

雄安新区（马蹄湾村）航空高光谱遥感影像分类数据集

岑奕¹，张立福^{1*}，张霞¹，王跃明²，戚文超¹，汤森林¹，张鹏¹

1. 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；

2. 中国科学院上海技术物理研究所 中国科学院空间主动光电技术重点实验室，上海 200083；

摘要：高光谱遥感影像分类数据集主要用于辅助高光谱遥感分类算法的精度验证、效率评价及性能评估，一般包括高光谱遥感影像、影像对应地物类别标注以及相关信息文档等内容。常用的高光谱遥感影像分类数据集以欧美为主，如 India Pines、Salinas、KSC 等。随着我国高光谱遥感传感器技术发展和学术交流机制的日臻完善，国内也发布了高光谱遥感分类数据集，如江苏省常州的茶树数据集。相较于欧美高光谱遥感分类数据集的广泛应用，我国高光谱遥感分类数据集的发布与应用仍偏少。近年来，我国高质量高光谱遥感数据获取能力大幅增强，提升了我国高光谱遥感共享数据源的数量及质量，为促进我国高光谱遥感应用研究及业务化能力提供了支撑。本分类数据集包括雄安新区马蹄湾村高光谱影像数据，由上海技术物理研究所研制高分专项航空系统全谱段多模态成像光谱仪采集，光谱范围为 400-1000nm，波段 250 个，影像大小为 3750×1580 像元，空间分辨率 0.5m；同步实地调研地类分布 19 种，包括水稻茬、草地、榆树、白蜡、国槐、菜地、杨树、大豆、刺槐、水稻、水体、柳树、复叶槭、栎树、桃树、玉米、梨树、荷叶、建筑。利用随机森林分类方法对该数据进行了分类验证，分类精度可达 97%。该数据集可为我国经济作物高光谱精细分类研究提供良好的数据支持，更可为我国高光谱遥感载荷业务化应用发展提供有力促进。

关键词：高光谱遥感；雄安新区；航空影像；影像分类；数据集；

1 引言

高光谱遥感影像分类数据集主要用于辅助高光谱遥感分类算法的精度验证、效率评价及性能评估，一般包括高光谱遥感影像、影像对应地物类别标注、以及相关信息文档等内容（Zhang, 2016; Zhong, 2015; Du, 2016）。研究人员可通过对数据集给定地物样本的空间信息和光谱信息进行分析，选取可分性大的特征，利用分类算法对不同地物进行类别的划分和属性的判定，通过和数据集中地物类别标注结果对比验证，从而验证及评估分类算法精度。

常用的高光谱遥感影像分类数据集以欧美为主。India Pines 是最早用于高光谱图像分类的测试数据，由 AVIRIS（Airborne Visible Infrared Imaging Spectrometer）于 1992 年对美国印第安纳州西北部一片印第安松树试验区成像；Salinas 数据集，由 AVIRIS 对美国加利福尼亚州的 Salinas 山谷成像；Kennedy Space Center（KSC）数据集，由 AVIRIS 于 1996 年对美国佛罗里达州肯尼迪航天中心成像；Cuprite 数据集，由 AVIRIS 于 1997 年对美国内华达州矿区成像；Pavia University 数据集，由德国 ROSIS（Reflective Optics Spectrographic Imaging System）在 2003 年对意大利的帕维亚城成像；Botswana 数据

集，由 Hyperion 在 2001-2004 年对非洲奥卡万戈三角洲成像（Huang, 2017; Zhang, 2012; Li, 2016）。随着我国高光谱遥感传感器技术发展和学术交流机制的日臻完善，国内也发布了高光谱遥感分类数据集，如茶树等十种地类高光谱遥感数据集，由 PHI（Pushbroom Hyperspectral Imagery）于 1999 年对江苏省常州方麓村茶树种植基地成像（参数详见表 1）（Zhang, 2017）。相较于欧美高光谱遥感分类数据集的广泛应用，我国高光谱遥感分类数据集的发布与应用仍偏少。近年来，我国高质量高光谱遥感数据获取能力大幅增强，提升了我国高光谱遥感共享数据源的数量及质量，为促进我国高光谱遥感应用研究及业务化能力提供了支撑（Tong, 2014; Tong, 2016; Wang, 2016; Zhang, 2017; Zhao, 2016）。

