

文章编号 1004-924X(2012)03-0668-07

基于内容的高光谱图像无损压缩

汤 毅*, 辛 勤, 李 纲, 万建伟

(国防科技大学 电子科学与工程学院, 湖南 长沙 410073)

摘要:提出了一种基于内容的高光谱图像无损压缩算法。采用自适应波段选择算法对高光谱图像进行降维,引入 C-means 算法对降维后的光谱矢量进行无监督分类。利用单调后向排序算法确定波段的预测顺序,并根据相邻波段的相关系数大小进行自适应波段分组。针对每一类地物,选取类内部分像素进行最优预测系数的训练,采用多波段线性预测的方案去除同类像素的谱间相关性,预测残差进行 JPEG-LS 无损压缩。对机载可见光/红外成像光谱仪(AVIRIS)与实用型模块化成像光谱仪(OMIS)获取的高光谱图像分别进行实验,并与未进行分类预测的算法比较。结果显示,提出的算法的平均压缩比分别提高约 0.11 和 0.7,验证了该算法在无损压缩方面的有效性。

关键词:高光谱图像;图像压缩;无损压缩;波段选择;光谱分类

中图分类号:TP391.4;TH751.1 文献标识码:A doi:10.3788/OPE.20122003.0668

Lossless compression of hyperspectral images based on contents

TANG Yi*, XIN Qin, LI Gang, WAN Jian-wei

(College of Electronic Science and Engineering, National University of Defense Technology,
Changsha 410073, China)

* Corresponding author, E-mail: lantange@163.com

Abstract: A lossless compression algorithm based on contents was proposed for hyperspectral images. An adaptive band selection algorithm was introduced to reduce the dimensionality of hyperspectral images, and a C-means algorithm was used to classify the spectral vectors resulting from dimensionality reduction unsupervisedly. Then, the reverse monotonic ordering method was taken to determine the prediction ordering, hyperspectral images were divided into groups adaptively according to the correlation between each adjacent bands, and the scheme of multi-band linear prediction was used to eliminate the spectral redundancy of the identical class. For each class, partial pixels within this class were selected to train optimal predictive coefficients, and predictive errors were compressed in lossless by JPEG-LS standard. Experiments were performed for the hyperspectral images acquired by an Airborne Visible/Infrared Imaging Spectrometer(AVIRIS) and an Operational Modular Imaging Spectrometer(OMIS). Experiential results show that the average compression ratio of the proposed algorithm can be improved about 0.11 and 0.7 respectively as compared with above algorithms without classification prediction.

收稿日期:2011-09-14;修订日期:2011-10-31.

基金项目:国家自然科学基金资助项目(No. 60572135);国家 863 高技术研究发展计划资助项目(军口);武器装备预研基金资助项目

Key words: hyperspectral imagery; image compression; lossless compression; band selection; spectral classification

1 引言

遥感技术在20世纪80年代进入了高光谱遥感时代。由于高光谱遥感影像能够提供更为丰富的地表信息,在军事侦查和国民经济等领域得到了广泛应用。纳米量级的光谱分辨率使得高光谱图像从可见光到近红外的光谱区间波段数达上百个。随着光谱分辨率与空间分辨率的不断提高,高光谱图像数据量持续增加^[1]。目前,高光谱遥感正在向航天航空高光谱遥感方向发展,在高分辨率对地观测中发挥着重要作用。对于卫星数据链路,日益增长的数据量给存储和传输带来了负担,严重制约着航天航空高光谱遥感技术的发展,因此,研究高效的数据压缩方法势在必行。

