

“十二五”国家重点图书

Springer 精选翻译图书

卫星数据压缩

Satellite Data Compression

[美] Bormin Huang 主编

陈浩 周爽 译



哈尔滨工业大学出版社
HARBIN INSTITUTE OF TECHNOLOGY PRESS

“十二五”国家重点图书

Springer 精选翻译图书

卫星数据压缩

Satellite Data Compression

[美] Bormin Huang 主编

陈浩 周爽 译



哈尔滨工业大学出版社
HARBIN INSTITUTE OF TECHNOLOGY PRESS

内 容 提 要

随着星载传感器空间、光谱、时间等分辨率的不断提高,获取的传感器数据量海量地增长,卫星数据压缩已成为遥感数据传输和存储的必要手段,也一直是遥感领域的研究热点之一。全书共分为 14 章,涵盖了目前国际上关于卫星数据压缩的有损、近无损、无损压缩等多种典型方法,既有理论算法也包括部分硬件实现。可以使读者对该领域的研究具有较深入和全面的认识。

本书可作为高等院校信息、通信、遥感及相关专业的教材或教学参考书,也可供相关科研人员参考。

黑版贸审字 08-2012-049 号

Translation from English language edition:

Satellite Data Compression

by Bormin Huang

Copyright © 2011, Springer New York

Springer New York is a part of Springer Science+Business Media

All Rights Reserved

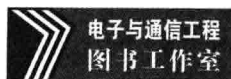
图书在版编目(CIP)数据

卫星数据压缩/(美)黄柏铭(Bormin Huang)主编;陈浩,周爽译. —哈尔滨:
哈尔滨工业大学出版社,2014.5

ISBN 978-7-5603-4733-2

I. ①卫… II. ①黄…②陈…③周… III. ①遥感卫星-数据
压缩 IV. ①V474.2

中国版本图书馆 CIP 数据核字(2014)第 095178 号



责任编辑 李长波

封面设计 刘长友

出版发行 哈尔滨工业大学出版社

社 址 哈尔滨市南岗区复华四道街 10 号 邮编 150006

传 真 0451-86414749

网 址 <http://hitpress.hit.edu.cn>

印 刷 哈尔滨市石桥印务有限公司

开 本 660mm×980mm 1/16 印张 19.75 字数 355 千字

版 次 2014 年 5 月第 1 版 2014 年 5 月第 1 次印刷

书 号 ISBN 978-7-5603-4733-2

定 价 40.00 元



(如因印装质量问题影响阅读,我社负责调换)

译者前言

在 2011 年,我有幸与威斯康星大学麦迪逊分校空间科学与工程中心高级研究员 Bormin Huang(黄柏铭)博士取得了联系,黄博士现为国际光学工程学会(SPIE)的 Fellow,是卫星遥感图像压缩等领域的著名专家。至今,已经荣幸地邀请黄博士多次来到哈尔滨工业大学进行学术交流。在 2012 年的交流期间,获悉黄博士于 2011 年主编出版了 *Satellite Data Compression* 一书。在黄博士的推荐下,拜读了该书,并达成了翻译该书的协议。该书共分为 14 章,包含了目前国际上卫星数据压缩领域的典型方法,涉及了有损、近无损、无损压缩的理论算法和硬件实现,具有较强的学术研究和实践参考价值。该书对于我一直从事的数据压缩的研究具有很大的推动作用。2013 年 2 月至 2014 年 2 月,我作为访问学者到威斯康星大学麦迪逊分校空间科学与工程中心,与黄博士研究团队继续卫星数据压缩方向的研究,该阶段的研究过程进一步加深了我对该书内容的理解。

全书由哈尔滨工业大学陈浩对译稿进行统稿,周爽负责了部分章节的翻译工作。在整个翻译过程中,博士研究生石翠萍、ALIS-SOU SIMPLICE,硕士研究生雷威、陶宇、李亚乾、刘东、王佳斌、张孟、滑艺、王怡然等做了许多有价值的工作,在此一并表示感谢。

