

全国地表类型遥感影像样本数据集

赵理君^{1*}, 郑柯¹, 史路路^{1,2}, 白洋^{1,2}, 唐吉文^{1,2},

张伟^{1,2}, 饶梦彬^{1,2}, 邹松^{1,2}, 李艳艳^{1,2}

1. 中国科学院遥感与数字地球研究所, 北京 100101

2. 中国科学院大学, 北京 100049

摘要: 全国地表类型遥感影像样本数据集是对全国(除港澳台地区外)31个省、自治区、直辖市的十米级Landsat 8多光谱影像数据和米级GF-1、QuickBird多光谱影像数据进行人工解译采集而得。本数据集使用了2013–2017年的影像数据,覆盖夏季和冬季两个典型时相,采用了面向测绘地物波谱的分类体系。所构建数据集可为土地覆盖分类提供先验训练和测试样本,指导待分类影像中同类影像样本点的采集,为应用示范研究提供样本数据支持。为了更好地利用这些数据,采用统一规范的数据处理方法、样本采集规则和质量控制体系,对地表类型遥感影像样本数据进行采集和质量控制,并已经公开共享和提供下载。

关键词: 地表类型; 全国; 遥感影像样本; 十米级多光谱; 米级多光谱

数据库(集)基本信息简介

数据库(集)名称	全国地表类型遥感影像样本数据集
数据作者	赵理君、郑柯、史路路、白洋、唐吉文、张伟、饶梦彬、邹松、李艳艳
数据通信作者	赵理君(zhaolj01@radi.ac.cn)
数据时间范围	2013–2017年
地理区域	全国(除港澳台地区外)31个省、自治区、直辖市等区域
空间分辨率	2.4–30 m
数据量	647 MB (解压缩后)
数据格式	*.tif, *.jpg, *.txt, *.xml
数据服务系统网址	http://www.sciencedb.cn/dataSet/handle/663
基金项目	科技基础性工作专项(2014FY210800)
数据库(集)组成	本数据集由2个分卷压缩文件构成,解压后包含了米级空间分辨率样本点的GF1和QuickBird这两个文件夹,以及十米级空间分辨率样本点的Landsat文件夹,各个文件夹下又由以不同采样地区命名的子文件夹组成。每个子文件夹下,存储了覆盖土壤、水体、岩矿、植被、冰雪冻土、人工目标6大类地表类型的样本点压缩文件,每个压缩文件中仅压缩存储同一景影像中同一种地表类型的多个样本点数据。解压缩后,每个样本点数据又由4个数据文件组成,分别为:样本点原始影像数据文件(*.tif),样本点预览图文件(*.jpg),样本点的影像波段DN值文件(*.txt),样本点的元数据文件(*.xml)。

文献 DOI:

10.11922/csdata.2018.0058.zh

数据 DOI:

10.11922/sciencedb.663

文献分类: 地球科学

收稿日期: 2018-09-17

开放同评: 2018-10-26

录用日期: 2019-03-26

发表日期: 2019-06-05

* 论文通信作者

赵理君: zhaolj01@radi.ac.cn

引言

地表覆盖是地球表面各种物质类型及其自然属性与特征的综合体，其空间分布直接影响着地球表面的物质和能量的循环过程^[1]。利用遥感影像开展地表覆盖监测是生态环境变化研究、土地资源管理和可持续发展的重要基础，在全球资源监测、全球变化检测中发挥着重要作用^[2]。目前，遥感影像的计算机分类成为主要的发展方向，涌现出多种分类方法，如统计模式识别方法、人工智能分类法、遥感与 GIS 结合法、面向对象的分类法、多源信息复合分类法等^[3]。遥感影像分类就是把图像中的每一个像元或区域划分为若干类别中的一种，即通过对各类地物的光谱特征分析来选择特征参数，将特征空间划分为互不重叠的子空间，然后将影像内各个像元划分到各子空间中去，从而实现分类^[4]。在遥感影像地表覆盖分类的过程中，影像样本的采集可支持监督分类模型的构建，为地表类型的遥感影像分类提供类别先验知识。目前，国家基础地理信息中心发布了 2010 年 GlobalLand 30 全球 30 米地表覆盖分类产品 (<http://www.webmap.cn/commres.do?method=globeIndex>)，可免费下载使用；中国科学院遥感与数字地球研究所发布了连续多年的中国土地覆盖遥感监测产品；清华大学发布了 FROM-GLC2010 地表覆盖分类产品 (http://data.ess.tsinghua.edu.cn/landsat_pathList_fromglc_0_1.html) 以及全球范围的验证样本点数据 (http://data.ess.tsinghua.edu.cn/data/temp/GlobalLandCoverValidationSampleSet_v1.xlsx)，但训练样本并未完全公开。