高光谱图像的无损压缩方法可分为基于预测的方法、基于变换的方法与基于矢量量化的方法等。基于预测的方法^[2]能够在压缩性能与运算复杂度之间达到一个较好的折中,在无损压缩中获得了广泛应用。武筱林提出了基于上下文的自适应无损编码(Context-based, Adaptive, Lossless Image Code, CALIC)^[3],其压缩效果优于JPEG-LS无损压缩标准,在自然图像的压缩中取得了较好效果,但由于未考虑谱间相关性,用于高光谱图像无损压缩的效果较差。在文献[4]中,Wu和Memon将其推广到3D CALIC,后来Magli等人又对该方法进行了改进^[5],提出了一种可实时实现的三维CALIC压缩算法。Magli在线性预测过程中引入了卡尔曼滤波,取得了较高的无损压缩比^[6]。Zhang等提出了有效的波段分组排序方案,利用参考波段的预测误差线性反馈给当前波段,取得了一定的压缩效果^[7]。Mielikainen提出了差分脉编码调制(Differential Pulse Code Modulation, C-DPCM)算法^[8],引入了光谱分类作为预处理,结合多波段预测的方式,取得了极高的压缩比。文献[9]提出了一种有效的高光谱图像分布式信源编码方案。万建伟在文献[10]中对国内外有关高光谱图像压缩技术的研究进展情况进行了综述,这里不再过多概述。在借鉴C-DPCM的基础上,本文提出了一种基于内容的无损压缩算法。该算法首先利用C-means对光谱矢量进行

无监督分类,然后进行自适应波段分组;对于各组中的波段,分别选取各类地物的部分像素训练最优的预测系数,采用多波段预测的方式去除各类像素的相关性。若波段组中仅有一个波段,只进行类内预测。对AVIRIS(Airborne Visible/Infrared Imaging Spectrometer)与OMIS(Operational Modular Imaging Spectrometer)两种类型数据的仿真实验验证了算法的有效性。

2 高光谱图像的无监督分类

遥感图像分类是根据各像元的性质分为若干类别的过程,高光谱图像的分类是以每个像元矢量为基础进行的,目前已成为高光谱遥感领域研究的热点^[11]。星载遥感数据通常需要传输到地面才能进行相应的处理,遥感卫星只负责获取数据并传输至地面。目前,许多在轨的军事或商业遥感卫星已经能够在星上进行影像处理,美军的NEMO(Naval Earth Map Observer)卫星、欧空局的PROBA(Project for On-Board Autonomy)卫星以及德国宇航局的BIBD(Bispectral Infrared Detection)卫星都具备了地物分类、分析以及特征选择与提取等功能^[12],其优势是可以最大程度地减少人工干预。未来第四代对地观测卫星(“智能星”)的发展规划也将星上数据处理系统作为重要发展方向。

高光谱图像无监督分类的结果应用于无损压缩技术,实际上是一种基于内容的压缩方式。对于星载高光谱数据,事先无法获知地物类别的先验知识,因此,只能进行遥感影像的无监督分类。无监督分类是在无任何先验知识的前提下,仅仅利用遥感影像中各类地物光谱特征的分布规律,随其自然地进行分类,最终结果只是对地物类别进行了划分,无法确定类别的属性。

3 波段分组

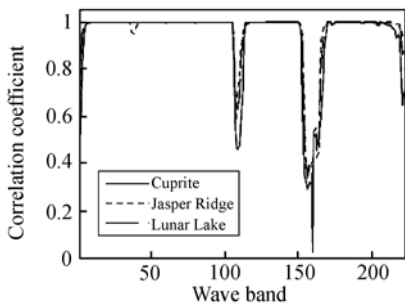
高光谱数据的格式可分为BSQ、BLP以及BIP3种格式,该文的实验数据为BSQ格式,即 $I_k(i, j)$ ($1 \leq k \leq L, 1 \leq i \leq M, 1 \leq j \leq N$)表示第 k 波段空间位置为 (i, j) 的像素,其中 L 为波段数, M 与 N 分别为波段的高度与宽度。若 $R_{u,v}$ 表示波

段 u 与 v 之间的互相关系数,其表达式为

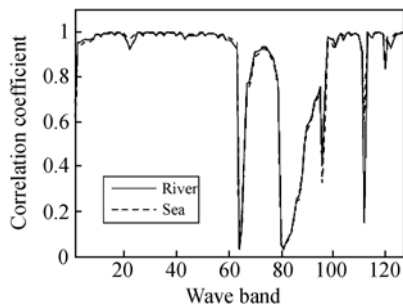
$$R_{u,v} = \frac{\sum_{i=1}^M \sum_{j=1}^N (I_u(i,j) - \bar{I}_u)(I_v(i,j) - \bar{I}_v)}{\sqrt{\sum_{i=1}^M \sum_{j=1}^N (I_u(i,j) - \bar{I}_u)^2 \sum_{i=1}^M \sum_{j=1}^N (I_v(i,j) - \bar{I}_v)^2}} \quad (1)$$