需特别感谢的是,张晔教授对于该书翻译的支持。

由于该书原著涉及多个国家学者的研究成果,理论、方法、研究内容较多,在全书翻译过程中在某些理解和用词上难免有不恰当或者不足之处,敬请有关专家和学者批评指正。

陈 浩

2014 年 3 月

于哈尔滨工业大学

目 录

第 1 章 加拿大航天局星载数据压缩技术的进展	1
1.1 加拿大航天局对卫星数据压缩方面研究的回顾	1
1.1.1 无损压缩	1
1.1.2 小波变换有损数据压缩	2
1.1.3 矢量量化数据压缩	2
1.2 近无损压缩技术: SAMVQ 和 HSOCVQ	5
1.2.1 连续近似多级矢量量化(SAMVQ)	6
1.2.2 分等级的自组织聚类矢量量化(HSOCVQ)	6
1.3 评估 SAMVQ 和 HSOCVQ 的近无损特征	7
1.4 异常对压缩性能的影响	9
1.5 预处理和辐射转换对压缩性能的影响	11
1.6 Keystone 和 Smile 现象对压缩性能的影响	13
1.7 压缩数据的多学科用户可接受性研究	15
1.8 压缩技术适应比特错误能力的增强	16
1.9 星载原型压缩机的发展	17
1.10 参与卫星数据系统国际标准的发展	20
1.11 结论	22
参考文献	24
第 2 章 法国航天局对高分辨率卫星图像星载压缩的研究	30
2.1 引言	30
2.2 星载压缩算法的历史及现状	32
2.2.1 最初的压缩器	32
2.2.2 基于离散余弦变换的压缩器	33
2.2.3 基于小波变换的压缩器	34
2.2.4 空间应用标准: CCSDS 推荐标准	34
2.2.5 图像质量评价	35

2.3 星载压缩算法的现状与发展	36
2.3.1 多光谱压缩	36
2.3.2 小波的限制	36
2.3.3 一种新的星载压缩变换方法	37
2.3.4 星载压缩的商用成品	38
2.3.5 离散小波变换算法的特殊处理	40
2.3.6 选择性压缩	40
2.3.7 固定质量:可变数据率压缩	43
2.4 结论	44
参考文献	45
第3章 低复杂度的无损及近无损高光谱图像压缩方法	47
3.1 引言	47
3.2 背景	48
3.2.1 压缩技术	48
3.2.2 国际标准	49
3.3 星上压缩要求	50
3.4 基于块的编码方法	52
3.4.1 预测	52
3.4.2 近无损压缩	53
3.4.3 熵编码	54
3.4.4 波段重排序	55
3.4.5 复杂度	56
3.5 压缩性能	56
3.5.1 数据集描述	56
3.5.2 AVIRIS	57
3.5.3 AIRS	58
3.5.4 近无损压缩	59
3.6 硬件实现	60
3.7 结论	62
参考文献	62
第4章 星载图像压缩无链表 SPIHT 的 FPGA 设计	66
4.1 引言	66

4.2	SPIHT	67
4.2.1	渐进图像传输	68
4.2.2	集分割排序算法	68
4.2.3	编码算法	70
4.3	无链表 SPIHT 及 FPGA 实现	71
4.3.1	LZC	71
4.3.2	提出的无链表 SPIHT	72
4.3.3	性能分析	74
4.3.4	FPGA 实现	77
4.4	结论	82
	参考文献	83
第 5 章	适应异常值的熵编码	85
5.1	引言	85
5.2	Rice 和空间数据咨询委员会 (CCSDS 121.0) 的限制	86
5.3	适应异常值码	88
5.3.1	子指数码	88
5.3.2	有限长度有符号 Rice 编码	89
5.3.3	预测误差编码器 (PEC)	90
5.4	分析编码	93
5.4.1	输入数据建模	93
5.4.2	性能度量	95
5.4.3	Rice-Golomb 编码	95
5.4.4	子指数码	96
5.4.5	有限长度有符号 Rice 编码	96
5.4.6	PEC 编码	97
5.5	自适应编码器	99
5.5.1	自适应子指数	99
5.5.2	自适应有限长度有符号 Rice 编码器	100
5.5.3	完全自适应 PEC (FAPEC)	100
5.6	测试编码器	101
5.6.1	合成测试	101
5.6.2	数据压缩集	102

