专题: 遥感与全球变化研究

论 文

基于快速聚类方法的 30 m 分辨率中国土地覆盖遥感制图

胡雯运^①, YanLei CHEN^②, 徐玥^③, 赵圆圆^①, 俞乐^①, 王杰^④, 宫鹏^{①②④*}^① 清华大学地球系统科学研究中心, 地球系统数值模拟教育部重点实验室, 北京 100084;^② Department of Environmental Science, Policy and Management, University of California, Berkeley, CA 94720-3114, USA;^③ 北京师范大学全球变化与地球系统科学研究院, 北京 100875;^④ 中国科学院遥感与数字地球研究所, 遥感科学国家重点实验室, 北京 100101

* 联系人, E-mail: penggong@mail.tsinghua.edu.cn

收稿日期: 2013-07-22; 接受日期: 2013-12-12; 网络版发表日期: 2014-07-11

国家高技术研究发展计划前沿项目(编号: 2009AA12200101)和清华大学国际合作项目(编号: 2012Z02287)资助

摘要 大范围遥感土地覆盖制图通常费时费力. 本文采用一种基于特征空间变换的快速聚类方法(CBEST)对覆盖中国的 508 景 Landsat Thematic Mapper(TM)影像进行土地覆盖制图. 将中国划分为 50 个生态分区, 每个生态分区内采用 CBEST 方法对 TM 影像进行快速聚类, CBEST 平均聚类时间约为每 2 min 一景; 通过对自动获得的光谱簇人工土地覆盖类型识别, 归并光谱簇, 绘制 30 m 空间分辨率的中国土地覆盖图, 统计土地覆盖类型面积, 生成主要土地覆盖类型专题图. 采用 2159 个分布全国的验证样本对制图结果进行精度验证, 得到总体精度为 71.7%, Kappa 系数为 0.641. 与当前两套全球地表覆盖产品(FROM-GLC 和 GlobCover 2009)的局部对比表明 CBEST 分类结果在各类型土地覆盖面积统计及视觉效果上存在优势.

关键词

土地覆盖

遥感制图

CBEST

Landsat TM

地表覆盖是研究全球气候变化、能量循环、碳氮循环和水循环的重要参数(Turner 等, 1995; Sellers 等, 1997; Hibbard 等, 2010; Gong 等, 2013), 为生态、水文和气候模型提供数据支持; 土地覆盖作为经济、政治等决策分析的重要因素, 能帮助有效掌握生态系统状态和健康状况, 了解空间模式的多样性, 为土地管理政策提供依据(Homer 等, 2007); 土地覆盖变化与人类活动密切相关, 影响着生物栖息地以及物种的多样性(Brooks 等, 2002; Foley 等, 2005; Buchanan 等, 2008; 宫鹏, 2012), 改变着人类生活环境. 因此,

及时获得特定时间地表覆盖状况及其随时间的变化特征具有重要意义.

通过分析数据特征、合理处理数据和选择分类方法等途径, 中国已经完成了不同时期的全国土地利用和土地覆盖制图产品. 刘纪远等(1997)通过人工解译的方法完成了以 Landsat Thematic Mapper(TM)为主要信息源的中国土地覆盖/土地利用制图(Liu, 2002; Liu 等, 2005). 徐文婷等(2005)在欧盟联合研究中心 2000 年全球土地覆盖计划背景下, 利用 2000 年的 1 km 空间分辨率的 SPOT-4-VGT S10 数据与 DEM、

中文引用格式: 胡雯运, Chen Y L, 徐玥, 等. 2014. 基于快速聚类方法的 30 m 分辨率中国土地覆盖遥感制图. 中国科学: 地球科学, 44: 1621-1633

英文引用格式: Hu L Y, Chen Y L, Xu Y, et al. 2014. A 30 meter land cover mapping of China with an efficient clustering algorithm CBEST. Science China: Earth Sciences, 57: 2293-2304, doi: 10.1007/s11430-014-4917-1

积温和降水等数据通过层次分析法合成的自然因子数据, 采用非监督分类方法制作了中国 2000 年的土地覆盖图. 张炳智等(2002)结合中国 1:250 万林业资源分布图和 1:250 万土地利用现状图等数据, 采用最大似然法对 Landsat TM 和 Advanced Very High Resolution Radiometer(AVHRR)遥感图像进行分类, 完成了 1 km 中国土地利用、中国土地覆盖以及中国植被分类图. 潘耀忠等基于 1 km 分辨率的连续时间序列的 AVHRR-NDVI(Normalized Difference Vegetation Index)数据与气候综合指标可能蒸散, 采用 ISODATA 非监督分类算法对中国土地覆盖类型进行测量和制图(Pan 等, 2003). 张永红等(2009)基于 Moderate Resolution Imaging Spectrometer(MODIS)数据采用 ISODATA 非监督分类与决策树分类相结合的方法得到了中国 2003 年 1 km 分辨率的中国土地覆盖分类结果. 张增祥(2010)建立了 1:25 万比例尺土地覆盖数据库, 对不同的土地覆盖类型采用不同的方法提取, 完成了中国 2005 年 1 km 分辨率的土地覆盖图. 冉友华等采用多源数据决策融合的方法, 产生了国际地圈生物圈计划(IGBP)分类体系的 1 km 分辨率的中国 2000 年土地覆盖数据(Ran 等, 2012). 总体而言, 除了刘纪远等(1997, 2002, 2005)和张增祥(2010)采用人工

解译 30 m 分辨率的影像完成中国土地覆盖制图以外, 其他的土地覆盖产品均是基于 2005 年前的低空间分辨率的遥感图像. 随着人们对土地覆盖数据更广泛的需求与应用, 在中国或更大区域实现中等分辨率土地覆盖分类制图十分必要(Gong 等, 2013).