其中 \bar{I}_u 与 \bar{I}_v 分别为波段 u 和 v 的均值。

图 1 给出了 AVIRIS 与 OMIS-I 数据相邻波段的相关性曲线。从该图可知,高光谱图像中部分波段之间相关性较差,这部分波段基本上属于强噪声或者水吸收波段。图 2 给出了 Cuprite 与 River 在谱间预测前后的各波段熵值,可以看出, Cuprite 的 2~4、107~113、154~161 以及 165~167 等波段预测后熵值增加, River 的 65~69 以及 80~97 等波段预测后熵值增加。对照图 1 可以发现,这部分波段之间的相关性普遍较差,说明此类波段并不适合进行谱间预测,应采用谱内预测的方式去除冗余。因此,为了充分消除谱间相关性,需要在预测之前对高光谱图像进行自适应



(a) AVIRIS 数据
(a) AVIRIS data

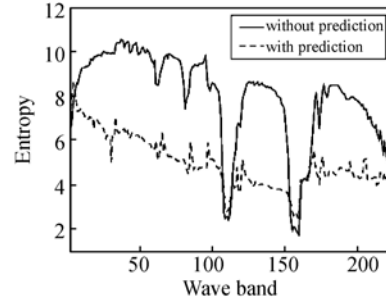


(b) OMIS-I 数据
(b) OMIS-I data

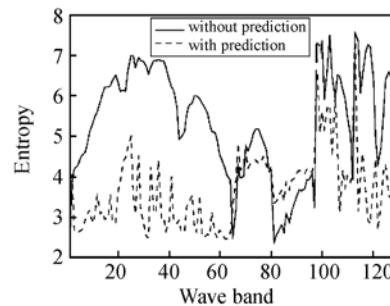
图 1 相邻波段的相关性曲线

Fig. 1 Correlation coefficient curves of adjacent bands

波段分组,其分组结果是将相关性较强的波段分到同一组,相关性较差的波段位于不同的组。选取互相关系数作为波段分组的准则,具体方法为:若 $R_{u,v} \geq T$,其中 T 为阈值,则波段 u 与 v 属于同一波段组(Group of Bands, GOB);否则,划分到不同的 GOB。



(a) AVIRIS 的 Cuprite 数据
(a) Cuprite data of AVIRIS



(b) OMIS-I 的 River 数据
(b) River data of OMIS-I

图 2 波段预测前后的熵值

Fig. 2 Predictive entropies of adjacent bands

4 波段选择与光谱分类

与全色图像相比,高光谱图像的空间分辨率相对较低,但其较高的光谱分辨率能够为不同地物的识别提供有利条件。若能够预先对高光谱图像进行无监督分类,根据不同地物的统计特性分别进行预测,可以显著改善预测性能。高光谱图像波段数量较多,为降低运算量,需要对高光谱图像进行降维。降维可看作是一个特征选择的过程,可以有效降低高光谱图像应用过程的不确定性。基于预测的方法在处理过程中不存在数据变换的问题,因此,采用基于波段选择的方法实施降维。

最简单的波段选择方法是在所有波段中任意选取部分波段,但无法充分挖掘高光谱图像的光谱信息。自适应波段选择(Adaptive Band Selection, ABS)方法充分考虑了每个波段的信息富集程度与相邻波段的相关性,所选取的波段更利于地物分类^[11]。

图 3 给出了各波段的标准差曲线,对于同一种成像光谱仪获取的不同数据,其标准差曲线以及相邻波段的相关性曲线是类似的。因此,各数据经过波段选择后的结果存在交集。针对高光谱数据的压缩,可以对该光谱仪获取的一定有代表性的数据进行波段选择,取其交集作为最终选择的波段,并可直接应用于该光谱仪获取的其它数据,从而避免了重复计算。采用 C-means 算法对选取的波段进行光谱分类,即可得到分类标识与类别索引。分类标识为 $H(i, j) = c, c \in \{1, 2, \dots, C\}$, 其中 C 为分类数。