5.7 结论	107
参考文献	109
第6章 通过频谱自适应 DPCM 高光谱图像压缩的质量问题	112
6.1 引言	112
6.2 无损/近无损图像压缩算法	113
6.2.1 基于预测的 DPCM	114
6.2.2 基于上下文模型的熵编码	115
6.3 基于 DPCM 的高光谱数据压缩	115
6.4 明确/模糊分类 3D/光谱预测	116
6.4.1 S-FMP 和 S-RLP 基本定义	116
6.4.2 初始化	117
6.4.3 预测的迭代细化	118
6.4.4 基于上下文的算术编码	121
6.5 基于 LUTs 的无损高光谱图像压缩	122
6.6 近无损压缩	124
6.6.1 失真测量	125
6.6.2 线性和对数量化	127
6.7 几乎无损压缩	128
6.8 实验结果与比较	130
6.8.1 无损压缩性能比较	130
6.8.2 近无损压缩性能比较	134
6.8.3 材料辨别	137
6.9 结论	138
参考文献	138
第7章 基于预测下三角变换的超光谱探测器数据压缩	144
7.1 引言	144
7.2 数据	145
7.3 压缩系统	147
7.4 实验结果	153
7.5 结论	159
参考文献	160

第 8 章 基于查找表的高光谱数据压缩	162
8.1 引言	162
8.2 高光谱图像的无损压缩	163
8.3 LUT 方法	164
8.4 预测指引的 LUT 方法	166
8.5 对应位置像素的均匀量化	167
8.6 多波段 LUT	170
8.7 实验结果	170
8.8 结论	176
参考文献	176
第 9 章 用于有损到无损高光谱图像压缩的无乘法器可逆整数 TDLT/KLT	180
9.1 引言	180
9.2 多重提升方案	182
9.2.1 提升方案和应用	183
9.2.2 基于多重提升方案的可逆整数变换	184
9.3 可逆的整数到整数 TDLT/RKLT	185
9.3.1 矩阵分解和多重提升	186
9.3.2 可逆的整数到整数的 TDLT/RKLT 设计	189
9.4 基于 RTDLT/RKLT 的高光谱图像压缩的实验结果和讨论	194
9.5 基于 3D-RLT 的高光谱图像压缩	201
9.6 基于 3D-RLT 的高光谱图像压缩实验结果	201
9.7 结论	205
参考文献	206
第 10 章 基于分治法去相关的高光谱数据压缩	210
10.1 引言	210
10.2 Karhunen-Loève Transform 变换	211
10.2.1 中心化和协方差特殊性	212
10.2.2 空间划分	213
10.2.3 可逆 KLT	213
10.3 分治策略	215
10.3.1 普通分集	215

10.3.2	递归结构	216
10.3.3	两层结构	217
10.3.4	多层结构	218
10.4	实验结果	219
10.5	结论	224
	参考文献	225
第 11 章	基于分段主成分分析的高光谱图像压缩	228
11.1	引言	228
11.2	分段主成分分析	232
11.3	SPCA+JPEG 2000	233
11.4	SPCA+JPEG 2000 的数据分析性能	238
11.4.1	光谱保真度	238
11.4.2	无监督分类	239
11.4.3	异常检测	240
11.4.4	无监督线性解混	241
11.5	SPCA+JPEG 2000 的特定传感器次最优分段方法	242
11.6	结论	246
	参考文献	246
第 12 章	基于优化比特分配的快速预计算矢量量化超光谱探测器数据无损压缩	250
12.1	引言	250
12.2	数据	251
12.3	压缩方法	252
12.4	实验结果	256
12.5	结论	261
	参考文献	262
第 13 章	有损压缩对高光谱分类的影响	264
13.1	引言	264
13.2	压缩和分类精度	265
13.3	特征提取和压缩	268
13.4	保持判别特征	269
13.4.1	判别特征预增强	269

13.4.2 增强判别信息光谱波段·····	273
13.5 结论·····	277
参考文献·····	278
第 14 章 基于投影寻踪法的高光谱图像降维 ·····	282
14.1 引言·····	282
14.2 基于投影寻踪成分分析法的降维·····	284
14.3 基于指标投影的优先级排序 PP ·····	285
14.4 驱动初始化 PIPP ·····	287
14.5 真实的高光谱图像实验·····	288
14.5.1 具有随机初始条件的 PIPP ·····	290
14.5.2 PI-PRPP ·····	294
14.5.3 ID-PIPP ·····	297
14.6 结论·····	301
参考文献·····	302