本文拟以经济作物类型丰富的雄安新区马蹄湾村作为分类样区，基于机载高光谱成像仪获取的影像数据，利用辐射校正、几何校正、影像镶嵌、裁剪等数据预处理技术，结合同步开展地面实地调研调查，构建了雄安新区（马蹄湾村）航空高光谱遥感影像分类标准数据集。相较于已有分类标准数据集，该数据集具有光谱分辨率高、空间分辨率高、地物类别多等特点，可为高光谱精细分类研究提供良好的数据支持，更可服务于我国高光谱遥感应用载荷设计与论证。

表 1 目前公开发表的代表性高光谱遥感影像分类标准数据集

收稿日期：2019-12-22；

基金项目：国家重点研发计划项目（2017YFC1500900）；国家自然科学基金重点项目（41830108）

第一作者简介：岑奕，1979 年生，中国科学院遥感与数字地球研究所副研究员，研究方向：高光谱遥感应用，E-mail: cenyi@radi.ac.cn

通信作者简介：张立福，1967 年 5 月生，男，研究员，博导，研究方向为高光谱遥感应用技术，邮箱：zhanglf@radi.ac.cn

Table 1 public released standard dataset for hyperspectral remote sensing image classification

数据集	空间分辨率 (m)	谱段范围 (μm)	波段数	地类 (种)	影像大小 (pixel)	地物类别
India Pines	20	0.4-2.5	220	16	145×145	玉米等农作物
Salinas	3.7	0.4-2.5	224	16	512×217	休耕地、芹菜等农作物
Pavia University	1.3	0.43-0.86	115	9	610×340	树木、沥青道路、砖块、牧场等
Cuprite	20	0.4-2.5	224	30 (Jin, 2009)	250×191	蒙脱石等矿物
KSC	18	0.4-2.5	224	13	512×614	松树、阔叶树等植被
Botswana	30	0.4-2.5	242	14	610×340	混合豆木、刺槐、芦苇、泛滥平原草地等植被
茶树	2.25	0.42-0.85	80	10	348×512	马尾松、竹林、茶树等

2 数据研发方法

2.1 雄安新区航空高光谱遥感数据获取

2017年10月,中国科学院遥感与数字地球研究所联合中国科学院上海技术物理研究所河北省雄安新区开展了航空高光谱遥感数据获取试验。雄安新区位于中国河北省保定市境内,地处北京、天津、保定腹地,规划范围涵盖河北省雄县、容城、安新及周边部分区域,是继深圳经济特区和上海浦东新区之后又一具有划时代意义的新区。雄安新区地处冀中腹地,北距北京108公里,东距天津100公里,西距保定70公里,东西长26公里,南北阔25.5公里(图1)。全境西北地势较高,东南略低,海拔标高7—19米,为缓倾平原,土层深厚,地形开阔,植被覆盖率很低。该区地处中纬度地带,属暖温带季风型大陆性气候。



图1 雄安新区 google earth 影像图

Fig.1 Google Earth map of Xiongan New Area

上海技术物理研究所研制高分专项航空系统全谱段多模态成像光谱仪,本次飞行仅使用可见光、

近红外谱段光谱仪进行数据采集,该载荷总视场角40.6度,瞬时视场0.25mrad,有效推扫像元数2834,可适应最大速高比为0.04,具有较高的作业效率(Jia, 2018)。本次航飞航高2000m,测区东西长约48km,南北宽约27.5公里,总面积约1320平方公里,测区覆盖了雄县、安县、容县和白洋淀主要水域。飞行共规划航线21条,航向为东西方向,马蹄湾村地区位于第10、11条带(图2)。航飞天气晴朗无云,通视条件较好。



图2 雄安新区航带规划(橙色框为数据集位置)

Fig.2 Airline Planning of Xiongan New Area (the orange square shows the classification dataset location)

2.2 雄安新区航空高光谱遥感数据预处理

➤ 辐射校正

影像辐射校正通过遥感图像处理平台(the environment for visualizing images, ENVI)完成的,具体步骤如下:

步骤一,导入数据,利用载荷实验室定标参数,将高光谱遥感影像DN值转换为辐亮度值;

