效的压缩方法对其进行压缩。高光谱图像的获取非常昂贵, 具有长期的保存价值, 同时由于高光谱数据经常应用于特征提取、目标检测与分类等领域<sup>[1]</sup>, 有损压缩会对后续应用造成无法估量的影响, 因而无损压缩成为首选方案。

高光谱图像无损压缩方法可以分为基于预测的方法、基于整数变换的方法、基于矢量量化的方法或者多种方法结合使用。对于高光谱图像无损压缩而言, 基于预测的方法成为首选<sup>[2]</sup>。F. Rizzo 等人设计了相对简单的 SLSQ(Spectral-oriented Least Squares)算法, 若增加预测器选择的复杂度(SLSQ-OPT), 可进一步提高压缩比<sup>[3]</sup>。S. K. Jain 利用前一波段对当前波段进行非线性预测, 并首次将边缘分析应用于

2008-05-30 收到, 2008-10-14 改回

国家自然科学基金(60572135)和国防科技大学优秀研究生创新基金资助课题

高光谱图像的无损压缩<sup>[4]</sup>。Zhang Jing 等人在谱内预测基础上,将预测误差反馈机制引入到谱间预测<sup>[5]</sup>。P. L. Dragotti 等人在文献[6]中首次提出了基于3D SPIHT算法的高光谱图像无损压缩方案。柴焱等人在文献[7]中详细分析了各种变换方法在谱内和谱间去相关上的性能,并提出了基于混合整型变换和3D SPIHT的无损压缩算法。由于高光谱图像内容比较丰富,以非监督分类作为预处理手段的压缩方法表现出较大的优势。本文提出一种基于谱向矢量聚类的高光谱图像无损压缩算法。首先利用自适应波段选择算法对高光谱图像进行降维处理,对降维后的波段谱向矢量进行k-means聚类;在波段分组基础上,采用多波段谱间预测方式,仿真结果证明了算法的有效性。

## 2 波段分组与排序

波段分组是使得相关性较强的波段分到同一组(Group Of Bands, GOB),而相关性较差的波段分到不同组,进而提高谱间预测性能。文献[5]根据相邻波段相关性大小对进行非等长分组,不但提高了预测性能,而且降低了波段排序算法的运算量。AVIRIS型高光谱图像在第113波段、第160波段和第221波段附近相关性较差,这些波段之间并不适合进行谱间预测,因而需要进行波段分组。根据上述分析,给出如下波段分组方法

(1)计算相邻波段的相关系数 $R$ ,若 $R \geq T$  ( $0 < T < 1$ ),将该两个波段归为同一GOB,否则置于不同的GOB;

(2)设置参数 $B$ ,若GOB中的波段数大于 $B$ ,将GOB中每 $B$ 个波段作为一个子集(Subset of Bands, SOB),若有剩余波段,则单独作为一个SOB。

为了获得较好的压缩效果,往往需要对各GOB(SOB)的波段重新排序。文献[8]的研究结果表明,根据段波长递增或递减的顺序进行预测,能够取得较好压缩效果。本文对各GOB(SOB)按照波长递增的原则进行简单排序。由于波段之间的相关性取决于成像光谱仪,因而上述波段分组和排序的结果,可应用于大部分该光谱仪获取的图像。

## 3 高光谱图像降维与谱向聚类

非监督分类方法在遥感图像的应用中具有非常重要的作用,它是指对分类过程不施加任何先验知识,仅凭遥感图像地物光谱特征的分布规律,随其自然地进行“盲目”分类。采用非监督分类方法预先对图像进行“粗”分类,对各个分类中的像素,分别训练其预测系数,可显著改善预测编码的性能。高光谱图像通常包括上百个谱向波段,而聚类是一种计算量较大的算法,对每个波段分别聚类不仅计算量较大,而且聚类后会产生大量的附加信息。针对这一不足,首先对高光谱图像进行降维处理,从中获取信息量较大的波段。本文采用文献[9]提出的自适应波段选择算法(Adaptive Band Selection, ABS)进行降维,其数学模型为