第 1 章 加拿大航天局星载数据压缩技术的进展

摘要 本章回顾和概括了加拿大航天局在与加拿大其他政府部门、学术机构和工业机构的合作下,对卫星传感器数据压缩的研究和发展情况。本章通过以下几节对该问题进行阐述。

1.1 加拿大航天局对卫星数据压缩方面研究的回顾

加拿大航天局(Canadian Space Agency, CSA)在 20 世纪 90 年代开始研究数据压缩算法作为一项使能技术用于高光谱卫星,这种研究包括无损和有损压缩技术。在实际运行中使用的压缩技术也得到了研究^[1-16],研究的核心是基于矢量量化(Vector Quantization, VQ)的近无损数据压缩技术。由于矢量量化非常简单,并且其矢量性质可用于保存与场景中单个地面样本有关的光谱特征,因此矢量量化是对高光谱图像数据压缩的一种有效编码技术。对大多数遥感应用和算法而言,都能够取得合理的高压缩比($>10:1$)以及高压缩保真度。

1.1.1 无损压缩

在 20 世纪 90 年代,采用了一种基于预测的方法。它使用在空间域或/和光谱域中的线性或非线性预测器来产生预测,并应用 DPCM 产生残余,其后进行熵编码。研究了包括 1D、2D 和 3D 情况的共 99 个固定系数的预测器^[15],包括一种从预先选定的预测器组中选择最好的预测器的自适应预测器。空间数据系统咨询委员会(Consultative Committee for Space Data System, CCSDS)推荐的无损算法(主要是 Rice 算法^[17])被选作熵编码器,用于编码预测残差。为了评估 CCSDS 无损算法的性能,使用一种被称为基础比特加溢出比特编码(Base-bit Plus Overflow-bit Coding, BPOC)^[18]的熵编码器来比较熵编码效率。

对使用机载可见光/近红外成像光谱仪(Airborne Visible/Near Infrared Imaging Spectrometer, AVIRIS)和轻便机载光谱成像仪(Compact Airborne Spectrographic Imager, CASI)获得的三个高光谱数据立方体进行测试。两个在

3D 空间分别使用五个和七个最近邻像素的 3D 预测器,产生了最佳的残差。在由熵编码器编码产生编码残差之后,取得了 2.4 左右的压缩比,而其他的预测器产生的压缩比均小于 2 : 1。BPOC 略优于 CCSDS 无损算法。

1.1.2 小波变换有损数据压缩

我们也使用小波变换技术为高光谱数据研究了一种压缩技术^[16]。由 Shapiro 提出的创建零树的方法^[19]被修改为用于高光谱数据,提出了一种优化的多层查找表的方法,它改善了嵌入零树小波算法的性能。为了评估该算法的性能,将算法与 SPIHT^[20]和 JPEG 相比较。经比较,这种新的算法等同于或优于已发表的算法,因为这些算法计算复杂度更高,不适于高光谱数据的压缩。对 AVIRIS 和 CASI 数据进行了如 1.1.1 节提到的实验,得到了超过 32 : 1 的压缩比,保真度大于 40.0 dB。

1.1.3 矢量量化数据压缩

我们选择矢量量化技术用于遥感图像的数据压缩,是由于矢量量化非常简单并且其矢量性质可用于保存与场景中单个地面样本有关的光谱特征。矢量量化是一种有效的编码技术。广义 Lloyd 算法 (Generalized Lloyd Algorithm, GLA) 有时也称 LBG,是用于图像压缩最经常使用的 VQ 算法。

在应用中,将用于 VQ 算法的矢量和高光谱数据立方体中每个地面样本的全光谱维(被称为光谱矢量)联系起来。因为数据立方体场景中的目标数目是有限的,用于训练的光谱矢量的数目应比数据立方体中光谱矢量的总数要小得多。因此,可以使用码书表示所有的谱向量,该码书具有相对少的码向量,并取得较好的重建保真度。