步骤二,利用FLASSH快速大气校正软件,进行大气校正,得到影像反射率数据,并乘以1000保

存。

➤ 几何校正

根据影像数据获取时，利用 GPS/INS 模块获取的定位定向数据，利用共线方程，解算出每个像元对应的地面点的地理编码。

步骤一，定位定向数据匹配。影像数据获取速度为 30HZ，而定位定向数据可达 50HZ，二者之间没有硬件上的同步装置，只能根据时间进行匹配。

将图像采集时间从影像中提取出来；根据图像采集时间和定位定向数据中的最接近该时间的两个定位姿态数据的时间进行计算，得到一个线性加权系统：

$$a = \frac{t_2 - t}{t_2 - t_1}, b = \frac{t - t_1}{t_2 - t_1} \quad (1)$$

其中，t 为图像采集时间，t1, t2 为最接近图像采集时间的两个定位姿态数据的时间。

根据加权系数，计算每个扫描行的定位定向数据。

步骤二，求取角元素与线元素。根据内方位元素与定位定向数据，分别计算出角元素与线元素。

步骤三，根据共线方程求取像元对应地面点坐标；

步骤四，重采样。

采用距离的加权分配，根据原始影像四个角点校正后的坐标与地面分辨率建立栅格，每个栅格的数据的灰度值由离其最近的四个像元按距离加权求得。

➤ 影像配准

根据提供的高分辨率的航空正射影像，将经过粗校正的单条带影像与其进行配准，以提高几何精度为后期影像镶嵌做准备。

影像配准采用的是遥感图像处理平台（the environment for visualizing images, ENVI），registration-Image Registration Workflow 功能，具体步骤如下，

步骤一，导入数据，以高分辨率航空正射影像作为 Base Image，单条带高光谱影像作为 Warp Image；

步骤二，通过与软件交互，选择两个影像的连接点，研究共选择 54 个地面控制点，均匀分布于整个区域；

步骤三，利用一阶多项式进行转换，并进行保存。

配准误差控制在 2 个像元以内。

➤ 影像拼接裁剪

使用经过配准后的高光谱数据，将单条带数据进行镶嵌成图。

步骤一，使用 ENVI 中，Mosaicking-Seamless Mosaic 功能对高光谱进行拼接处理。

步骤二，使用 ENVI 中，利用另存为功能选取目标区进行裁剪保存。

2.3 地物类别实地调查

为更准确的获取影像区域的地物类别，和航空飞行同步开展实地地物类别调研，利用 GPS 对不同地物类别区块进行定位，并对周边位置信息加以描述记录。野外共调查样地 57 块，拍摄实地照片 39 张，样地在 google earth 上的展点如图 3 所示。调查地物类别共计 19 类，包括经济作物如国槐、栾树、杨树、柳树、白蜡、以及大豆、玉米和水稻等。借助 GPS 匹配，通过人工勾绘，得到马蹄湾村地物类别标注图。



图 3 地物类别实地调研样点分布

Fig.3 Sample point distribution of field investigation

3.标准数据集描述

雄安新区（马蹄湾村）航空高光谱遥感影像如图 4 所示，其光谱范围为 400-1000nm，波段 250 个，影像大小为 3750×1580 像元，空间分辨率 0.5m。通过地物类别实地调研，该影像共勾绘标注地物类别 19 类（图 5），主要以经济作物为主，各地类所占像元数如表 2 所示。