$$S_k = 2\sigma_k / (R_{k-1,k} + R_{k,k+1}) \quad (1)$$

其中 $S_k$ 为波段 $k$ 的指数, $\sigma_k$ 为波段 $k$ 的标准差, $R_{u,v}$ 为波段 $u$ 和 $v$ 之间的互相关系数,本文利用式(1)分别计算各个波段的指数,并按照由大到小排序,选出前 $D$ 个波段进行聚类。若高光谱图像第 $k$ 个波段表示为 $[f_k(i,j)]_{M \times N}$ ,其中 $M \times N$ 为每个波段大小。将上述 $D$ 个波段表示为谱向矢量形式

$$\mathbf{H} = [f_1'(i,j), f_2'(i,j), \dots, f_D'(i,j)]_{M \times N} \quad (2)$$

采用k-means算法对 $\mathbf{H}$ 进行聚类,分类标识表示为

$$L(i,j) = \ell, \quad \ell \in \{1, 2, \dots, C\} \quad (3)$$

其中 $C$ 为分类数,同类像素点的集合表示为

$$\{f_k(i,j) | L(i,j) = \ell, i = 1, \dots, M, j = 1, \dots, N\} \quad (4)$$

## 4 多波段谱间预测方案

考虑到高光谱图像波段间较强的相关性,在完成分类的基础上,对当前波段采用多个参考波段进行谱间预测。各个分类中的像素按照顺序扫描的方式,通过选取与当前像素相邻的已编码部分同类像素点进行训练,可以获得当前像素的最优谱间预测系数。对于当前波段的像素 $f_k(i,j)$ ,其谱间预测值为

$$f_k(i,j) = \left\lfloor \sum_{l=1}^L \alpha_l f_{k-l}(i,j) + 0.5 \right\rfloor \quad (5)$$

其中 $L$ 为预测器阶数。选取与当前像素 $f_k(i,j)$ 相邻的 $K$ 点局部邻域,表示为 $\mathbf{X} = [x_{k,1}, \dots, x_{k,K}]^T$ 。最佳预测系数 $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_L]^T$ 的选择满足式(6)最小化

$$G = (\mathbf{P}\alpha - \mathbf{A})^T (\mathbf{P}\alpha - \mathbf{A}) \quad (6)$$

其中

$$\mathbf{P} = \begin{bmatrix} x_{k-1,1} & \dots & x_{k-L,1} \\ \vdots & \ddots & \vdots \\ x_{k-1,K} & \dots & x_{k-L,K} \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} x_{k,1} \\ \vdots \\ x_{k,K} \end{bmatrix} \quad (7)$$

最佳线性预测器的预测系数为

$$\alpha = (\mathbf{P}^T \mathbf{P})^{-1} (\mathbf{P}^T \mathbf{A}) \quad (8)$$

由式(8)可知,求解 $\alpha$ 的运算量随着 $K$ 和 $L$ 的增大而迅速增加。文献[10]的研究结果表明,对于多波段谱间预测而言,预测器阶数为3时可获得较好的预测效果。综合考虑压缩性能和运算量两方面因素,选取 $K=4$ , $L=3$ 。图1所示为参考波段与预测波段相同位置上的邻域结构。

对于各个分类中的像素,图1所示邻域结构中的像素并非都是当前像素的同类点,很多情况下同类点之间位置比较分散。为解决这一问题,本文给出了一种同类邻点搜索方案,为提高搜索速度,首先定义像素之间的距离公式为

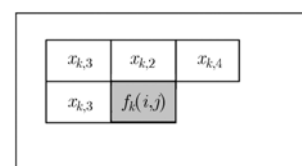


图1 当前像素的局部邻域结构

$$d(f_g(m, n), f_g(p, q)) = |m - p| + |n - q| \quad (9)$$

对于各个分类, 具体的搜索算法为

(1)按从左到右、从上到下的顺序扫描分类图。对于当前像素点  $f_k(i, j)$  且  $L(i, j) = \ell$ , 按照左、上、左上和右上的顺序进行判断, 若均为  $f_k(i, j)$  的同类点, 直接进行谱间预测, 否则转入步骤(2);