VQ 压缩技术较好地利用了光谱域波段之间的高相关性,取得了高压缩比。然而,从运行使用的角度看,用于高光谱图像的 VQ 压缩技术面临很大的挑战,主要在于它需要大量的运算资源,特别是在码书的生成阶段。因为高光谱图像数据立方体大小是传统遥感数据的几百倍,因此用于训练码书或者使用码书对数据立方体编码的处理时间也应是传统遥感数据的几十到几百倍。在一些应用中,通过仅仅对码书训练一次,训练时间的问题可以很大程度地被避免,以后重复地对后面所有需要压缩的数据立方体应用该码书,就像在传统的 2D 图像压缩中所采取的方法一样。若待压缩的数据立方体是由用于训练码书的训练集为限的,上述方法会工作得很好。

然而,在高光谱遥感中,获取一个能够使多个数据立方体同时达到需要保

真程度的所谓“通用的”码书非常困难。部分原因是由于目标特性(季节、位置、亮度、视角、精确的大气影响的需要)以及仪器配置(光谱分辨率和空间分辨率、光谱范围、仪器的 SNR)会使数据立方体具有高的变化性,还有部分原因是在其下游使用中高重建高保真度的需要。基于这些原因,倾向于为每一个待压缩的数据立方体生成一个新的码书,并与索引图一起作为压缩数据传到解码器。因此,压缩高光谱图像的基于 VQ 方法发展的主要目的是寻找一种更快、更有效的压缩算法来克服这种挑战,特别是在星载应用中。在这一章中,使用传统的 GLA 算法和一种用于数据立方体的新训练码书进行高光谱数据立方体的压缩,称为 3DVQ。

加拿大航天局开发了一种代表高光谱数据立方体光谱向量的有效方法,被称为基于光谱特征的二进制编码(Spectral Feature Based Binary Code, SFBBC)^[1]。使用 SFBBC 码和汉明距离,而不是欧几里得距离,可以用于高光谱图像 VQ 压缩过程中的码书训练和码向量匹配(编码)。汉明距离是逻辑按位异或运算的简单的和,这比计算欧几里得距离要快得多。与 3DVQ 相比,基于 SFBBC 的 VQ 压缩技术可以在保真度损失 PSNR<1.5 dB 的前提下加速压缩过程 30~40 倍。这种压缩技术也称为基于 SFBBC 的 3DVQ。

后来,提出了相关矢量量化(Correlation Vector Quantization, CVQ)^[2]。它使用一个可移动的窗口去覆盖高光谱数据立方体中相邻地面 2×2 的光谱矢量块,并同时移除数据立方体中的光谱相关和空间相关性。编码时间(CT)可以通过因子 $1/(1-\beta)$ 来改善,这里 β 表示窗口中的一个光谱矢量可以用窗口中三个已编码光谱矢量中的一个来近似的概率。实验结果表明,采用 2 左右的因子,可以提高编码的时间,并且与使用 3DVQ 相比,压缩比提高了 30%。CVQ 可以与 SFBBC 结合使用,以进一步提高 3DVQ 的编码时间^[3]。

码书的生成阶段是一个迭代的过程,并且决定着 3DVQ 整个的处理时间,因此减少码书生成的时间(Codebook Generation Time, CGT)非常重要。由于 CGT 大致和训练集的大小成正比,可以通过减少训练集的大小来得到一个更快的压缩系统。为码书生成构建小而有效的训练集,加拿大航天局研发了三种光谱矢量选择方案以下采样待压缩的数据立方体中的光谱矢量。数值分析表明,4% 的下采样率对于保持重建保真度以及减少 CGT 来说是最优的。实验结果表明,当训练集是由以 2.0% 的下采样的数据立方体组成,利用因子为 15.6~17.4,处理时间能得到改善,PSNR 损失 0.6~0.7 dB^[4]。

使用包含在待压缩的高光谱数据立方体中的遥感信息,加拿大航天局进一步改善了 3DVQ。引入一种光谱索引,例如归一化植被指数(Normalized Difference Vegetation Index, NDVI),来改善压缩算法。提出了一种新的基于