对遥感影像自动分类来获得土地覆盖图是最常用的方法. 大地理范围遥感制图更是如此. 相较于监督分类, 非监督分类(也称聚类)的制图方法具有人工干预少和高效聚类的优点. 但传统的非监督分类算法通常需要逐像元的多次迭代处理以获得较高精度, 计算效率较低, 无法满足越来越高分辨率以及越来越大尺度的遥感影像聚类处理(Gong 和 Howarth, 1992; Chen 和 Gong, 2013; 付薇等, 2013), 因此本文采用一种新的基于特征空间变换的快速聚类方法 Clustering based on Eigenspace Transformation(CBEST) (Chen 和 Gong, 2013)对覆盖中国区域的 508 景中等分辨率 Landsat TM 的遥感图像进行快速聚类试验.

1 研究方法

本研究主要由以下几个部分组成(图 1). 首先, 从 FROM-GLC (Finer Resolution Observation and

图 1 CBEST 土地覆盖分类流程图

Monitoring of Global Land Cover)项目中收集经大气校正和几何校正后的覆盖中国的 Landsat TM 影像(Gong 等, 2013); 然后基于生态分区思想, 采用 CBEST 聚类方法对 Landsat TM 影像进行聚类; 按照预先设计的分类系统, 参考 Google Earth 高分辨率影像确定每景影像中的光谱簇类别, 拼接分类图像, 完成 2010 年 30 m 分辨率的中国土地覆盖图; 并且根据全球 30 m 地表覆盖制图项目 FROM-GLC(Gong 等, 2013)的解译方法对验证样本进行更正解译, 验证分类精度; 最后分别获取 FROM-GLC 土地覆盖数据、GlobCover 2009 土地覆盖数据以及部分地区统计年鉴土地资源数据, 局部分析 CBEST 分类效果.

1.1 CBEST 聚类方法

1.1.1 CBEST 基本原理

CBEST 是一种快速的非监督分类方法, 由 Chen 等(2013)提出, 它集成 Eigen-Based Gray Level Vector Reduction(Gong 和 Howarth, 1992)和 K-means 算法, 缩短了传统算法对大量数据的处理时间. CBEST 算法主要分为压缩与聚类两个步骤(Chen 和 Gong, 2013).

CBEST 采用了 Eigen-Based Gray Level Vector Reduction 的方法对数据进行压缩, 其核心是将 $n \times p$ 的空间数据压缩成为用 N 个灰度级表示空间数据. 首先, Eigen-Based Gray Level Vector Reduction 方法基于主成分分析将原始的相关性很高的数据转换成彼此相互独立的数据, 并得到相应主成分的特征值和特征向量; 然后, 将主成分特征空间分成期望的 N 个灰度级, 每一个主成分的本征根(λ_i)决定了该主成分能够分成灰度级的数量(式(1)~(3)), 并且仅采用灰度级大于 2 的 q 个主成分($q < p$); 最后将数据压缩成新的灰度矢量, 计算特征向量的均值向量和权重用于聚类.

$$\frac{N_1}{\lambda_1} = \frac{N_2}{\lambda_2} = \dots = \frac{N_p}{\lambda_p}, \quad (1)$$

$$N_1 N_2 \dots N_p = N, \quad (2)$$

$$N_i = \lambda_i^p \sqrt{N / \lambda_1 \lambda_2 \dots \lambda_p}. \quad (3)$$

式中, λ_i , 第 i 个主成分对应的本征根; N_i , 第 i 个主成分对应的灰度级; N , 期望的特征空间能够分成多少个灰度级.

CBEST 聚类采用经过数据压缩得到的 q 维的均

值向量($m_1, m_2, \dots, m_N$)和对应的向量的权重($w_1, w_2, \dots, w_N$)代替特征向量($u_1, u_2, \dots, u_p$). 与 k -means 聚类算法类似, CBEST 的目的也是寻找 k 个聚类中心($\mu_1, \mu_2, \dots, \mu_k$)图像分成 V 个类别 $V = \{V_1, V_2, \dots, V_k\}$, 每一类别中的点都距离该部分的聚类中心距离最短.

$$J = \sum_{i=1}^k \sum_{v_j \in V_i} \|v_j - \mu_i\|^2. \quad (4)$$

与 k -means 每一次迭代遍历所有数据不同, CBEST 对数据只进行两次搜索. 并且对于每一次迭代, CBEST 将时间从 $O(knp)$ 次迭代减少到了 $O(kNq)$ 次迭代($q \leq p$). 因此, 对于大批量数据的处理, 压缩后的数据量为 N , 能够大量减少运算时间.