(2)定义搜索区域。若  $L(i, j-1) \neq \ell$ , 搜索区域定义为  $\{(i', j') \mid i' = i, j' < j\}$ , 在该区域中搜索与  $f_k(i, j)$  距离最小的同类点; 若  $L(i-1, j) \neq \ell$ , 在  $\{(i', j') \mid i' < i, j' = j\}$  中搜索最小距离同类点; 若  $L(i-1, j-1) \neq \ell$ , 搜索区域定义为  $\{(i', j') \mid (i' \leq i, j' < j) \cup (i' < i, j' \leq j)\}$ , 若  $L(i-1, j+1) \neq \ell$ , 相应的搜索区域为  $\{(i', j') \mid (i' \leq i, j' > j) \cup (i' < i, j' \geq j)\}$ ;

(3)若上述某个搜索区域中未找到同类点, 需在其它搜索区域中补充搜索相同数量的同类点, 对于同一搜索区域中最小距离相等的同类点, 按扫描先后进行选取。

利用上述多波段预测算法对各 GOB 进行预测编码, 对于 GOB 中包含多个 SOB 的情况, 除第一个 SOB 外, 其它 SOB 中的第 1 个波段均由其上 1 个相邻的 SOB 中的最后一个波段进行预测。由于 JPEG-LS 无损压缩标准具有较高的压缩性能和运算效率, 采用该标准对参考波段进行无损压缩, 预测残差直接进行 Gombolb-Rice 编码。综上所述, 本文提出的无损压缩算法框图可由图 2 表示。

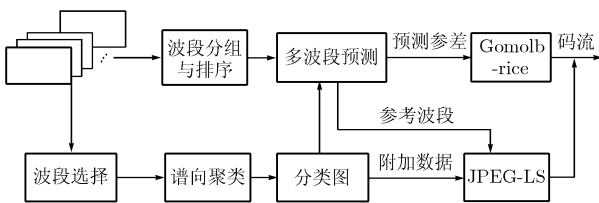


图 2 基于聚类的高光谱图像无损压缩算法框图

### 5 实验结果与讨论

为验证所提算法的有效性, 利用 matlab 软件对 224 波段 AVIRIS 型高光谱图像进行了仿真实验。AVIRIS 数据来自于美国国家航空和航天管理局(NASA), 包括 Cuprite, Jasper ridge, Lunar lake 和 Low altitude, 每个像素保存为

16bit, 每个波段图像大小为  $614 \times 512$ 。本实验将所有图像大小均截取为  $256 \times 256$ , 图 3 给出了 Cuprite, Jasper ridge, Lunar lake 和 Low altitude 各 1 个波段图像。