VQ 的压缩技术,被称为基于光谱索引的多子码书算法(Multiple Sub-Codebook Algorithm, MSCA)^[5]。首先为待压缩的数据立方体创建一个光谱索引图,然后基于索引值将其分割成 n (通常是 8 或 16) 个区域,根据分割了的索引图,数据立方体被分成 n 个子集,每个子集对应一个区域(或类)。为每个子集训练一个独立的码书,并用于压缩对应的子集。通过一个大小为 n 左右的因子,MSCA 能够加速 CGT 和 CT。实验结果表明,当测试数据立方体的场景被分割成 16 个区域,因子为 14.1 和 14.8 时,CGT 和 CT 都得到了改善,而重建保真度和使用 3DVQ 的结果几乎相同。

通过结合以往研究的快速 3DVQ 技术,创建和测试用于高光谱图像的三种 VQ 数据压缩系统:SFBBC、下采样和 MSCA。仿真结果表明,CGT 减少了超过三个数量级,而码书的质量依然很好。通过一个 1 000 左右的因子,3DVQ 的总体处理速度得到提高,平均 PSNR 损失小于 1.0 dB^[6]。

文献[7]提出了一种用于基于 VQ 压缩算法的快速搜索方法。该方法利用了这样一个事实:在 GLA 的全搜索中,一个训练矢量,如果和前面的迭代相比,得到分割的距离在当前迭代中提高了,则不需要搜索以发现最小距离分割。该方法的优点是简单,节省了大量运算时间,并且产生的压缩保真度和 GLA 一样好。对涵盖多种场景类型的四个高光谱数据立方体进行测试。实验结果表明,对于四个测试高光谱数据立方体,码书的大小为 16 ~ 2 048,提出的方法提高了 3.08 ~ 27.35 倍的压缩时间。码书的尺寸越大,节约的时间越多,而由压缩导致的光谱信息损失可以通过光谱角匹配和遥感应用来评价。

接着文献[7]中的工作,加拿大航天局进一步为矢量压缩技术改善了搜索方法^[8]。它利用了这样一个事实:在 GLA 中,训练序列中的一个矢量要么位于与前一次迭代相同的最小距离分割(Minimum Distance Partition, MDP),要么位于一个很小的分割子集内。提出的方法仅仅在这个分割子集内及上次迭代得到的 MDP 中为训练矢量寻找 MDP。由于子集的大小比码向量的总数要小,搜索过程能得到显著加速。提出的方法生成的码书与使用 GLA 生成的码书一致。实验结果表明,当训练从 16 到 2 048 大小的码书时,对于两个测试数据集,码书训练的计算时间分别从 6.6 倍提高到 50.7 倍以及从 7.7 倍提高到 58.7 倍。当结合文献[7]提出的快速搜索方法时,对两个测试数据集,计算时间分别从 7.7 倍提高到 58.7 倍,从 13.0 倍提高到 128.7 倍。

图 1.1 给出了与 3DVQ 相比,前面提到的各种快速 VQ 技术以及这些技术结合后的处理速度改善情况。采用单一的快速技术,处理速度能够提高 70 倍。当三种加速技术结合使用时,处理速度能够提高 1 000 倍左右。当使用 SFBBC 技术时,与 3DVQ 相比,快速技术的 PSNR 保真度损失小于 1.5 dB。当