本文构建的数据集是依托科技基础性工作专项“测绘地物波谱本底数据库建设”(2014FY210800)完成，与项目中的典型水体反射率数据集、典型地物反射率数据集、典型地表类型全波段波谱数据集、作物生长季时间序列反射率数据集、森林和作物多尺度反射率波谱数据集、多频率多角度森林微波辐射亮温数据集、全国岩矿标本反射率数据集等其他数据集共同构成了项目最终的测绘地物波谱本底数据库，所有数据将通过平台网站(<http://210.72.27.78/spectrum/>)向公众发布。需要指出的是，项目中其他数据集数据均来自地面实测波谱，而本数据集样本数据则是从遥感影像的角度采集影像样本，以期为研究影像与地面波谱之间的关联关系提供支撑，同时也期望能够为遥感影像的分类提供补充样本。本数据集利用了全国范围内与项目其他数据集试验区域相近的遥感影像，通过人工判读与比对得到最终的影像样本。相比已有的影像样本数据集(如 GlobeLand 30, FROM-GLC 数据集^[5-6])，本数据集特色在于：(1) 时相更新(2013–2017 年)；(2) 包含影像样本空间分辨率更高(最高 2.4–8 m)；(3) 样本类型更精细(面向测绘地物类型)；(4) 服务方式完全免费(所有样本数据在平台完全公开)。

1 数据采集和处理方法

本数据集包括十米级和米级两种影像的样本数据，其中十米级影像以 Landsat 8 卫星的 30 m 空间分辨率数据为主，米级影像以国产 GF-1 卫星的 8 m 空间分辨率数据为主，补充了部分 QuickBird 卫星的 2.4 m 空间分辨率数据。所有的影像数据通过地理空间数据云平台 (<http://www.gscloud.cn/>)、购买等渠道获取下载。其中，Landsat 8 数据使用 L1T 产品，据 USGS 提供的信息，该产品已使用地面控制点和数字高程模型数据进行精确校正，几何校正精度为 12 m，小于 0.5 个像元；GF-1 数据使用 L1 级产品，无几何信息，利用 RPC 参数文件进行了 RPC 几何精校正；QuickBird 数据使用 L2A 级产品，已经过了几何精校正处理。上述所有数据源影像均经过了了几何精校正处理，不做辐射和大气校正处理，保留原始像元值(Digital Number, DN)数据，但需要注意的是样本点的 DN 值一般不

能直接应用于分类模型的构建。这里更多地考虑几何信息主要源于在实际应用中，不同传感器会因波段设置上的差异造成像元波谱信息无法保持一致性，在遥感影像分类过程中，可以很容易地根据地理点位和类别信息构建训练样本。通过对待分类影像进行辐射和大气校正便可以得到训练样本点的光谱反射率特征。在影像空间范围选择上，十米级影像覆盖全国（除港澳台地区外）31个省/自治区/直辖市的区域，每个行政区划单元随机选择至少2景影像；米级影像覆盖全国的7大分区，即东北、华北、华东、华中、华南、西南、西北地区，每个地区随机选择至少2景影像。在影像时相选择上，主要考虑夏季和冬季两个差异度最大的时相，其中冬季时相的时间跨度设定为12-3月，夏季时相的时间跨度设定为7-9月。在影像的获取年份上，设定为最近5年时间内的影像，以保证样本数据的时效性。最终，获得待采样的影像数据情况如表1所示。

表1 不同卫星影像数据获取情况表

卫星影像数据源	分辨率等级	数量(景)	空间范围划分
Landsat 8	十米级	66	全国(除港澳台地区外)31个省/自治区/直辖市
GF-1	米级	14	东北、华北、华东、华中、华南、西南、西北等地区
QuickBird 02	米级	1	华北地区

基于上述预处理后的遥感影像，通过人工解译开展地表类型影像样本采集工作。类别体系严格遵从项目内部制定的标准规范《测绘地物波谱库地物分类编码规范(草案)》，该草案主要参考了已有的国家相关标准和公认度较高的分类原则、标准，根据测绘部门和遥感数据分类研究的实际需要制定，覆盖地表类型种类包含植被、土壤、岩矿、冰雪冻土、水体、人工目标6大类一级类，植被向下划分至6级分类体系，冰雪冻土向下划分至5级分类体系，土壤、岩矿、水体、人工目标向下划分至4级分类体系。根据遥感影像对地表区分的能力，对上述分类体系进行了精简，表2给出本数据集构建中使用的类别体系划分情况。

表2 类别体系简表

1级类		2级类		3级类		4级类	
编号	类别	编号	类别	编号	类别	编号	类别
1	植被	11	农林用地	1101	耕地	110101	稻田
						110102	旱地
						110103	菜地
						110104	水生作物
		1103	林地				
		1104	天然草地	110402	草地		
		12	城市绿地	1201	人工绿地		
2	土壤	206	漠土				
3	岩矿	31	岩石				
4	冰雪冻土	41	冰	412	湖冰		
				413	河冰		
		42	雪				