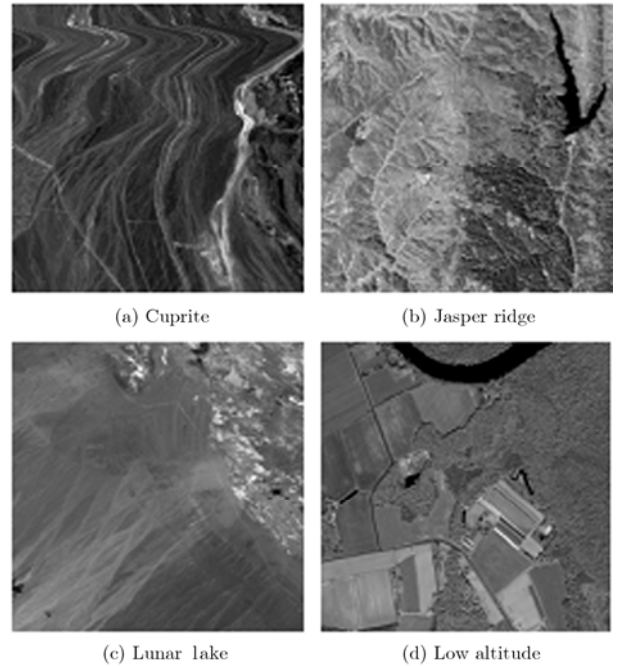


图 3 高光谱图像的原始波段

仿真过程中相应的参数选取如下: 对于波段分组, 选取  $T = 0.9$ ,  $B = 32$ ; 对于谱向聚类, 选取  $D = 28$ , 分类数  $C = 8$ 。表 1 给出了本文算法与几种具有代表性无损压缩算法的结果比较, bpp 表示压缩后每个像素的比特数。JPEG2000 和 JPEG-LS 标准均只进行谱内压缩, 压缩性能较低, 但由于 JPEG-LS 标准是专为图像无损压缩而设计, 其压缩性能要优于 JPEG2000。Differential JPEG-LS 算法是先将相邻波段图像进行差分运算, 再分别进行谱内 JPEG-LS 压缩, 压缩效果较前两者有所提高。SPIHT-Multi 算法是对高光谱图像进行等长分组, 利用一维小波变换对各组去除谱间相关性, 之后采用 SPIHT 算法对各个波段进行无损压缩, 这是一种准三维的压缩方法; 3D SPIHT 算法是基于三维小波变换的等级树编码; 文献[10]研究了预测阶数对无损压缩性能的影响, 并给出了多波段最优谱间预测方

表 1 各种算法压缩结果比较(单位: bpp)

	JPEG2000	JPEG-LS	Differential JPEG-LS	SPIHT-Multi	3D SPIHT	多波段最优预测	文献[11]	本文算法
Cuprite	8.188	7.497	6.119	5.647	5.357	5.325	5.096	4.818
Jasper ridge	8.533	7.709	5.650	5.734	5.439	5.143	4.984	4.732
Lunar lake	7.996	7.370	6.171	5.736	5.390	5.503	5.161	4.901
Low altitude	8.796	8.001	6.314	6.272	5.913	5.708	5.316	5.126
平均 bpp	8.378	7.644	6.064	5.847	5.525	5.419	5.139	4.894

案;文献[11]提出了一种基于同类邻点多波段预测的无损压缩方案。本文算法在聚类结果的基础上进行多波段谱间去相关,极大地降低了压缩后的平均 bpp,压缩性能明显优于上述几种算法。

一般而言,基于预测方法的压缩效果要优于基于变换的方法,主要原因是后者利用特定的变换方法去除谱间相关性,将波段之间孤立地看待;而基于预测的方法是从波段之间整体或局部的统计最优上获得最佳去相关性能,去相关性要好于基于变换的方法。图4为 Jasper ridge 压缩后的 bpp 曲线比较,可以看出,多波段最优预测算法的个别波段压缩效果较差,原因是这些波段与其相邻波段之间相关性较弱,本文进行了波段分组,较好地解决这一问题。

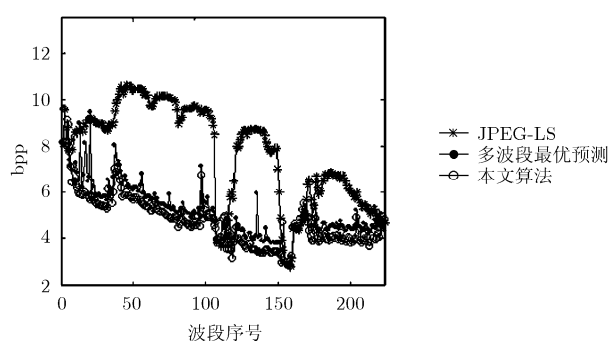


图4 AVIRIS 型高光谱图像压缩后的 bpp 曲线图

## 6 结束语

本文提出了一种基于聚类的高光谱图像无损压缩算法,该算法对高光谱图像谱向矢量进行聚类,在波段分组基础上,采用多波段预测方案,并提出了同类邻点搜索方案,取得了较好的压缩效果。本文在波段分组的基础上,只是简单地按照波长顺序进行了排序,并没有采用特殊的排序方法,一方面是由于排序算法的运算量都比较大,另一方面由于采用了多波段谱间预测,使得波段排序算法在提高压缩性能方面的作用非常有限。此外,合理分类数的选择也是需要进一步研究的问题。分类数选择太少,压缩性能较差,而分类数选择较多,分类结果越精细,对同类点的搜索和预测系数的训练都带来困难,反而不利于压缩。此外,采用性能更好的聚类方法,可以进一步提高压缩效果。