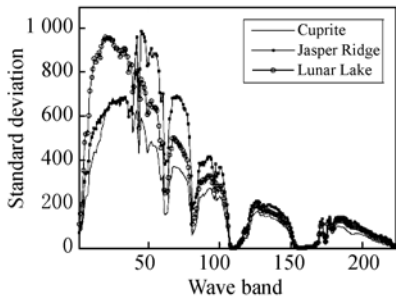


图 3 各波段的标准差

Fig. 3 Standard deviation of each band

5 谱间预测

高光谱图像谱间预测方式可分为单波段预测与多波段预测。前者仅仅利用单个波段进行预测,较少的参考信息使得其对参考波段的选取较为严格,往往需要通过高性能的波段排序算法获取理想的参考波段,而性能较好的波段排序算法一般都具有较高的运算复杂度;为了充分利用较高的光谱分辨率,对当前波段采用多波段线性预测的方案。在光谱分类的基础上,针对每个分类,需要选取该类的像素训练预测系数。若第 k 波段为当前波段, $I(k, i, j) (H(i, j) = c, 1 \leq c \leq C)$ 为当前待预测像素,利用与第 k 波段相邻的前 K 波段

进行预测, $I(k, i, j)$ 的预测值为

$$\hat{I}(k, i, j) = \lfloor \sum_{l=1}^K \alpha_l^c I(k-l, i, j) + 0.5 \rfloor, \quad (2)$$

其中 $\alpha_c = [\alpha_1^c, \alpha_2^c, \dots, \alpha_K^c]^T$ 为第 c 类的预测系数。由于同类像素具有相近的光谱特性,因此,在训练样本的选择上,只需在第 c 类中选取部分像素即可。若选取的像素数为 λ , 表示为 $\mathbf{Y} = [y_{k,1}^c, y_{k,2}^c, \dots, y_{k,\lambda}^c]^T$, 构造矩阵 \mathbf{P} 与 \mathbf{Q}

$$\mathbf{P} = \begin{bmatrix} y_{k-1,1}^c & \cdots & y_{k-K,1}^c \\ \vdots & \ddots & \vdots \\ y_{k-1,\lambda}^c & \cdots & y_{k-K,\lambda}^c \end{bmatrix}, \mathbf{Q} = \begin{bmatrix} y_{k,1}^c \\ \vdots \\ y_{k,\lambda}^c \end{bmatrix}, \quad (3)$$

根据最小二乘准则可得第 $c (1 \leq c \leq C)$ 类的最优预测系数为

$$\alpha_c = (\mathbf{P}^T \mathbf{P})^{-1} (\mathbf{P}^T \mathbf{Q}), \quad (4)$$

求解 α_c 的运算量随着 γ 和 K 的增大而增加。C-DPCM 在光谱分类的基础上,参考波段数量为 20 时达到最优,该算法的计算量过大,实际工作中难以应用。本文借鉴 C-DPCM 的基础上,综合考虑压缩性能和运算复杂度两方面因素,选取 $K=2$ 。预测残差为

$$e_c(k, i, j) = I(k, i, j) - \hat{I}(k, i, j). \quad (5)$$

对于各个分类,分别得到预测残差。针对只有一个波段的 GOB,采用类内预测的方式去除该波段的空空间相关性,类内预测算法采用的是 JPEG 标准中的无损压缩部分。

除了预测残差外,分类图与预测系数需要作为附加信息传送到解码端,因此,需要选取一种高效的无损压缩算法对以上信息实施无损压缩。目前,在诸多无损压缩标准中, JPEG-LS 取得了较为理想的无损压缩效果^[14],在无损压缩性能上优于 JPEG2000 标准,并且具有较高的运算速度。因此,选取 JPEG-LS 对预测残差与附加信息进行无损压缩。综上所述,谱间预测流程如图 4 所示,其中灰色像素为同类地物。