图 4 马蹄湾村航空高光谱遥感影像

Fig.4 Aerial hyperspectral remote sensing image of Matiwang Village

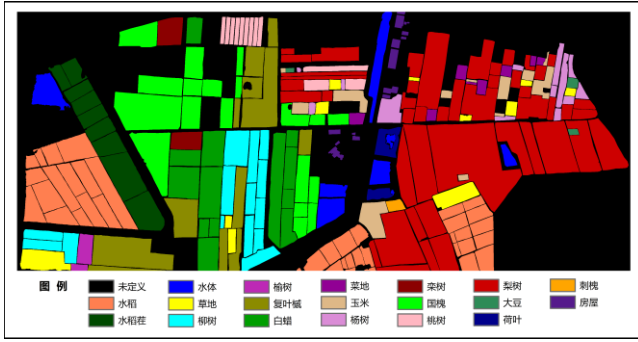


图5 马蹄湾村地物类别标注图

Fig.5 Real land cover map of Matiwan Village

表2 马蹄湾村各地类样本数

Table 2 Sample number of different land cover in Matiwan Village

类别	水稻	水稻茬	水体	草地
样本数	26138	187425	124862	91518
类别	柳树	榆树	复叶槭	白蜡
样本数	197218	19663	296538	276755
类别	栎树	国槐	桃树	菜地
样本数	44232	372708	67210	29763
类别	玉米	杨树	梨树	大豆
样本数	85547	68885	986139	7456
类别	荷叶	刺槐	建筑	
样本数	27178	6506	26140	

数据集包括：(1) 马蹄湾村航空高光谱反射率图像及示意图；(2) 水稻茬、草地、榆树、白蜡、国槐、菜地、杨树、大豆、刺槐、水稻、水体、柳树、复叶槭、栎树、桃树、玉米、梨树、荷叶、建筑等 19 种典型地类的 ROI 及示意图。数据集存储为 .img 和 .tif 格式，压缩后数据量为 2.4GB。该数据适于遥感地物的识别和高光谱遥感分类算法的研究和验证等。

表3 雄安新区（马蹄湾村）航空高光谱遥感分类数据集
Table 3 Aerial Hyperspectral Remote Sensing Classification Dataset of Matiwan Village

文件名称	类型	备注
farm	hdr	马蹄湾村航空高光谱遥感数据 envi 标准格式头文件
farm	img	马蹄湾村航空高光谱 envi 标准格式遥感图像
farm_roi	hdr	马蹄湾村地物类别样本 envi 标准格式头文件
farm_roi	img	马蹄湾村地物类别样本 envi 标准格式遥感图像
马蹄湾村航空高光谱遥感影像图	png	
马蹄湾村地物类别标注图	png	

4. 数据分类结果

基于雄安新区马蹄湾村航空高光谱遥感影像，

分别提取影像的前三个主成分以及每个主成分对应的八种空间纹理特征（均值、二阶矩、异质性等），以及对植被和水体敏感的植被指数 NDWI 和 NDVI。利用集成学习方法中的随机森林分类算法对研究区遥感影像进行分类研究。

纹理特征主要通过灰度共生矩阵方法，灰度共生矩阵是图像灰度级间联合概率密度函数的矩阵形式，常用 $P(i, j, d, \theta)$ 表示。均值（Mean）、方差（Variance）均质性（Homogeneity）、对比度（Contrast）、异质性（Dissimilarity）、熵（Entropy）、二阶矩（Second Moment）和相关性（Correlation）等 7 种纹理特征。

均值主要反映图像局部平均灰度值的变化，计算公式为

$$f_{\text{mean}} = \frac{1}{L^2} \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P(i, j, d, \theta) \quad (2)$$

方差反映像素值和均值偏差的度量，计算公式为

$$f_{\text{variance}} = \sqrt{\frac{1}{m \times n} \sum_{i=1}^m \sum_{j=1}^n [F(i, j) - \bar{F}]^2} \quad (3)$$

其中， m 和 n 分别表示融合图像的行列数， \bar{F} 表示融合图像所有像元值的均值， $F(i, j)$ 表示融合图像 (i, j) 位置的像元值。

均质性反映图像局部的均匀程度，图像越均匀，均质性数值越大，计算公式为

$$f_{\text{domo}} = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \frac{1}{1 + (i - j)^2} P(i, j, d, \theta) \quad (4)$$

对比度反映了图像清晰度和纹理沟纹的深浅程度。图像的纹理沟纹越深，说明其对比度数值越大，其视觉效果也越清晰。对比度的计算公式为

$$f_{\text{cont}} = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i - j)^2 P(i, j, d, \theta) \quad (5)$$

异质性反映了图像局部像元灰度的差异，图像上地物灰度差异越大，异质性数值越大。计算公式为

$$f_{\text{diss}} = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} |i - j| P(i, j, d, \theta) \quad (6)$$