1 级类		2 级类		3 级类		4 级类		
编号	类别	编号	类别	编号	类别	编号	类别	
		43	冻土					
5	水体	51	河流	511	常年河			
				513	干枯河（干河床）			
		52	沟渠	522	干渠			
		53	湖泊	531	常年湖、塘	53101	湖泊	
						53102	池塘	
		533	干枯湖					
		54	水库					
55	海洋要素	551	海域					
		552	海岸线					
		557	海岛					
6	人工目标	61	水系	6101	沟渠			
				62	居民地及设施	6201	居民地	
		6202	工矿及其设施					
		63	交通	6302	城际公路			
				6303	城市道路			

在人工解译采样过程中参考了已有的专题产品和历史资料以确保样本类别标注的准确性。这里的专题产品和历史资料包括植被、水体等单一类别的分类专题产品，Google Earth 高分辨率历史影像，项目组其他课题采集的地面样本实测数据等。针对不同分辨率的影像样本数据（十米级和米级），每景影像中的每一个子类别，采样数量控制在 50–700 个，采样大小为 7 像元×7 像元的区域，并在采样过程中避免边缘像元的选择。按照项目内部数据入库规范与要求，每一个样本数据采用标准统一的数据组织和存储格式进行记录，每一个样本数据对应 4 个文件，如表 3 所示。

表 3 样本记录文件

文件明细	文件内容
image_<top class>_<sensor type>_<acquired time>_Axxx.tif	样本点原始影像数据
view_<top class>_<sensor type>_<acquired time>_Axxx.jpg	样本点预览图
pixel_<top class>_<sub class>_<acquired time>_Axxx.txt	影像各波段 DN 值
pixel_<top class>_<sub class>_<acquired time>_Bxxx.xml	样本点的元数据描述信息

在表 3 中，<top class>、<sub class>、<sensor type>和<acquired time>有统一的命名规则。其中，<top class>为地物大类，限定为 vege, soil, rock, snow, water, manmade, 分别对应表 2 中的植被、土壤、岩矿、冰雪冻土、水体和人工目标等类别；<sub class>为子类英文简写名称，长度不超过 15 个字符；<sensor type>为传感器的英文名，限定为 OLI、PMS、QuickBird, 分别对应 Landsat 8、GF-1 和 QuickBird 02 卫星；<acquired time>为数据获取时间，记录年月日时分秒，格式为 YYYYMMDDHHMMSS；XXX 为文件编号范围从 001–999；A 和 B 为文件标识，A 是对样本点配套

文件的标识，B 是对元数据文件的标识。

2 数据样本描述

全国地表类型遥感影像样本数据集包含十米级和米级两个空间分辨率等级，其中十米级影像样本采集了 118 324 个样本（以省/自治区/直辖市为单元采样，每个地区至少两个时相），其中夏季 58 317 个，冬季 60 007 个；米级影像样本采集了 29 551 个（以全国大的区划为单元采样，每个区划至少两个时相），其中夏季 15 792 个，冬季 13 759 个。采样点的空间分布如图 1 所示。图 2 和图 3 分别给出了数据集中十米级影像样本和米级影像样本的组成情况。



审图号：GS（2019）1381 号

图 1 全国地表类型遥感影像样本数据集空间分布图

本数据集以“图像+描述文档”的方式组织存储，包含了影像数据和文本数据，如表 3 所示。图 4 以农林用地这一类别的一个样本点数据为例，展示了其对应的存储文件列表。