1.1.2 CBEST 聚类速度

随着影像大小和簇数量的增加, CBEST 加速性能更加明显(Chen 和 Gong, 2013). 本文选取了云贵高原地区和东北平原地区两景 TM 影像(图 2)比较 CBEST 与 k -means, ISODATA 的聚类速度, 采用 ENVI/IDL 完成 k -means 和 ISODATA 聚类, 所使用的个人计算机 CPU 主频速度为 2.90 GHz. 3 种聚类方法的参数设置如表 1 所示.

从表 2 的 3 种聚类方法所需时间对比可以发现, 在相同聚类类别个数和最大迭代次数下, CBEST 的聚类时间明显优于其他聚类方法. 除此之外, 从表 3 本研究所使用的 CBEST 聚类参数可以发现, 当最大迭代次数增加到 100, 聚类类别个数增加到 100 类时, 聚类时间为 1~2 min 左右, 聚类速度很快. 可见, CBEST 方法适用于大批量数据的快速聚类.

1.1.3 CBEST 聚类过程

空间异质性是大尺度制图中的常见现象, 对整个中国进行统一的聚类无法体现地表的多样性, 聚类结果易混淆; 但单独对单景影像进行聚类, 相邻相似影像不能很好融合, 容易出现空间不连贯的“马赛克”现象. 因此, 本研究采用了生态分区思想对影像进行了分析.

聚类过程主要包括 4 个步骤(图 3): (1) 获取中国科学院生态环境研究中心的 50 个生态分区(<http://www.ecosystem.csdb.cn/ecoass/ecoplanning.jsp>), 按照地理位置将中国 508 景影像划分到不同的生态区内, 当一景影像横跨多个生态区时, 采用所占面积最大原则划分影像生态区位置; (2) 统计每景影像各波段

图 2 聚类速度对比的研究区域及其 Landsat TM 影像

(a) L5130043_04320061118(8039×6951 pixel); (b) L5118028_02820070930(8511×7521 pixel)

表 1 CBEST, *k*-Means 和 ISODATA 聚类参数设置

聚类方法	聚类类别个数	最大迭代次数	其他参数
CBEST	10	4	空间分块细度: 1000000; 变化率(Change rate): 0
<i>k</i> -Means	10	4	变化阈值: 5%
ISODATA	10	4	变化阈值: 5%; 最大簇标准差: 2; 最小簇间距离: 5; 最大簇合并数量: 2

表 2 CBEST, *k*-Means 和 ISODATA 聚类时间对比

影像名称	CBEST 聚类时间	<i>k</i> -Means 聚类时间	ISODATA 聚类时间
L5130043_04320061118	62"	297"	277"
L5118028_02820070930	61"	313"	316"

表 3 本研究 CBEST 聚类参数设置及平均聚类速度

参数	设置
聚类类别个数	100
最大迭代次数	100
空间分块细度	1000000
变化率(Change rate)	0
间隔采样(# of skip in PCA)	100
单景平均聚类时间	约 2 min

的标准差, 并分别计算每个生态区中两两影像标准差之间的 Pearson 相关系数 R , 将 Pearson 相关系数 $R>0.9$ 的影像划分为一个分区, 这样一个生态区将划分为更加细致的多个分区. 当出现一景影像无法划入该生态区任何一分区时, 会考虑邻近生态区中的分区, 计算其与分区各影像的相关系数, 当相关系数全部大于 0.9 时, 划入该分区. 经过这样的处理, 每个分区内的影像具有较高的相似性; (3) 采用 CBEST 聚类方法对每景 Landsat TM 影像进行聚类, 聚类类

别设置为 100 簇; (4) 依据每景影像 CBEST 聚类的 100 簇的各波段均值, 采用 CBEST 软件对每个分区的所有影像进行 K-means 再聚类, 将每个分区的影像统一为 50 类簇.

1.2 分类系统

本文采用与 FROM-GLC 全球制图产品(Gong 等, 2013)一致的分类系统. 该系统依据地表的覆盖类型, 植被功能类型, 植被的覆盖度以及植被高度等 4 个特征来定义二级类, 最后将二级类归并为一级类. 该分类系统主要表征地表覆盖的类型, 根据地球系统模式的需求只有耕地属于土地利用类型的范畴; 目前很多研究团队都研究了云探测及去除方法, 但是其方法均需要大量候补数据用于修改所用分类数据(Zhu 和 Woodcock, 2012), 由于本研究只收集了单时相影像, 做云探测也不能用无云数据替补, 所以在本文中暂不能做逐块修补数据. 为减少云对于分类的干扰, 只能按所见即所得的方法把云作为一类分出

图 3 CBEST 聚类过程

来, 日后可以在此数据基础上进行修补, 逐块修改. 本研究的土地覆盖分类系统共包括 25 类二级类和 9 类一级类(表 4).