熵是对图像局部信息量的一种有效度量，主要是度量图像局部灰度级分布的随机性，反映图像局部纹理的复杂程度。图像局部的纹理越复杂，熵值越大。熵的计算公式为

$$f_{\text{entr}} = - \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P(i, j, d, \theta) \log P(i, j, d, \theta) \quad (7)$$

二阶矩也称为能量矩，反映图像局部灰度级分布的均匀程度和纹理的粗细，纹理越粗，二阶矩的数值越大。具体计算公式为

$$f_{Sm} = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P(i, j, d, \theta)^2 \quad (8)$$

相关性主要度量灰度共生矩阵中各元素在行或列方向上的相似度。图像局部元素值越接近，相关性值越大。相关性的计算公式为

$$f_{\text{corr}} = \frac{\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} ijP(i, j, d, \theta) - u_1 u_2}{\sigma_1^2 \sigma_2^2} \quad (9)$$

其中, $u_1 = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} iP(i, j, d, \theta)$,

$$u_2 = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} jP(i, j, d, \theta),$$

$$\sigma_1 = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i - u_1)^2 P(i, j, d, \theta),$$

$$\sigma_2 = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (j - u_2)^2 P(i, j, d, \theta).$$

其中 i 与 j 分别为起点与终点灰度值, d 和 θ 分别为给定的空间距离和方向。

$$\text{NDVI 计算公式: } (\rho_{800} - \rho_{670}) / (\rho_{800} + \rho_{670}) \quad (10)$$

$$\text{NDWI 计算公式: } (\rho_{560} - \rho_{860}) / (\rho_{560} + \rho_{860}) \quad (11)$$

随机森林的构建是一个重复构建决策树的过程。假设训练样本的总数为 N , 每个样本具有 M 个属性, 构建决策树的总数为 T , 则将随机森林应用于图像分类的基本思想如下:

①从原始训练样本中随机有放回地抽取 N 个训练样本, 形成样本集来训练一个决策树, 并将训练样本集作为根节点处的样本。

②在决策树的每一节点分裂时, 都随机地从 M 个特征中选取 m 个特征 ($m < M$), 利用分割函数从 m 个属性中选取最佳的分裂特征。

③重复步骤②, 直到决策树中的任意叶子节点不可再分为止, 即构建了一棵决策树, 在整个决策树生成过程中没有进行剪枝;

④重复以上三个步骤 T 次, 即构建了 T 棵独立的决策树, 形成了森林。

在生成随机森林的过程中, 对每一棵决策树不进行剪枝, 让其最大限度的增长, 最终将所有的决策树组合成随机森林。完成随机森林的构建后, 将样本输入分类器中, 对于每个样本每棵决策树都输出对应的预测值对其类别进行投票, 最终投票数最多的一类为该样本最终确定的类别。

本研究分类训练样本选择采用分层抽样方式, 训练样本与验证样本比例约为 2:1。分类结果如图 6 所示。各类别的分类精度以及相应的混淆矩阵分别

如表 4 和表 5 所示。

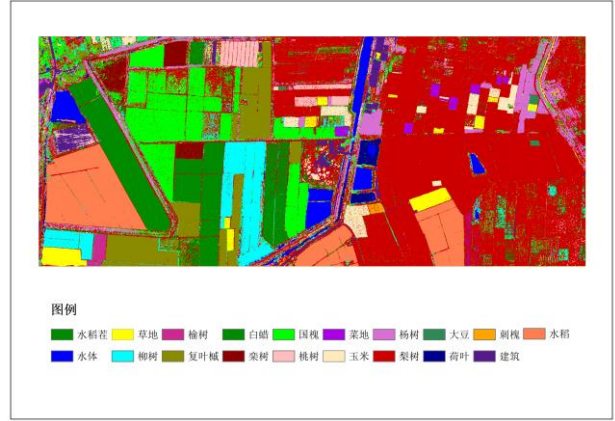


图 6 马蹄湾村航空高光谱遥感影像分类图

Fig.6 Land cover classification map of Matiwan Village using aerial hyperspectral remote sensing data

表 4 基于随机森林分类算法各类别精度统计
Table 4 Accuracy statistics of different land cover using random forest classification algorithms