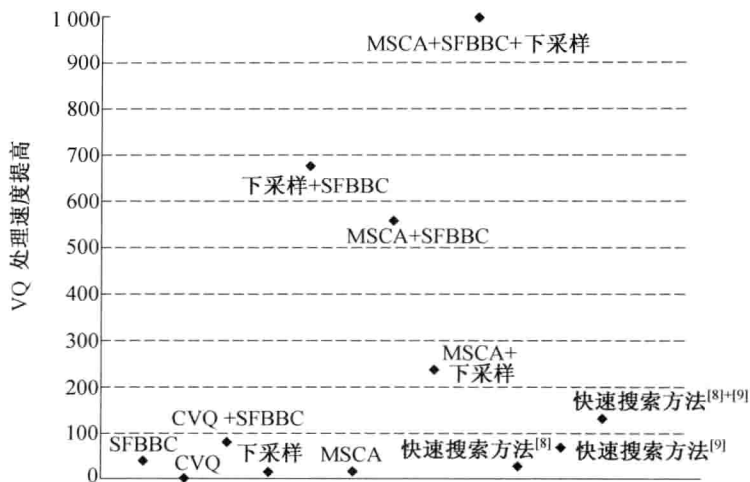


图 1.1 加拿大航天局快速算法及其组合的 3DVQ 处理速度的提高

使用下采样时,快速技术的保真度损失小于 1.0 dB。其他快速技术几乎没有保真度损失。

加拿大航天局为星载处理制定并取得了两种近无损 VQ 数据压缩技术专利:连续近似多级矢量量化 (Successive Approximation Multi-stage Vector Quantization, SAMVQ)^[10,11] 以及分等级的自组织聚类矢量量化 (Hierarchical Self-Organizing Cluster Vector Quantization, HSOCVQ)^[12,13]。这两种技术都是简单而有效的,已被明确指明用于多维传感数据星载使用。尽管这些算法为星载高光谱数据进行了优化,但它们也同样可用于地面数据的压缩。在接下来的两节里,我们将简要介绍这两种压缩技术以及使它们能够被称为近无损压缩的特点。

1.2 近无损压缩技术: SAMVQ 和 HSOCVQ

在数据压缩的发展中,我们将引入有损压缩过程中的压缩误差限制到原始数据集内部噪声的水平。这里的内部噪声指的是包含在一个原始数据集内的总体噪声或误差,它们是由仪器噪声以及数据集的其他误差源,例如在数据处理链中的误差或不确定性(如探测器暗电流的移除、非均匀性校准、辐射定标、大气校正等)产生的。与原始数据相比,希望压缩误差对数据集遥感应用的影响能小到可忽略的程度,这种无损压缩就称为近无损压缩。这不同于文

献[22,23]给出的医学图像的视觉近无损与文献[24,25]给出的多光谱/高光光谱图像的视觉近无损。

1.2.1 连续近似多级矢量量化(SAMVQ)

连续近似多级矢量量化(Successive Approximation Multi-stage Vector Quantization, SAMVQ)是一种多级VQ压缩算法,能够使用非常小的码书以一种连续近似的方式压缩数据立方体。随着码书的大小 N 比以前减小了超过两个数量级,传统VQ方法面临的计算负担不再是问题。假设SAMVQ使用四个码书以多级近似的过程压缩数据立方体,每个过程包含八个码矢量。等效的传统VQ码书需要 $N=8^4=4\,096$ 个码矢量来取得类似的重建保真度,而SAMVQ码书仅仅需要 $N'=8\times 4=32$ 个码矢量。由于码书的训练时间和编码时间都与码书的大小成正比,两者都得到了改善,其因子约为 $N/N'=4\,096/32=128$ 。由于码矢量总数很小,重建数据保真度相同时,SAMVQ的压缩比大于传统的VQ方法。与之等效,在和传统的VQ方法有相同压缩比的情况下,SAMVQ比传统的VQ方法具有更高的重建保真度。

此外,SAMVQ能够根据光谱特征的相似性,自适应地将待压缩的数据立方体分类/划分为聚类(子集),并且分别对每个子集进行压缩。这一特点进一步提高了星载应用的处理速度,因为它可以通过将每个子集分配到一个单独的处理单元,实现硬件处理中的并行运算。例如,如果一个数据立方体被分为8个子集,处理时间可以进一步提高大约8倍。这一特点也会提高重建的保真度,因为每一聚类的光谱矢量是相似的,当使用相同数目的码矢量时,可以以更小的编码失真进行编码。

压缩比和保真度可以通过合理地选择码书大小和级数容易地进行控制,级数越多,压缩的保真度就越高。在压缩的过程中,算法可以自适应地在每个逼近阶段选择码书的尺寸,以使得失真最小化,并使压缩比最大化。对于星载应用,SAMVQ可以用于压缩比(Compression Ratio, CR)模式或保真度模式。在压缩比模式中,可以在压缩之前设定参数得到期望的CR,然后不同的数据立方体的压缩保真度也变化。在保真度模式中,在压缩前设定一个RMSE阈值,然后算法会保证压缩误差小于等于设定的阈值。当设定的阈值与原始数据的内在噪声一致时,能够实现近无损压缩。SAMVQ算法的详细描述参见文献[10,11,14]。

1.2.2 分等级的自组织聚类矢量量化(HSOCVQ)

对于期望的保真度测度(例如RMSE),应用分等级的自组织聚类矢量量