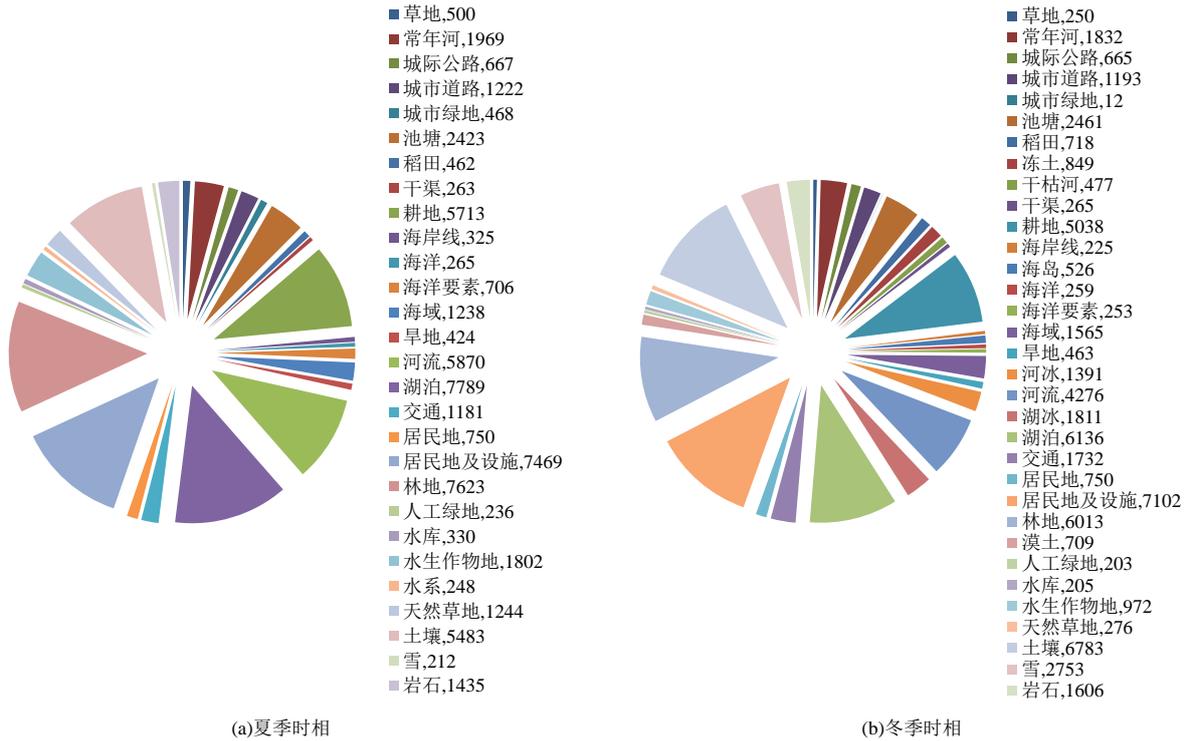


图2 十米级影像样本数量统计

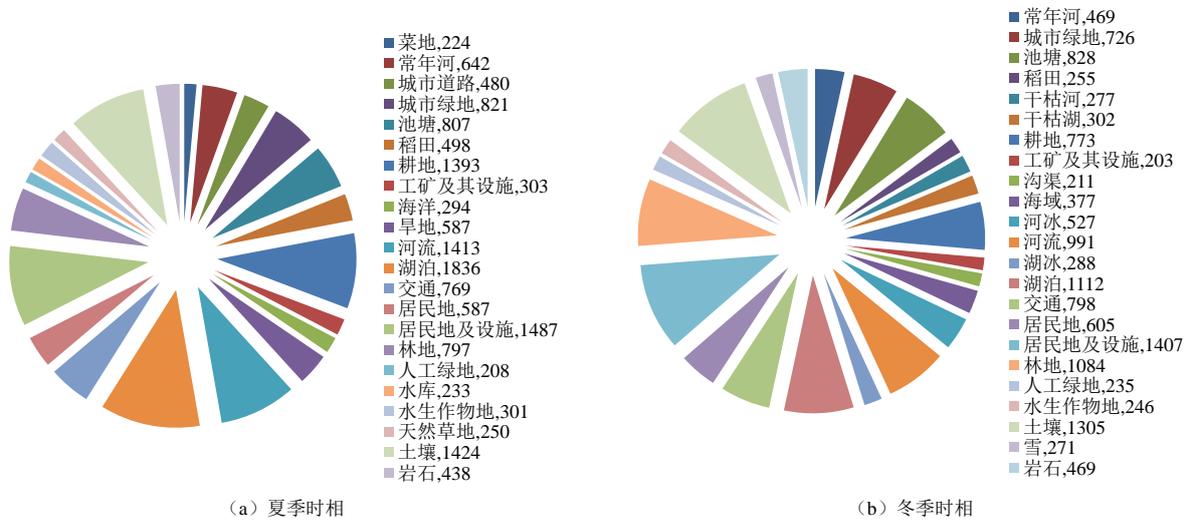


图3 一米级影像样本数量统计



图4 影像样本数据组织存储文件示例

其中, tif 文件为 7 像元×7 像元大小的原始影像文件; jpg 文件为 7 像元×7 像元大小的预览图像, 与原始影像对应; txt 文件为样本点中心像元的 DN 值文件, 内容的格式为两列数值, 以 Tab 键