类别	水稻	水稻茬	水体	草地
精度	0.99	0.99	0.99	0.96
类别	柳树	榆树	复叶槭	白蜡
精度	0.98	0.93	0.98	0.96
类别	栎树	国槐	桃树	菜地
精度	0.96	0.96	0.94	0.92
类别	玉米	杨树	梨树	大豆
精度	0.92	0.95	0.99	0.93
类别	荷叶	刺槐	建筑	Kappa
精度	0.96	0.84	0.99	0.9764

由表 4 可知, 随机森林分类算法得到分类图的总体分类精度可达 97%, kappa 系数为 0.98。在 19 类地物中, 刺槐的分类精度偏低。由混淆矩阵 (表 5) 可知, 刺槐、梨树以及复叶槭混分较为严重。三种植被的反射率光谱如图 7 所示。

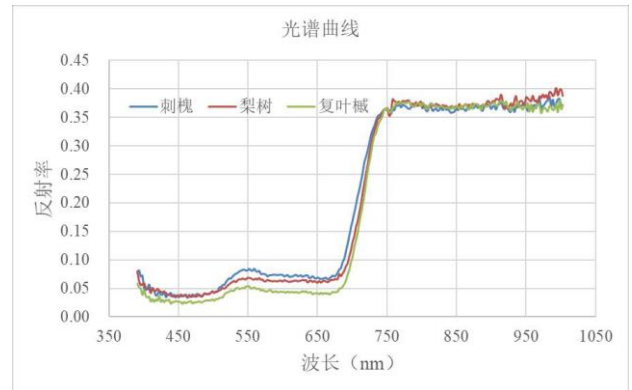


图 7 刺槐、梨树以及复叶槭的反射率光谱

Fig. 7 The reflectance of Robinia pseudoacacia, pear and Acer

由图 7 可知, 三种混分植被的光谱曲线基本重合, 光谱相似度较高, 从而导致错分现象严重。

5 结 论

雄安新区（马蹄湾村）航空高光谱遥感影像分类标准数据集以经济作物类型丰富的雄安新区马蹄湾村作为分类样区，基于机载高光谱成像仪航飞获取的影像数据，利用辐射校正、几何校正、影像镶嵌、裁剪等数据预处理技术，结合同步开展地面实地调研地类分布构建。该数据集高光谱影像光谱范围为 400-1000nm，波段 250 个，影像大小为 3750 × 1580 像元，空间分辨率为 0.5m，勾绘标注有经济作物等地物类别 19 类，具有光谱分辨率高、空间分辨率高、地物类别多等特点。基于影像前三个主成分及其对应八种空间纹理和植被指数，利用随机森林分类方法对该数据进行了分类验证，分类精度可达 97%。本分类标准数据集为国内外同行提供了目前国内外空间分辨率、光谱分辨率最高的航空高光谱共享数据，不仅可为高光谱分类技术的研究提供良好的数据支持，更可服务于我国高光谱遥感应用载荷设计与论证。

数据下载方式：

<http://www.hrs-cas.com/a/share/shujuchanpin/2019/0501/1049.html>

表 5 基于随机森林分类结果对应的混淆矩阵

Table 5 Confusion matrix of different land cover using random forest classification algorithms

实际/预测	水稻	水稻茬	水体	草地	柳树	榆树	复叶槭	白蜡	栎树	国槐	桃树	菜地	玉米	杨树	梨树	大豆	荷叶	刺槐	建筑
水稻	21223	0	0	2	1	0	1	0	0	1	1	3	0	0	3	0	0	0	0
水稻茬	0	9326	0	2	0	1	0	1	0	0	2	0	0	0	2	0	0	0	0
水体	0	0	6237	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
草地	0	0	0	4390	2	0	0	4	0	2	0	0	2	1	10	0	0	0	0
柳树	4	6	0	26	9658	29	48	14	0	50	0	12	24	11	31	0	9	1	0
榆树	0	0	0	0	2	918	1	0	0	2	0	0	0	0	0	0	0	0	0
复叶槭	34	24	0	11	47	14	14549	66	2	55	4	2	58	2	98	1	22	1	0
白蜡	1	2	0	14	4	0	61	13327	46	97	6	0	23	2	53	5	3	0	0
栎树	0	0	0	0	1	0	0	2	2125	5	1	0	0	0	2	0	0	0	0
国槐	5	2	0	9	19	6	10	26	0	17945	6	9	29	12	68	1	7	1	0
桃树	0	0	0	0	0	0	0	1	0	1	3159	0	0	0	2	1	0	0	0
菜地	1	0	0	1	0	0	0	0	0	0	0	1366	0	0	2	1	0	0	0
玉米	0	1	1	3	1	0	0	0	0	2	0	0	3938	4	5	0	0	0	0
杨树	0	0	1	1	0	0	14	1	0	5	1	0	6	3301	14	0	0	0	0
梨树	39	10	1	116	124	15	142	394	39	467	181	96	197	110	49014	14	5	48	6
大豆	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	350	0	0	0
荷叶	0	0	3	1	2	0	0	1	0	3	0	0	0	0	1	0	1312	0	0
刺槐	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	274	0
建筑	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	0	0	0	13011