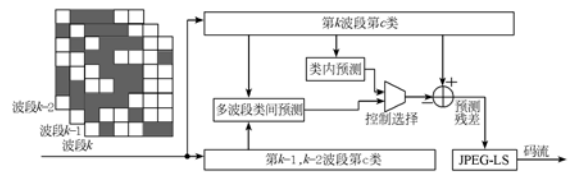


图 4 谱间预测流程图

Fig. 4 Flowchart of spectral prediction

6 实验结果

为验证所提算法的有效性,利用 VC++ 软件对 AVIRIS 采用推扫模式获取的高光谱数据进行压缩测试。AVIRIS 是由美国宇航局所属的喷气动力实验室开发,所获取的数据在波长 $0.4\sim 2.5\ \mu\text{m}$

共有 224 个波段,光谱分辨率为 $10\ \text{nm}$,空间分辨率为 $20\ \text{m}\times 20\ \text{m}$,每个像素存为 2 个字节。选取 Cuprite、Jasper Ridge (J. R.)、Lunar Lake (L. L.)、Low Altitude (L. A.) 以及 Moffett Field (M. F.) 5 个场景,波段大小为 512×512 ,图 5 给出了这 5 个场景的第 40 个波段图像,图 6 给出了光谱分类后的结果。

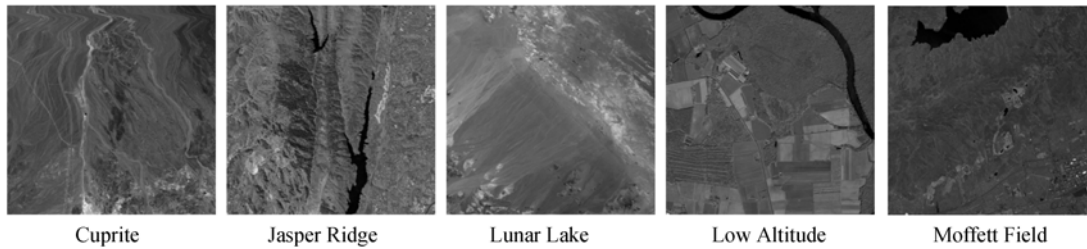


图 5 AVIRIS 数据的第 40 个波段原始图像

Fig. 5 AVIRIS images of band 40

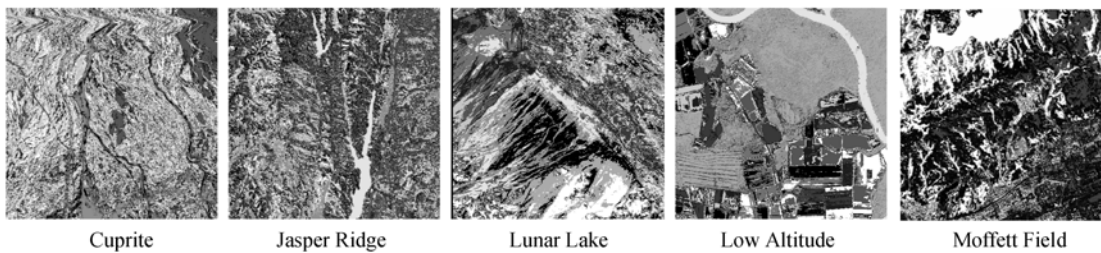


图 6 AVIRIS 数据光谱分类后的结果

Fig. 6 Classified results of AVIRIS

实验参数为: $T=0.8$ 、 $C=10$,波段选择后的结果为 $16\sim 35$,在每个类中选取像素总数的 30% 进行预测系数的训练。采用压缩比(Compression Ratio, CR)作为评价无损压缩性能的标准,表 1 给出了几种具有代表性的无损压缩算法的性能比较。JPEG-LS 所对应的是各波段单独压缩后的结果;Differential JPEG-LS 将相邻波段进行差分,再分别进行 JPEG-LS 压缩;3D 多级树集合分裂(SPIHT)算法是三维小波变换基础上的等级树编码;分布式信源编码(Distributed Source Coding, DSC)是文献[9]提出的高光谱图像分布式无损编码;基于内容的多波段预测(Content, Multi-bands Prediction, CMP)为本文提出的方案;多波段预测(Multi-bands Prediction, MP)是一种多波段预测方案,与 CMP 相比,MP 未进行分类预测,其它步骤与 CMP 相同。