分隔，第一列数据为波长，第二列数据为 DN 值；xml 文件为元数据描述文件，采用格式化存储方式，内容如图 5 所示：

```
<?xml version="1.0" encoding="UTF-8"?>
<影像波谱样本>
  <像元光谱>
    <像元光谱文件名> pixel_vege_AgriculLand_20160822032811_A001.txt</像元光谱文件名>
  </像元光谱>
  <图像文件名> image_vege_PMS_20160822032811_A001.tif</图像文件名>
  <快视图文件名> view_vege_PMS_20160822032811_A001.jpg</快视图文件名>
  <观测目标描述 />
  <经度> 124.066327</经度>
  <纬度> 46.156588</纬度>
  <地物类别> 11</地物类别>
  <像元光谱>
    <获取图像传感器>
      <传感器型号> PMS</传感器型号>
      <生产商 />
      <标定时间 />
      <文件格式说明 /> GeoTiff</文件格式说明>
    </获取图像传感器>
    <波段设置>
      <中心波长> 502;576;680;810</中心波长>
      <半波宽 />
      <信噪比 />
      <光谱分辨率 />
      <光谱漂移值 />
      <波段数> 4</波段数>
    </波段设置>
    <测量平台信息>
      <仪器平台> GF1</仪器平台>
      <测量高度 />
      <移动速率 />
      <备注 />
    </测量平台信息>
    <测量情况描述>
      <观测日期> 2016/08/22</观测日期>
      <观测时间> 03:28:11</观测时间>
      <视场角 />
      <瞬时视场角 />
      <每行像元数> 7</每行像元数>
      <图像行数> 7</图像行数>
      <投影方式 />
      <扫描速率 />
      <空间分辨率> 8</空间分辨率>
      </测量情况描述>
      <配套大气参数>
        <气溶胶光学厚度 />
        <大气水汽含量 />
        <湿度垂直廓线 />
      </配套大气参数>
      <人员信息>
        <处理人姓名> 唐言文</处理人姓名>
        <单位名称> 中国科学院遥感与数字地球研究所</单位名称>
        <联系电话> 010-64855178</联系电话>
        <负责人姓名> 赵理君</负责人姓名>
        <负责人单位名称> 中国科学院遥感与数字地球研究所</负责人单位名称>
        <负责人联系电话> 010-64855178</负责人联系电话>
      </人员信息>
      <几何纠正>
        <几何纠正精度 />
      </几何纠正>
      <大气纠正>
        <大气纠正精度 />
      </大气纠正>
      <辐射纠正>
        <辐射纠正精度 />
      </辐射纠正>
      <光谱采样方式>
        <在图像中行号> 4</在图像中行号>
        <在图像中列号> 4</在图像中列号>
        <观测天顶角> 0</观测天顶角>
        <观测方位角> 0</观测方位角>
        <太阳天顶角 />
        <太阳方位角 />
        <跨元面积比 />
        <光谱类型> 46</光谱类型>
        <样地名称 />
        </光谱采样方式>
      </影像波谱样本>

```

图 5 影像样本元数据描述文件内容示例

3 数据质量控制和评估

针对遥感影像样本数据，数据质量控制过程包括对待采样影像数据的检查整理，单个样本数据点的检查，样本点配套文件的生成和入库，以及元数据的编写、检查和入库。对待采样影像数据的检查包括影像投影信息、波段数、存储损坏等明显的的数据问题以及文件格式等。单个样本数据点的检查，主要针对类别标注错误的样本点数据进行修正、剔除。样本点配套文件以及元数据文件的检查包括文件命名、文件格式、文件内容中字段标准化命名以及文件内容的完整性等。

针对遥感影像样本数据集的构建，在数据入库阶段建立了完善的质量控制过程（图 6），保证已入库数据的正确性、完整性和一致性。在数据采集整理和入库过程的质量控制方面，进行了原始遥感影像数据和影像样本点数据的整理和格式统一，同时通过一系列质量控制方法，如正确性检查、数据一致性检查等，以保障数据的质量。影像样本点配套文件和元数据文件由影像样本数据采集人员根据项目制定的数据标准格式（包含图像文件名、经纬度、传感器型号、观测时间、空间分辨率、光谱类型人员信息等）进行填写，为减少人工填写引起的错误，所有元数据信息均通过程序编写自动从原始影像数据和采样点数据中读取填写得到。