参考文献(References)

- Du P.J., Xia J.S., Xue Z.H., Tan K., Su H.J., and Bao R. 2016. Review of hyperspectral remote sensing image classification. *Journal of Remote Sensing*, 20(2) : 236-256. (杜培军, 夏俊士, 薛朝辉, 谭琨, 苏红军, 鲍蕊. 2016. 高光谱遥感影像分类研究进展. *遥感学报*, 20(2):236-256) [DOI 10.11834/jrs.20165022]
- Huang S.G., Zhang H.Y., Pizurica A. 2017. A Robust Sparse Representation Model for Hyperspectral Image Classification, *Sensors*, 17(9) :1-18. [DOI: 10.3390/s17092087]
- Jia J.X., Wang Y.M., Cheng X.Y., Yuan L.Y., Zhao D., Ye Q., Zhuang X.Q., Shu R., Wang J.Y. 2018. Destriping Algorithms Based on Statistics and Spatial Filtering for Visible-to-Thermal Infrared Pushbroom Hyperspectral Imagery. *IEEE Transactions on Geoscience and Remote Sensing*.doi: 10.1109/TGRS.2018.2889731
- Jin Q.H., Zhu L.L., Zhang L.X., Jiang Y.H., 2009. Examples of using hyperspectral remote sensing technology for mineral resource evaluation and mining environment monitoring, *Geological Bulletin of China*, 28, 278-284. (金庆花, 朱丽丽, 张立新, 江永宏. 2009. 矿产资源评价与矿山环境监测中高光谱遥感技术方法应用的实例. *地质通报*, 28:278-284) [DOI]
- Li X.K., Wu T.X., Liu K., Li Y., and Zhang L.F. 2016. Evaluation of the Chinese Fine Spatial Resolution Hyperspectral Satellite TianGong-1 in Urban Land-Cover Classification. *Remote Sensing*, 8(5), pp: 438-455.
- Tong Q.X., Xue Y.Q., Zhang L.F. 2014. Progress in Hyperspectral Remote Sensing Science and Technology in China Over the Past Three Decades, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 7: 70-91. [DOI 10.1109/JSTARS.2013.2267204]
- Tong Q.X., Zhang B., Zhang L.F. 2016. Current Progress of hyperspectral remote sensing in China. *Journal of Remote Sensing*. 20(5):689-707. (童庆禧, 张兵, 张立福. 2016. 中国高光谱遥感的前沿进展. *遥感学报*, 20(5):689-707) [DOI:10.11834/jrs.20166264]
- Wang Y.M., Jia J.X., He Z.P., and Wang J.Y., Key technologies of advanced hyperspectral imaging system. *Journal of Remote Sensing*, 20(5) : 850-857. (王跃明, 贾建鑫, 何志平, 王建宇. 2016. 若干高光谱成像新技术及其应用研究. *遥感学报*, 20(5) : 850-857) [DOI:10.11834/jrs.20166206]
- Zhang, L.F., Zhang L.P., Tao D., and Huang X. 2012. On Combining Multiple Features for Hyperspectral Remote Sensing Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 50(3): 879-893.
- Zhang, L.F., Jiao, W.Z., Zhang H.M., Huang, C.P., and Tong, Q.X. 2017. Studying drought phenomena in the Continental United States in 2011 and 2012 using various drought indices. *Remote Sensing of Environment*, 190:96-106.
- Zhang X., Sun Y.L.*, Shang K., Zhang L.F., Wang S.D. 2016. Crop Classification Based on Feature Band Set Construction and Object-Oriented Approach Using Hyperspectral Images, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(9): 4117-4128.
- Zhang X., Zhang B., Zhang L.F., Sun Y.L., 2017. Hyperspectral remote sensing dataset for tea farm, *Global Change Research Data Publishing & Repository*. (张霞, 张兵, 张立福, 孙艳丽. 茶树等十种地类高光谱遥感数据集. 全球变化科学研究数据出版系统) [DOI: 10.3974/geodb.2017.03.04.V1]
- Zhao B., Zhong Y.F., Zhang L.P. 2016. A spectral-structural bag-of-features scene classifier for very high spatial resolution remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*. 116: 73-85. [DOI: 10.1016/j.isprsjprs]
- Zhong Y.F., Wu Y., Xu X., and Zhang L.P., An Adaptive SubPixel Mapping Method Based on MAP Model and Class Determination Strategy for Hyperspectral Remote Sensing Imagery, *IEEE Transactions on Geoscience and Remote Sensing*, 53(3): 1411-1426.