1.3 光谱簇类型识别

图像光谱簇识别的目的是将光谱簇赋予特定的地表覆盖类别. 目视辨别光谱簇类别的工作平台在 ENVI+IDL 5.0, 采用 432 波段的彩色合成, 参照 Google Earth 高清分辨率影像, 选取 FROM-GLC 制图过程中有经验的操作人员完成. 考虑 CBEST 整个

聚类方法的特殊性, 光谱簇识别主要包括以下几个部分: 首先对通过相关系数 R 确定的聚类分区的 50 簇进行判别, 依据前两景的 50 类簇综合确定聚类分区的 50 类簇类别, 填写表 5 地物类型代码列. 然后, 检验分区中的影像 3, 4, 5……对应的 50 簇类别是否与分区的簇类别相同, 如果不同, 在表 5 对应的影像名称中填写新类别, 同时判别影像的每一簇是否为纯类别, 非纯类别采用类别 1 代码+类别 2 代码+……(例如: 13+23+32)的形式填写在对应影像名称列中(表 5), 并返回原始的 100 类簇中, 识别不是纯类别簇

表 4 土地覆盖分类系统

一级类	二级类						
耕地	水稻田	大棚养殖	其他农地				
森林	阔叶林	针叶林	混交林	果园			
草地	牧草地	其他草地	沼泽湿地	苔原			
灌丛地							
水体	湖泊	水库/鱼塘	河流	海水			
人造覆盖	高反照率	低反照率					
裸地	盐碱地表	沙地	砾/岩石地	裸耕地	干河/湖床	裸土地	
积雪或冰	积雪或粒雪	冰盖或冰川					
云							

表 5 CBEST 图像 50 类光谱簇解译表

分区编号: 解译时间: 解译人:

CBEST_50 类编码	地物类型代码	是否为纯类别	备注	影像名称 1	影像名称 2	影像名称 3	...
--------------	--------	--------	----	--------	--------	--------	-----

对应的 100 簇编码的类别, 当 100 簇中有些单个光谱簇是代表了多个地物类别的混合时, 确定光谱簇类别的原则是将该光谱簇中所占比例最大的地物作为该光谱簇的类别. 根据该光谱簇识别方法, 确定了每个分区的 50 类簇的类别, 并且详细确定了每景影像的 100 簇类别.

2 数据

2.1 影像数据

Landsat TM 数据对于不同土地覆盖类型的物理特征有较强的区分能力(Jensen, 2004), 除热红外第 6 波段以外, 第 1~5 波段和第 7 波段数据均具有 30 m×30 m 的空间分辨率. 因此本研究选取了 Landsat TM5 数据的第 1 波段蓝光波段(0.45~0.52 μm), 第 2 波段绿光波段(0.52~0.6 μm), 第 3 波段红光波段(0.63~0.69 μm), 第 4 波段近红外波段(0.76~0.9 μm), 第 5 波段中红外波段(1.55~1.75 μm), 第 7 波段中红外波段(2.08~2.35 μm)用于分类. 本研究收集了覆盖整个中国的 508 景 Landsat TM 数据(<http://glovis.usgs.gov/>; 王雷等, 2012), 并进行了大气校正和几何校正, 以 2010 年

为基准年, 从相近年份寻找合适影像, 其中 88% 的影像来自于 2007 年以后, 采用 2006 年前的影像来替代 2007 年后影像质量不好的影像, 并且影像主要分布在夏季. 这些遥感影像的获取年份和季节如图 4 所示.

2.2 验证数据

验证样本点位置来自于 FROM-GLC 全球制图产品(Gong 等, 2013)验证精度时所使用的验证样本位置, 共包含了 2159 个验证样本点. 从 508 景影像中提取验证样本点, 选用与光谱簇类型识别不同的有经验的操作人员通过参考 Landsat TM 影像(432 波段组合)、光谱曲线、MODIS 时间序列曲线和 Google Earth 上的高分辨率影像, 修正全球制图产品的验证样本点类型, 最终建立本研究的中国验证样本数据库(图 5). 该批验证样本点曾用于多篇文章(Gong 等, 2013; Yu 等, 2013), 精度可靠.

2.3 辅助数据

为了进一步评价本文方法的制图结果, 两套全球地表覆盖制图数据用于进行制图效果的对比: (1) 30 m 分辨率 FROM-GLC 产品(Gong 等, 2013;

图 4 中国 Landsat TM 影像时间分布图

图 5 中国验证样本空间分布图

data.ess.tsinghua.edu.cn)和(2) 300 m 分辨率 GlobCover 2009 产品 (Bontemps 等, 2011; <http://ionial.esrin.esa.int>). 另外, 我们收集了 3 个地区的统计年鉴资料 (天津市(苏长伟, 2011; 杜西平和董顺荣, 2011)), 湖北省宜昌市(马学军, 2010), 青海省海南藏族自治州(董洁人, 2010)辅助比较三套制图结果中不同地表覆盖类型面积与统计资料的差异。

3 结果与分析

3.1 分类结果及精度验证

采用图 3 中提出的方法对 Landsat TM 影像聚类, 通过人工识别光谱簇, 得到了中国 2010 年 30 m 分辨率的土地覆盖图(图 6 和 7), 并统计了本结果中各种土地覆盖类型的面积(表 6)。从图中可以看出中国土地覆盖的空间分布格局: 耕地和建设用地主要分布在中国东部地区以及四川盆地(图 7(a)和(f)); 森林主要分布在以大兴安岭-吕梁山-青藏高原东南沿线以东地区, 西部天山、阿尔泰山和祁连山等地有零散分