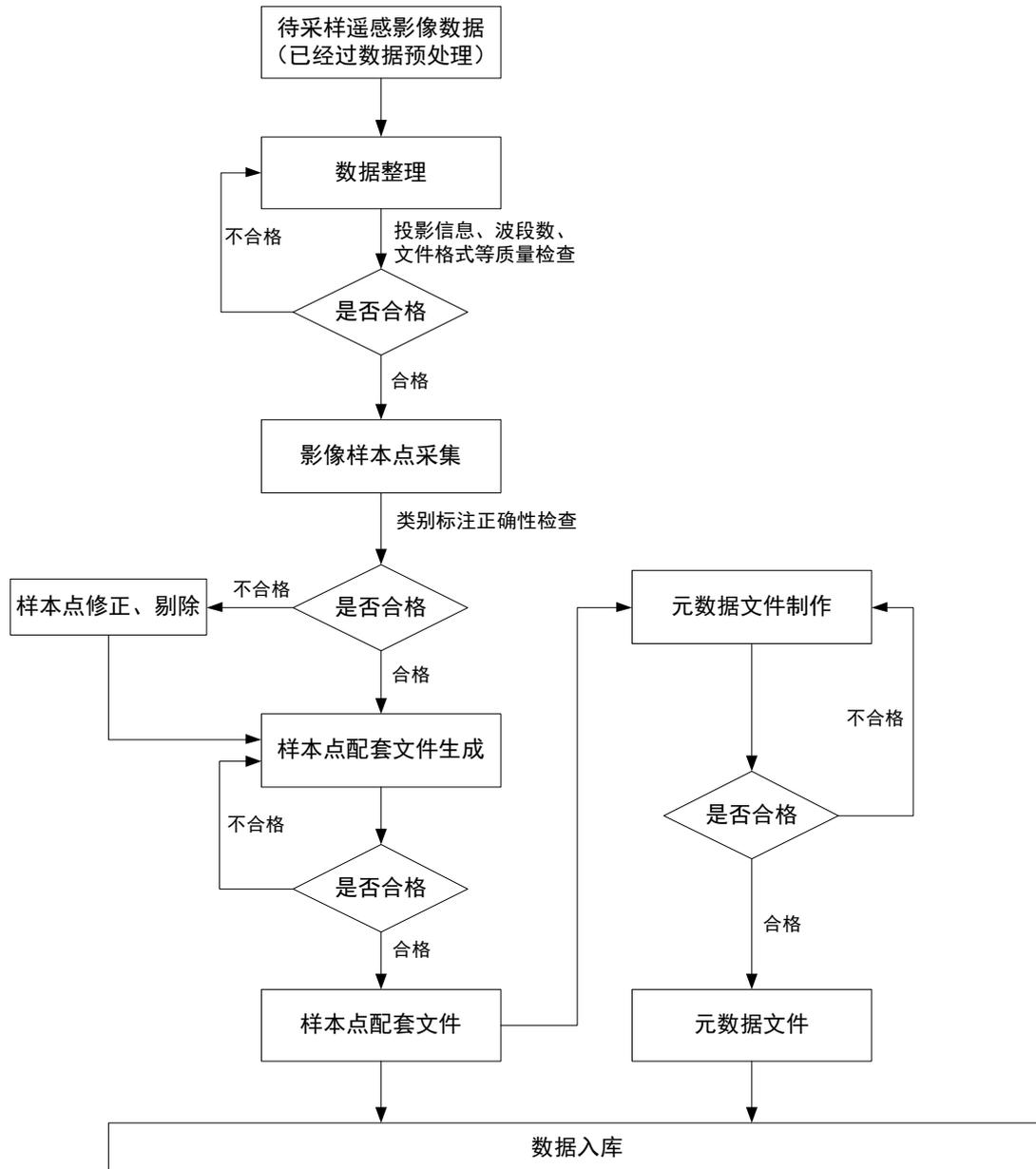


图 6 数据质量控制

为了对数据集的质量进行定量的精度评估，我们以样本采集过程中使用到的那景遥感影像的分类结果作为评价对象，将该影像中采集到的样本点随机划分为两部分，即训练样本和测试样本，使用训练样本训练支持向量机（Support Vector Machine, SVM）分类器，使用测试样本计算分类精度，统计总体分类精度和 Kappa 系数。我们随机选取了不同地区不同时相的样本进行评测（见表 4）。最终得到平均分类精度为 81.17%，平均 Kappa 系数为 0.78。从抽检结果可以看到，数据质量总体良好。

表 4 数据集质量精度评价

样点区域	季相	总体分类精度 (%)	Kappa 系数
天津地区	冬	82.25	0.80
新疆地区	冬	77.82	0.75

样点区域	季相	总体分类精度 (%)	Kappa 系数
天津地区	冬	82.25	0.80
黑龙江地区	冬	75.80	0.71
湖南地区	夏	84.75	0.82
广西地区	夏	85.25	0.81
平均值		81.17	0.78

4 数据使用方法和建议

本数据集是测绘地物波谱本底数据库平台的原始入库文件，可通过门户网站 (<http://210.72.27.78/spectrum/>) 检索查询获得检索区域内样本点的相关信息，也可以通过编程的方式批量解析读取本数据集 XML 文件中的样本点空间信息。数据集可以为遥感影像分类算法研究提供训练和测试样本数据，通过地理坐标与待分类影像像元坐标的转换，得到待分类影像中的样本点分布，使用最大似然 (Maximum Likelihood Classification, MLC)、支持向量机 (Support Vector Machine, SVM)、卷积神经网络 (Convolutional Neural Network, CNN) 等分类算法完成分类，使用流程如图 7 所示。由于本数据集并没有完全覆盖全国的每一个地区，如待分类影像的空间范围没有包含到任何影像样本点，建议相关研究者查询周边相邻地区的影像样本点，以筛选出的样本点为参考样本集，指导待分类影像中同类样本点的采集，进而利用所得到的样本点完成影像分类任务。