Aerial Hyperspectral Remote Sensing Classification Dataset of Xiongan New Area (Matiwan Village)

Cen Yi¹, Zhang Lifu¹, Zhang Xia¹, Wang Yueming², Qi Wenchao¹, Tang Senlin¹, Zhang Peng¹

1. *State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100101, China;*

2. *Key Laboratory of Space Active Opto-Electronics Technology, Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai 200083, China;*

Abstract: Aerial hyperspectral remote sensing dataset play an important role on the research of hyperspectral image, such as classification. However, few works has been done on the establishment of standard hyperspectral dataset. This paper introduced a standard hyperspectral dataset, which including hyperspectral remote sensing image, land cover map, and sensor parameters, the dataset was acquired by a newly designed airborne hyperspectral sensor, accompanied with the synchronous ground survey experiments.

(1) **Data acquisition:** Aerial hyperspectral remote sensing image of Xiongan New Area was acquired using the visible and near infrared imaging spectrometer designed by Shanghai Institute of Technical Physics, CAS in October 2017. The total field of view angle of the spectrometer is 40.6 degree, the instantaneous field of view is 0.25 mrad, effective push-scan pixel is 2834, and the maximum speed-to-height ratio is 0.04. The flight height is 2000 m, and the flight areas cover the Xiong County, An County, Rong County and Baiyangdian Lake, of which the East-West length is 48 km, the North-South width is 27.5 km, and the total area is 1320 km². There are 21 flight lines with east-west

direction, while the Matiwan Village is located in the 10th and 11th flight line. The flying weather is clear and cloudless, and the visibility condition is good. And radiation correction, geometric correction, image mosaic and clipping were done before data classification.

(2) **Dataset:** The spectral range of the aerial hyperspectral remote sensing image of Xiongan New Area (Matiwan Village) is 400-1000 nm, with 250 bands and a spatial resolution of 0.5m. The image size is 3750 x 1580 pixels. And the land cover types labeled here are 19 types, mainly cash crops.

(3) **Classification result:** With the first three principal components of the spectrum and its corresponding eight spatial texture features and the vegetation index, such as NDWI and NDVI, the aerial hyperspectral remote sensing image of Matiwan Village in Xiongan New Area were classified using random forest classification algorithm. The total classification accuracy is 97%, and the kappa coefficient is 0.98. According to the confusion matrix, the confusion of Robinia pseudoacacia, Pear tree and Acer complex is serious. that cause the classification accuracy of Robinia pseudoacacia is low.

(4) **Conclusion:** An aerial hyperspectral remote sensing dataset of Xiongan New Area (Matiwan Village) with high spatial resolution and spectral resolution was shared in this study. The dataset was classified using random forest classification algorithm, and the total classification accuracy is 97%. It shows that the dataset can provide good data support for hyperspectral classification research, and also can serve for the design and demonstration of hyperspectral imaging spectrometer.

Key words: hyperspectral remote sensing, Xiongan New Area, Aerial image, image classification, dataset

Supported by National Key R&D program of China (No. 2017YFC1500900) and National Natural Science Foundation of China (41830108)