表 1 AVIRIS 数据的压缩结果比较

Tab. 1 Compression ratio of various algorithms on the OMIS-I

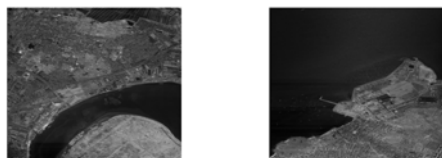
	JPEG-LS	Differential JPEG-LS	3D SPIHT	DSC	MP	CMP
Cuprite	2.12	2.77	2.86	2.99	3.05	3.17
J. R.	2.00	2.70	2.77	2.93	3.00	3.07
L. L.	2.14	2.71	2.85	2.94	2.96	3.08
L. A.	2.15	2.50	2.67	2.79	2.75	2.89
M. F.	1.98	2.51	2.86	2.98	3.05	3.16

通过表 1 可以看出, JPEG-LS 未考虑谱间相关性,压缩性能较低;Differential JPEG-LS 简单去除了谱间相关性,压缩性能优于 JPEG-LS。3D SPIHT 的压缩性能一般。DSC 的性能优于 3D SPIHT,表明基于预测的方法优于基于变换的方

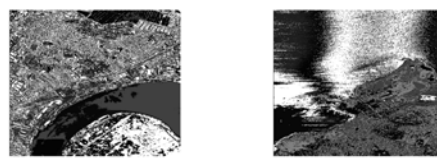
法,但 DSC 仅利用一个参考波段进行预测,MP 采用的是多波段预测方式,压缩性能有所提升,但 MP 利用整个波段的统计特性训练预测系数,压缩性能受到限制。CMP 的压缩性能明显优于 MP,充分说明分类预测可以提高预测性能,验证了所提算法的有效性。

为进一步验证 CMP 算法的有效性,对 OMIS-I 获取的高光谱图像进行仿真实验。OMIS-I 是上海技术物理研究所研制的中分辨率

成像光谱仪,其覆盖了从可见光到热红外的光谱范围,共 128 个波段,其中 $0.46 \sim 1.1 \mu\text{m}$ (10 nm),64 个波段; $1.06 \sim 1.70 \mu\text{m}$ (40 nm),16 个波段; $2.0 \sim 2.5 \mu\text{m}$ (15 nm),32 个波段; $3.0 \sim 5.0 \mu\text{m}$ (250 nm),8 个波段; $8.0 \sim 12.5 \mu\text{m}$ (500 nm),8 个波段。对 River 与 Sea 两个数据进行了仿真实验,图像大小为 512×512 ,每个像素存为一个字节。图 7 给出了 River 和 Sea 的第 40 个波段图像,图 8 给出了光谱分类后的结果。



River Sea
图 7 OMIS-I 数据的第 40 个波段
Fig. 7 OMIS-I images of band 40



River Sea
图 8 OMIS-I 数据光谱分类后的结果
Fig. 8 Classified results of OMIS-I

表 2 所示为不同算法对 OMIS-I 数据的压缩结果对比,显然,CMP 的压缩性能明显优于其它几种算法,与 AVIRIS 数据的压缩结果吻合,进一步验证了 CMP 算法的有效性。

表 2 OMIS-I 数据的压缩结果比较

Tab. 2 Compression ratio of various algorithms on the AVIRIS

	JPEG-LS	Differential JPEG-LS	3D SPIHT	DSC	MP	CMP
River	1.96	2.19	2.22	2.07	2.32	2.38
Sea	1.96	2.05	2.15	2.00	2.20	2.27

参考文献:

- [1] JENSEN J R. 遥感数字影像处理导论[M]. 北京:机械工业出版社,2007.
JENSEN J R. *Introductory Digital image Processing: A Remote Sensing Perspective* [M]. Beijing: China Machine Press, 2007. (in Chinese)
- [2] 孙蕾,谷德峰,罗建书,等. 最佳递归双向预测的高光谱图像无损压缩[J]. 光学精密工程,2009,17(11):2864-2870.