布(图 7(b)); 草地处于森林和裸地的过渡地带, 从东北地区西部到青藏高原, 呈带状分布, 小部分分布在新疆阿尔泰山和天山(图 7(c)); 水体主要分布在青藏高原地区和东部地区(图 7(d)); 裸地占据了我国大半西北(图 7(e))。总体来说, 从东南到西北植被覆盖逐渐减少, 土地覆盖由耕地和林地逐步过渡到草地和裸地。

利用收集到的 2159 个验证样本点计算得到混淆矩阵(表 7), 总体精度是 71.7%, Kappa 系数为 0.641。从混淆矩阵中可以看出, 积雪或冰有最高的用户精度(UA)(97.0%)和第二高的制图精度(PA)(86.5%)。水体具有最高的 PA(93.3%), 说明分类结果中对于水体漏分程度很小, 但由于山体阴影、图像边界无信号部分等的光谱信息与水体相似, 分类水体将会增加, 因此多分误差偏大(水体 UA 为 73.7%); 与水体相似, 耕地的 PA(77.5%)高于 UA(64.6%), 主要由于耕地与森林, 草地的光谱相似, 导致耕地面积偏高; 混淆矩阵中, 人造覆盖主要与耕地混淆, 耕地与城市的相辅相成的分布, 30 m 分辨率下, 城市与耕地的界限不易划

图 6 中国 30 m 分辨率土地覆盖图

表 6 CBEST 分类结果各类型面积及所占比例

类型	面积(km ²)	百分比(%)
耕地	2.06×10 ⁶	21.67
森林	2.28×10 ⁶	24.04
草地	2.59×10 ⁶	27.27
灌丛地	1.21×10 ⁵	1.28
水体	1.85×10 ⁵	1.95
人造覆盖	1.51×10 ⁵	1.59
裸地	1.91×10 ⁶	20.17
冰雪	1.34×10 ⁵	1.41
云	6.08×10 ⁴	0.64
总计	9.50×10 ⁶	100.00

定; 由于西部草原地带, 灌丛分布稀疏导致植被信息不突出, 易与裸地和草地混淆, 而东部地区, 灌丛与林地光谱信息相近, 导致土地覆盖图中灌丛精度偏低; 森林、草地和裸地用户精度分别为 84.3%, 61.7% 和 79.4%, 制图精度分别为 77.0%, 68.5% 和 74.2%,

三类土地类型的用户精度和制图精度相差较小, 识别出的森林、草地和裸地总面积越接近于真实值(朱秀芳等, 2007)。

3.2 局部地区分析

为了进一步分析 CBEST 分类结果的分类效果, 本文分别在中国东部, 西部, 中部地区选取具有代表性的地区, 对比 300 m 分辨率的 GlobCover 2009 土地覆盖产品、30 m 分辨率的 FROM-GLC 土地覆盖产品与 CBEST 分类结果图(土地覆盖分类系统对应表详见网络版附表 1(<http://earth.scichina.com>)). 图 8 是三种土地覆盖数据分别在天津市(A 组)、湖北宜昌市(B 组)、青海海南藏族自治州(C 组)的土地覆盖图。从图 8 中可以看出, 三种覆盖产品均能宏观上反映该地区的土地覆盖类型的分布特征, 但表达的局部细节特征各有不同。

图 7 中国主要土地覆盖类型专题图

表 7 CBEST 分类结果混淆矩阵^{a)}

	C1	C2	C3	C4	C5	C6	C7	C8	C9	总量	UA(%)
C1	337	94	59	9	0	20	3	0	0	522	64.6
C2	44	434	28	6	1	0	1	1	0	515	84.3
C3	29	22	330	39	0	1	110	2	2	535	61.7
C4	5	10	3	8	0	1	3	0	0	30	26.7
C5	3	3	1	0	28	0	3	0	0	38	73.7
C6	10	1	0	0	1	17	1	0	0	30	56.7
C7	4	0	61	20	0	1	351	2	3	442	79.4
C8	0	0	0	0	0	0	1	32	0	33	97.0
C9	3	0	0	0	0	0	0	0	11	14	78.6
总量	435	564	482	82	30	40	473	37	16	2159	
PA(%)	77.5	77.0	68.5	9.8	93.3	42.5	74.2	86.5	68.8		71.7

a) C1, 耕地; C2, 森林; C3, 草地; C4, 灌丛地; C5, 水体; C6, 人造覆盖; C7, 裸地; C8, 积雪或冰; C9, 云。OA=71.7%, Kappa 系数=0.641。单位: 像元

图 8 CBEST 土地覆盖图与 GlobCover 2009 土地覆盖图、FROM-GLC 土地覆盖图的对比

天津市位于华北平原东部, 三种土地覆盖数据均显示了天津市山地面积小, 主要位于北面燕山南部, 绝大部分为平原, 耕地面积大, 城市发展程度高(图 8(A)). 但从局部放大图可以看出, FROM-GLC 和 CBEST 分类结果对水体, 农田, 城市的分辨能力均高于 GlobCover 2009 的分类结果; 对比两种 30 m 分辨率土地覆盖数据, CBEST 能够细致地区分农田、农村居民点和交通道路, 并且形象地表达了土地破碎化严重, 人造覆盖的整体轮廓更加清晰. 参考天津土

地资源数据, CBEST 分类结果中林地、人造覆盖、裸地的面积比例均最接近统计年鉴中的土地资源数据(表 8).