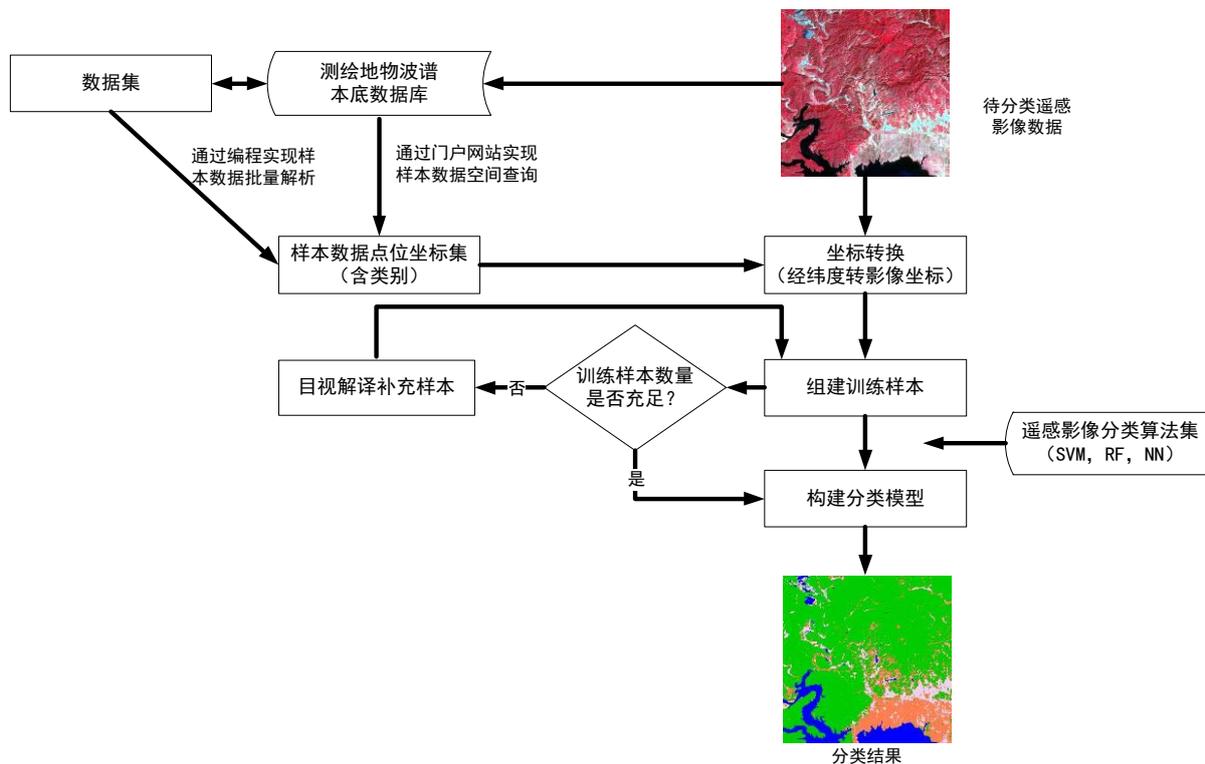


图 7 数据集用于影像分类的应用流程

另一方面，本数据集可以与地面实测波谱数据配合使用，通过与测绘地物波谱本底数据库平台提供的其他地面实测波谱数据集进行空间位置配对（相同和相近位置），研究和分析地面实测波谱与

影像波谱的关联关系，为相关算法的研究提供数据基础，为直接利用地面实测波谱数据指导影像样本自动采集提供可能。

数据作者分工职责

赵理君（1986—），男，河南省洛阳市人，博士，助理研究员，研究方向为遥感图像特征提取与分类。主要承担工作：影像样本数据集设计与采集规范、方案制定，数据质量检查，样本数据入库程序的编写。

郑柯（1963—），男，北京市人，博士，研究员，研究方向为遥感图像处理。主要承担工作：影像样本数据集制作方案可行性指导。

史路路（1991—），男，河南省周口市人，本科，硕士研究生，研究方向为遥感图像分类。主要承担工作：重庆、西藏、辽宁、吉林、黑龙江等地区十米级影像预处理与样本采集和东北地区米级影像预处理与样本采集以及数据质量检查与数据入库。

白洋（1991—），女，湖北省黄冈市人，本科，博士研究生，研究方向为遥感图像辐射处理。主要承担工作：北京、天津、河北、湖北、湖南等地区十米级影像预处理与样本采集和华北地区米级影像样本预处理与采集以及样本数据入库程序的优化。

唐吉文（1992—），男，湖南省永州市人，本科，博士研究生，研究方向为遥感图像目标检测与深度学习。主要承担工作：甘肃、新疆、贵州、四川、云南等地区十米级影像预处理与样本采集和西南地区米级影像预处理与样本采集以及数据汇集整理。

张伟（1992—），男，山西省运城市人，硕士，博士研究生，研究方向为遥感图像分类，机器学习与数据挖掘。主要承担工作：山东、江苏、上海等地区十米级影像预处理与样本采集和华东地区米级影像预处理与样本采集。

饶梦彬（1992—），女，江西省抚州市人，硕士，博士研究生，研究方向为高光谱遥感图像分类，机器学习。主要承担工作：江西、河南、海南、广西、广东等地区十米级影像预处理与样本采集和华南地区米级影像预处理与样本采集。