## 参考文献

- [1] 苏令华,李纲,衣同胜,等.一种稳健的高光谱图像压缩方法[J].光学精密工程,2007,15(10):1609-1615.  
Su L H, Li G, and Yi T S, et al. A robust hyperspectral image compression method [J]. *Opt. Precision Eng.*, 2007, 15(10): 1609-1615.
- [2] Mielikainen J and Toivanen P. Clustered DPCM for the lossless compression of hyperspectral images [J]. *IEEE Trans.*

- on Geoscience and Remote Sensing*, 2003, 41(12): 2943-2946.
- [3] Rizzo F, Carpentieri B, and Motta G, et al. Low-complexity lossless compression of hyperspectral imagery via linear prediction [J]. *IEEE Signal Processing Letters*, 2005, 12(2): 138-141.
- [4] Jain S K and Adjeroh D A. Edge-based prediction for lossless compression of hyperspectral images [C]. Data compression conference, Snowbird, USA, 2007: 153-162.
- [5] Zhang J and Liu G Z. An efficient reordering prediction-based lossless compression algorithm for hyperspectral images [J]. *IEEE Geoscience and Remote Sensing Letters*, 2007, 4(2): 283-287.
- [6] Dragotti P L, Poggi G, and Ragozini A R P. Compression of multispectral images by three-dimensional SPIHT algorithm [J]. *IEEE Trans. on Geoscience and Remote Sensing*, 2000, 38(1): 416-428.
- [7] 柴焱,计文平,沈兰荪.一种基于混合整型变换和3D-SPIHT的高光谱图像嵌入式无损压缩方法[J].电子学报,2007,35(9):1770-1773.  
Chai Y, Ji W P, and Shen L S. Hyperspectral Image lossless compression algorithm based on hybrid transform and 3D-SPIHT [J]. *Acta Electronica Sinica*, 2007, 35(9): 1770-1773.
- [8] Rao A K and Bhargava S. Multispectral data compression using bidirectional interband prediction [J]. *IEEE Trans. on Geoscience and Remote Sensing*, 1996, 34(2): 385-397.
- [9] 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].中国图像图形学报,2005,10(2):218-222.  
Liu C H, Zhao C H, and Zhang L Y. A new method of hyperspectral remote sensing image dimensional reduction [J]. *Journal of Image and Graphics*, 2005, 10(2): 218-222.
- [10] 孙蕾,罗建书.基于多波段谱间预测的高光谱图像无损压缩算法[J].电子与信息学报,2007,29(12):2876-2879.  
Sun L and Luo J S. Hyperspectral images lossless compression based on multi-bands prediction [J]. *Journal of Electronics and Information Technology*, 2007, 29(12): 2876-2879.
- [11] 苏令华,吕韶昱,万健伟.基于多预测器的高光谱图像无损压缩[J].国防科技大学学报,2007,29(1):44-48.  
Su L H, Lv S Y, and Wan J W. Lossless compression of hyperspectral image based on multi-predictor [J]. *Journal of National University of defense Technology*, 2007, 29(1): 44-48.

粘永健:男,1982年生,博士生,研究遥感图像处理等。  
苏令华:男,1979年生,博士,讲师,研究遥感图像处理等。  
孙蕾:女,1980年生,博士生,研究遥感图像处理等。  
万健伟:男,1964年生,博士,教授,博士生导师,主要研究图像处理与雷达信号处理等。