图 8(B)是位于云贵高原和渝东大巴山向汉江平原过渡地带的湖北宜昌市, 从西北向东南分别为山地、丘陵和平原. 森林覆盖率达 63.4%, CBEST 和 FROM-GLC 的土地覆盖数据均显示了该地区森林覆盖面积广, 而 GlobCover 2009 将部分森林分成了农地或草地. 放大图清晰地显示 FROM-GLC 将大量的山

表 8 三种土地覆盖数据与统计年鉴土地资源数据面积对比^{a)}

地区	土地类型	统计年鉴			CBEST			FROM-GLC			GlobCover 2009		
		面积 (km ²)	所占 比例(%)		面积(km ²)	所占 比例(%)	绝对 误差	面积 (km ²)	所占 比例(%)	绝对 误差	面积 (km ²)	所占 比例(%)	绝对 误差
天津	农地	6591	55.3		7294	62.9	7.6	8383	72.3	16.9	7160	50.6	-4.7
	林地	562	4.7		706	6.1	1.4	743	6.4	1.7	1458	10.3	5.6
	草地				29	0.3		92	0.8		3120	22.1	0.0
	水体	1358	11.4		1420	12.3	0.9	1380	11.9	0.5	1148	8.1	-3.3
	人造覆盖	2512	21.1		2057	17.8	-3.3	988	8.5	-12.6	1245	8.8	-12.3
	裸地	895	7.5		78	0.7	-6.8	13	0.1	-7.4	15	0.1	-7.4
	云				2	0.0		2	0.0				
	总量	11918	100.0		11587	100.0		11602	100.0		14146	100.0	
宜昌	农地	5487	25.8		4646	21.8	-4.0	3947	18.5	-7.3	8774	37.5	11.7
	森林	13450	63.4		15513	72.7	9.3	13205	62.0	-1.3	10968	46.9	-16.4
	草地	122	0.6		1	0.0	-0.6	914	4.3	3.7	3146	13.5	12.9
	水体	1094	5.2		639	3.0	-2.2	2512	11.8	6.6	305	1.3	-3.8
	人造覆盖	987	4.6		325	1.5	-3.1	118	0.6	-4.1	182	0.8	-3.9
	裸地	90	0.4		90	0.4	0.0	18	0.1	-0.3	0	0.0	-0.4
	云				134	0.6		570	2.7				
	总量	21230	100.0		21348	100.0		21285	100.0		23376	100.0	
海南 藏族 自治 州	农地	976	2.2		400	0.9	-1.3	362	0.8	-1.4	13923	26.3	24.1
	草地	35045	78.7		34014	75.0	-3.7	15258	33.7	-45.0	28387	53.6	-25.0
	森林	1843	4.1		640	1.4	-2.7	2571	5.7	1.5	302	0.6	-3.6
	水体	2982	6.7		4801	10.6	3.9	5143	11.3	4.6	5495	10.4	3.7
	人造覆盖	236	0.5		70	0.2	-0.3	11	0.0	-0.5	28	0.1	-0.5
	裸地	3465	7.8		5159	11.4	3.6	20535	45.3	37.5	4712	8.9	1.1
	冰雪				251	0.6		1310	2.9		70	0.1	
	云				15	0.0		154	0.3				
	总量	44546	100.0		45350	100		45342	100.0		52918	100.0	

a) 将统计年鉴的土地资源数据和 GlobCover 2009 的土地覆盖分类系统转化我们的分类系统, 统计各土地类型面积, 各类型的面积=各类型斑块数×分辨率的平方, 由于分辨率不同, 所以总面积存在一定的差异; 所占比例指各类型面积占总面积的比例; 绝对误差=土地覆盖产品中各类型所占比例-统计年鉴中对应类型所占比例; 天津统计数据为 2010 年, 宜昌统计数据为 2009 年, 海南统计数据为 2008 年; 天津市的土地资源数据综合了天津统计年鉴和天津各区县年鉴, 缺少了草地分布面积; 海南州的土地资源数据来自于海南年鉴, 由于海南州不同的统计年鉴对于青海湖面积的划定不一, 因此我们计算海南州面积时包含了全部青海湖面积, 故总体面积偏大, 水体面积所占比例大; 统计年鉴中灌木均划定了为了森林; 三个地区的 CBEST 分类结果基本采用 2009 年附近的影像, 组成影像具体信息详见网络版附表 2(<http://earth.scichina.com>)

体阴影分为水体, 而 GlobCover 2009 的土地覆盖图却损失了一部分的河流水体信息, 相对而言, CBEST 聚类结果能够很好地提取河流的形状, 保证水体信息的完整. 表 8 的面积数据对比证实了该分析结果, 除森林以外, CBEST 的分类类型面积的相对误差均为最小.