李艳艳（1990—），女，河北省唐山市人，硕士，博士研究生，研究方向为海洋遥感。主要承担工作：浙江、福建、安徽等地区十米级影像预处理与样本采集和华中地区米级影像预处理与样本采集。

邹松（1993—），男，湖北省仙桃市人，本科，硕士研究生，研究方向为遥感图像几何处理。主要承担工作：内蒙古、宁夏、青海、山西、陕西等地区十米级影像预处理与样本采集和西北地区米级影像预处理与样本采集。

参考文献

- [1] 乔五十, 郭喜绒, 刘妍, 等. 地表覆盖遥感制图耕地要素提取的方法与相关问题探讨[J]. 测绘标准化, 2013, 29(3): 21–23.
- [2] 张伟, 郑柯, 唐娉, 等. 深度卷积神经网络特征提取用于地表覆盖分类初探[J]. 中国图象图形学报, 2017, 22(8): 1144–1153.
- [3] 史泽鹏, 马友华, 王玉佳, 等. 遥感影像土地利用/覆盖分类方法研究进展[J]. 中国农学通报, 2012,

28(12): 273–278.

[4] 王昆, 戚浩平. 土地利用与土地覆盖遥感分类方法研究综述[J]. 山西建筑, 2008, 34(5): 353–354.

[5] GONG P, WANG J, YU L, et al. Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data[J]. International Journal of Remote Sensing, 2013, 34(7): 2607–2654.

[6] CHEN J, CHEN J, LIAO A P, et al. Global land cover mapping at 30 m resolution: A POK-based operational approach[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2015, 103: 7–27.

论文引用格式

赵理君, 郑柯, 史路路, 等. 全国地表类型遥感影像样本数据集[J/OL]. 中国科学数据, 2019, 4(2). (2019-02-04). DOI: 10.11922/csdata.2018.0058.zh.

数据引用格式

赵理君, 郑柯, 史路路, 等. 全国地表类型遥感影像样本数据集[DB/OL]. Science Data Bank, 2018. (2018-10-15). DOI: 10.11922/sciencedb.663.

Remote sensing image sample dataset of land cover types in China

Zhao Lijun^{1*}, Zheng Ke², Shi Lulu^{1,2}, Bai Yang^{1,2}, Tang Jiwen^{1,2}, Zhang Wei^{1,2}, Rao Mengbin^{1,2}, Zou Song^{1,2}, Li Yanyan^{1,2}

1. Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100101, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China

*Email: zhaolj01@radi.ac.cn

Abstract: This dataset of remote sensing images of land cover types in China is obtained through manual interpretation of ten-meter-level Landsat 8 multispectral image data and meter-level GF-1 and QuickBird multispectral image data, which covers 31 provinces, municipalities and autonomous regions of China excluding Hong Kong, Macao and Taiwan. Based on satellite images of the past five years (2013 – 2017), including summer and winter seasons, the study uses the classification system of the ground object spectral library. The dataset can offer priori training and testing samples for land cover classification, and provide sample data support for the research on the application of land cover classification. It can also be used to guide the collection of similar image sample points in the image. To promote convenient data usage, we adopted unified and standard data processing methods, sample collecting rules and quality control system, based on which the land cover image data are formed. So far, this dataset has been made publicly available online.

Keywords: land cover type; China; remote sensing image sample; ten meter level multispectral data; meter level multispectral data

Dataset Profile

Title	Remote sensing image sample dataset of land cover types in China
Data corresponding author	Zhao Lijun (zhaolj01@radi.ac.cn)
Data authors	Zhao Lijun, Zheng Ke, Shi Lulu, Bai Yang, Tang Jiwen, Zhang Wei, Rao Mengbin, Zou Song, Li Yanyan
Time range	2013 – 2017
Geographical scope	31 provinces, municipalities or autonomous regions of China excluding Hong Kong, Macao and Taiwan
Spatial resolution	2.4–30 m
Data volume	647 MB (after decompression)
Data format	*.tif, *.jpg, *.txt, *.xml
Data service system	http://www.sciencedb.cn/dataSet/handle/663
Sources of funding	Basic Research Foundation of Science and Technology (2014FY210800)
Dataset composition	<p>The dataset consists of two compressed files, made up of two folders storing meter-level sample data, namely GF1 and QuickBird, and one folder storing ten-meter-level sample data, namely Landsat. Each folder is comprised of several subfolders which are named after sampling regions. Each subfolder consists of compressed files that store data of six major land cover types in the sampling regions, including soil, water body, rock, vegetation, snow and ice, and man-made objects. Each of the compressed files includes only data of several samples of the same remote sensing image of the same land cover type. After being decompressed, each file is made up of four types of data files, including an original satellite image of the sample (*.tif), a sample image preview file (*.jpg), a text file of DN values of different spectral bands (*.txt), and a metadata file (*.xml).</p>