7 结 论

本文提出了一种基于内容的高光谱图像无损压缩算法,该算法首先对高光谱图像实施无监督分类;在自适应波段分组与单调后向预测的基础上,采用多波段同类点预测的方案去除同类像素的谱间相关性,取得了较好的无损压缩效果。随着航天高光谱遥感技术的不断发展,星载高光谱数据压缩技术已经成为研究的热点,受卫星上处理内存、计算能力以及功耗等的限制,需要在保证一定压缩性能的前提下,尽可能降低算法的运算量,同时,压缩码流还应具有良好的抗误码能力,这些工作需要在今后的工作中加以研究与解决。

- SUN L, GU D F, LUO J SH. Hyperspectral image lossless compression based on optimal recursive bi-directional prediction [J]. *Opt. Precision Eng.*, 2009, 17(11): 2864-2870. (in Chinese)
- [3] WU X L, MEMON N D. Context-based, adaptive lossless image coding [J]. *IEEE Transactions on Communications*, 1997, 45(4): 437-444.
- [4] MEMON X, WU N D. Context based lossless intraband adaptive compression-extending calic [J]. *IEEE Transactions on Geoscience and Remote Sens-*

- ing, 2000, 9: 994-1001.
- [5] MAGLI E, OLMO G, QUACCHIO E. Optimized onboard lossless and near-lossless compression of hyperspectral data using CALIC[J]. *IEEE Geoscience and Remote Sensing Letters*, 2004, 1(1): 21-25.
- [6] MAGLI E. Multiband lossless compression of hyperspectral images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2009, 47(4): 1168-1178.
- [7] ZHANG J, 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.
- [8] MIELIKAINEN J, TOIVANEN P. Clustered DPCM for the lossless compression of hyperspectral images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2003, 41(12): 2943-2946.
- [9] ABRARDO A, BARNI M, MAGLI E, *et al.*. Error-resilient and low-complexity onboard lossless compression of hyperspectral images by means of distributed source coding[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2010, 48(4): 1892-1904.
- [10] 万建伟, 粘永健, 苏令华, 等. 高光谱图像压缩技术研究进展[J]. *信号处理*, 2010, 26(9): 1397-1404. WAN J W, NIAN Y J, SU L H, *et al.*. Research progress on hyperspectral imagery compression technique[J]. *Signal Processing*, 2010, 26(9): 1397-1404. (in Chinese)
- [11] 高恒振, 万建伟, 粘永健, 等. 组合核函数支持向量机高光谱图像融合分类[J]. *光学 精密工程*, 2011, 19(4): 878-883. GAO H Z, WAN J W, NIAN Y J, *et al.*. Fusion classification of hyperspectral image by composite kernels support vector machine[J]. *Opt. Precision Eng.*, 2011, 19(4): 878-883. (in Chinese)
- [12] 张永生, 巩丹超, 刘军, 等. 高分辨率遥感卫星应用-成像模型、处理算法及应用技术[M]. 北京: 科学出版社, 2004. ZHANG Y S, GONG D C, LIU J, *et al.*. *High-resolution Remote Sensing Satellite Data: Imaging Model, Processing Algorithms, and Application Technology*[M]. Beijing: Science Press, 2004. (in Chinese)
- [13] 刘春红, 赵春晖, 张凌雁, 等. 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. LIU CH H, CHAO CH H, ZHANG L Y. A new method of hyperspectral remote sensing image dimensional reduction[J]. *Journal of Image and Graphics*, 2005, 10(2): 218-222. (in Chinese)
- [14] HP Labs LOCO-I/JPEG-LS Home Page. [OL]. <http://www.hpl.hp.com/loco>.

作者简介:



汤毅(1983—),男,湖北公安人,博士生,2008年与国防科学技术大学获得硕士学位,主要从事遥感图像处理方面的研究。E-mail: lantange@163.com



辛勤(1973—),男,江西南昌人,博士,副教授,2001年于国防科技大学信息与通信工程专业获工学博士学位,主要从事遥感图像处理方面的研究。E-mail: ym_xin@163.com



万建伟(1964—),男,江西南昌人,博士,教授,1998年于国防科技大学获信息与通信工程学科工学博士学位,主要从事雷达信号处理,现代信号处理,遥感图像处理等方面的研究。E-mail: kermittwan@hnmcc.com



李纲(1966—),男,湖南邵阳人,硕士,副教授,主要从事现代信号处理方面的研究。E-mail: ligang@nudt.edu.cn