相对华北平原地区、中部丘陵地区, 西部荒漠草原地区(图 8(C))的 30 m 高分辨率图像的显示效果并没有比 300 m 分辨率的视觉效果优越, 图 8(C)中可以看出, FROM-GLC 将大量的草地划分为裸地, 拼接痕迹明显; CBEST 与 GlobCover 2009 的裸地分布情况极其相似, 但 GlobCover 2009 草原地带出现了部分农地(农地面积 26.3%, 远远高于统计年鉴的 2.2%). 土地覆盖面积数据(表 8)显示, CBEST 的农地、草地、人造覆盖的面积精度最高, 对比土地覆盖图与面积数据可以看出, CBEST 在西部地区对草地识别能力强.

通过三种土地覆盖数据的视觉效果和面积统计对比可以看出: (1) 30 m 分辨率的图像表达信息能力明显优于 300 m 分辨率的图像; (2) 相比于 FROM-GLC, CBEST 更能细致地刻画土地覆盖的空间格局, 反映局部特征, 特别是东部耕地、道路、聚落的区分效果, 以及水体与其他地表的区分; (3) 根据 Pearson 相关系数在同一个生态区内划分了聚类分区, 对分区内的影像进行了再聚类, 统一了相似相邻影像的类别, 相对 FROM-GLC, 拼接痕迹明显减少, 提高了分类精度.

4 讨论与展望

本研究采用 CBEST 非监督分类完成了中国土地覆盖初步制图, 总体精度达到了 71.7%, 高于 30 m 分辨率的中国土地覆盖产品(FROM-GLC 中国区最高精度为 66.23%(Gong 等, 2013)), 并绘制了主要覆盖类型专题图, 其整体分布与面积基本符合中国现状. 该研究表明了 CBEST 分类方法的在大区域范围遥感数据分类的可行性; 对于单景 Landsat TM 影像的 CBEST 聚类 100 簇时间是 2 min 左右, 突显出 CBEST 方法在分类速度上的优势; 局部对比 FROM-GLC 和 GlobCover2009 两种土地覆盖产品的各类型面积比例及视觉效果, 反映了 CBEST 对 30 m 分辨率的图像分类效果的优越性, 能更细致地反映地表覆盖的空间格局.

基于 CBEST 完成的中国地表覆盖图反映了中国土地覆盖类型的空间分布特征, 但几个问题仍需改进: (1) 由于影像时间不同导致的“马赛克”效果; (2) 由于光谱相似性导致的干草与裸地易混淆、水体与阴影易混淆、草地与农地以及灌丛易混淆、耕地与城市易混淆等问题. 未来我们可以添加地形数据、时间序列数据以及纹理数据等进一步提高分类精度(例如: Yu 等, 2013). 另外, 利用 CBEST 的聚类结果, 可以服务于监督分类: 校正和扩充监督分类所使用的训练样本, 并最终提高分类精度及成图效果.

致谢 感谢 Arthur P Cracknell, Pauline Lovell 和李雪草对文章提出的宝贵意见; 感谢牛振国、李丛丛、王雷提供下载的 Landsat TM 数据.

参考文献

- 董杰人. 2010. 海南州年鉴(2009 版). 海南州: 海南州地方志编纂委员会. 38
- 杜西平, 董顺荣. 2011. 天津统计年鉴(2011 版). 北京: 中国统计出版社. 62
- 付薇, 陈焱磊, 施楣梧, 等. 2013. 卫星动态观测数据揭示中国地表色调分布及时序变化特征, 遥感学报, 17, doi: 10.11834/jrs.20132323
- 宫鹏. 2012. 拓展与深化中国全境的环境变化遥感应. 科学通报, 57: 1379-1387
- 刘纪远. 1997. 国家资源环境遥感宏观调查与动态监测研究. 遥感学报, 1: 225-230
- 马学军. 2010. 宜昌年鉴(2010 版). 武汉: 长江出版社. 76
- 苏长伟. 2011. 天津区县年鉴(2011 版). 天津: 天津社会科学院出版社. 64
- 王雷, 李丛丛, 应清, 等. 2012. 中国 1990~2010 年城市扩张卫星遥感制图. 科学通报, 57: 1388-1399
- 徐文婷, 吴炳方, 颜长珍, 等. 2005. 用 SPOT-VGT 数据制作中国 2000 年度土地覆盖数据. 遥感学报, 9: 204-214
- 张炳智, 张丽, 张继贤. 2002. 利用遥感技术进行全球测图项目中的中国土地利用覆盖与植被分类图的制作. 测绘科学, 27: 27-42
- 张永红, 张继贤, 郭健, 等. 2009. 全球地图第一版中国土地覆盖产品的生成. 武汉大学学报(信息科学版), 34: 914-918
- 张增祥. 2010. 中国土地覆盖遥感监测. 北京: 星球地图出版社

- 朱秀芳, 潘耀忠, 张锦水, 等. 2007. 训练样本对 TM 尺度小麦种植面积测量精度影响研究(I). 遥感学报, 11: 826–837
- Bontemps S, Defourny P, Van Bogaert E, et al. 2011. Globcover 2009 products description and validation report. E. UCLouvain. European Spatial Agency-UniversitéCatholique de Louvain
- Brooks T M, Mittermeier R A, Mittermerier C G, et al. 2002. Habitat loss and extinction in the hotspots of biodiversity. *Conserv Biol*, 16: 909–923
- Buchanan G M, Nelson A, Mayaux P, et al. 2008. Delivering a global, terrestrial, biodiversity observation system through Remote Sensing. *Conserv Biol*, 23: 499–502
- Chen Y L, Gong P. 2013. Clustering based on eigenspace transformation-CBEST for efficient classification. *Isprs-J Photogramm Remote Sens*, 83: 64–80
- Foley J A, Defries R, Asner G P, et al. 2005. Global consequences of land use. *Science*, 309: 570–574
- Gong P, Howarth P J. 1992. Frequency-based contextual classification and gray-level vector reduction for land-use identification. *Photogramm Eng Remote Sens*, 58: 423–437
- Gong P, Wang J, Yu L, et al. 2013. Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data. *Int J Remote Sens*, 34: 2607–2654
- Hibbard K, Janetos A, van Vuuren D P, et al. 2010. Research priorities in land use and land-cover change for the earth system and integrated assessment modelling. *Int J Climatol*, 30: 2118–2128
- Homer C, Dewitz J, Fry J, et al. 2007. Completion of the 2001 national land cover database for the conterminous United States. *Photogramm Eng Remote Sens*, 73: 337–341
- Jensen J R. 2004. *Introductory Digital Image Processing: A Remote Sensing Perspective*. New Jersey: Prentice Hall. 52
- Liu J Y. 2002. The land use and land cover change database and its relative studies in China. *J Geogr Sci*, 12: 275–282
- Liu J Y, Liu M L, Tian H Q, et al. 2005. Spatial and temporal patterns of China's cropland during 1990–2000: An analysis based on Landsat TM data. *Remote Sens Environ*, 98: 442–456
- Pan Y, Li X, Gong P, et al. 2003. An integrative classification vegetation in China based on NOAA AVHRR and vegetation climate indices of the Holdridge life zone. *Int J Remote Sens*, 24: 1009–1027
- Ran Y H, Lia X, Lua L, et al. 2012. Large-scale land cover mapping with the integration of multi-source information based on the Dempster-Shafer theory. *Int J Geogr Inf Sci*, 26: 169–191
- Sellers P J, Dickinson R E, Randall D A, et al. 1997. Modeling the exchanges of energy, water, and carbon between continents and the atmosphere. *Science*, 275: 502–509
- Turner B L, Skole D L, Sanderson S, et al. 1995. Land-use and land-cover change. *Science/Research Plan. Global Change Report*
- Yu L, Wang J, Gong P. 2013. Improving 30 meter global land cover map FROM-GLC with time series MODIS and auxiliary datasets: A segmentation based approach. *Int J Remote Sens*, 34: 5851–5867
- Zhu Z, Woodcock C E. 2012. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sens Environ*, 118: 83–94

附表

附表 1 GlobCover2009 分类系统与本研究分类系统对应表

本研究 分类系统	GlobCover 分类系统	
	编码	类别
耕地	11	Post-flooding or irrigated croplands (or aquatic)
耕地	14	Rainfed croplands
耕地	20	Mosaic cropland (50%~70%)/vegetation (grassland/shrubland/forest) (20%~50%)
耕地	30	Mosaic vegetation (grassland/shrubland/forest)(50%~70%)/cropland (20%~50%)
森林	40	Closed to open (>15%) broadleaved evergreen or semi-deciduous forest (>5 m)
森林	50	Closed (>40%) broadleaved deciduous forest (>5 m)
森林	60	Open (15%~40%) broadleaved deciduous forest/woodland (>5 m)
森林	70	Closed (>40%) needleleaved evergreen forest (>5 m)
森林	90	Open (15%~40%) needleleaved deciduous or evergreen forest (>5 m)
森林	100	Closed to open (>15%) mixed broadleaved and needleleaved forest (>5 m)
森林	110	Mosaic forest or shrubland (50%~70%)/grassland (20%~50%)
草地	120	Mosaic grassland (50%~70%)/forest or shrubland (20%~50%)
灌丛地	130	Closed to open (>15%) (broadleaved or needleleaved, evergreen or deciduous) shrubland (<5 m)
草地	140	Closed to open (>15%) herbaceous vegetation (grassland, savannas or lichens/mosses)
草地	150	Sparse (<15%) vegetation
森林	160	Closed to open (>15%) broadleaved forest regularly flooded (semi-permanently or temporarily)-Fresh or brackish water
森林	170	Closed (>40%) broadleaved forest or shrubland permanently flooded-Saline or brackish water
草地	180	Closed to open (>15%) grassland or woody vegetation on regularly flooded or waterlogged soil-Fresh, brackish or saline water
人造覆盖	190	Artificial surfaces and associated areas (Urban areas>50%)
裸地	200	Bare areas
水体	210	Water bodies
积雪或冰	220	Permanent

附表 2 CBEST 分类结果中天津市、宜昌市和海南藏族自治州所使用影像信息列表

地区	轨道号	行号	年	月	日
天津市	122	23	2009	8	30
	122	33	2009	8	30
	123	32	2010	6	5
	123	33	2009	9	22
宜昌市	124	38	2008	6	6
	124	39	2006	8	20
	125	38	2007	8	14
	125	39	2007	8	14
海南藏族自治州	132	35	2009	9	21
	132	36	2010	8	7
	133	34	2009	9	28
	133	35	2009	9	28
	133	36	2010	8	14