文章编号: 1672-8785(2023)08-0001-12

高光谱目标分类技术发展与展望

张 彬* 孙岩森 李晓杰 刘 珮 (海军航空大学航空作战勤务学院,山东 烟台 264000)

摘 要:高光谱成像技术源于遥感探测,具有谱图合一的独特优势,在农林、 地矿、防伪和环境保护等领域的应用日益广泛。作为高光谱技术最常见和最基 本的功能,目标分类在各个高光谱领域的应用均具有至关重要的作用。综述了 高光谱目标分类技术的发展现状,分析了高光谱数据的结构特点,归纳了高光 谱目标分类的一般流程,并详细阐述了数据读取、图像预处理和目标分类等高 光谱目标分类流程的主流方法及其基本原理。结合近年来的典型案例,分析了 高光谱目标分类的过程,并对其发展趋势进行了分析和展望。

关键词: 高光谱; 目标分类; 图像预处理; 特征提取

中图分类号: TP18 文献标志码: A DOI: 10.3969/j.issn.1672-8785.2023.08.001

Development and Prospect of Hyperspectral Target Classification Technology

ZHANG Bin*, SUN Yan-sen, LI Xiao-jie, LIU Pei

(Aviation Operations and Service Institute, Naval Aviation University, Yantai 264000, China)

Abstract: Hyperspectral imaging technology originated from remote sensing, has the unique advantage of spectral integration. It is increasingly widely used in agriculture and forestry, geology and mining, anti-counterfeiting and environmental protection. As the most common and basic function of hyperspectral technology, target classification plays a vital role in the application of hyperspectral fields. In this paper, the development status of hyperspectral target classification technology is reviewed, the structural characteristics of hyperspectral data are analyzed, the general process of hyperspectral target classification are described in detail. Combined with the typical cases in recent years, the process of hyperspectral target classification is analyzed, and the development trend is analyzed and prospected.

Key words: hyperspectral; target classification; image preprocessing; feature extraction

收稿日期: 2023-03-10

基金项目:国家自然科学基金项目(62005318)

作者简介:张彬(1981-),男,山东临沂人,讲师,主要研究方向为高光谱成像与图像处理。

*通讯作者: E-mail: yeaweam@163.com

1

0 引言

光谱成像技术是一种将成像技术与光谱技 术相结合的数据采集手段。采用该技术获得的 数据立方体既包含了二维空间信息,又包含了 一维光谱信息。基于传统的黑白二维照片,人 们能够获得目标的空间分布和形貌特征。随着 黑白照片升级为彩色照片,人们能够借助 R、 G、B三个波长通道的颜色信息区分形貌类似 的目标。而用光谱成像技术获得的数据立方体 将波长维度进一步细分,探测到的目标携带了 更多的信息,能够帮助人们进行更精细的区 分。高光谱成像技术是光谱成像技术的一个分 支。此外,光谱成像技术还包括几个到十几个 光谱通道数的多光谱成像技术和几百个光谱通 道数的超光谱成像技术。而高光谱成像技术通 常具有几十到上百个光谱通道[1]。由于多光谱 技术的光谱通道数相对较少,难以获取一条平 滑连续的光谱曲线,应用场景相对单一。而超 光谱技术的光谱通道数多,光谱分辨率极高, 设备研制成本高,因此它通常用于特殊需求的 场景。高光谱技术的光谱分辨率适中,通常在 几纳米到十几纳米之间。对于任意空间位置都 可以提取出一条完整平滑的光谱曲线。通过相 应的数据分析手段能够获取二维照片以及多光 谱数据无法得到的信息。因此,高光谱技术作 为应用最广、需求最高的成像光谱技术,自问 世以来就备受关注。

高光谱成像技术起源于高光谱遥感技术, 时至今日已同激光雷达技术、合成孔径雷达技 术并称为对地观测信息获取的三大核心技 术^[2]。针对高光谱成像技术的研究起始于 20 世纪 80 年代,美国喷气推进实验室(Jet Propulsion Laboratory, JPL)首次提出了在航空航 天领域进行 100~200 个波段的高光谱成像的 可行性^[3]。1983 年,在美国国家航空航天局 (National Aeronautics and Space Administration, NASA)的支持下, JPL 研制出国际上第 一台机载成像光谱仪(Airborne Imaging Spectrometer, AIS)^[4]。我国的高光谱成像技术则 起步于 20 世纪 90 年代初,中国科学院上海技术物理研究所最先研制出国内首台机载成像光 谱仪——实用型模块化成像光谱仪(Operational Modular Imaging Spectrometer, OMIS)^[5]以及 宽视场推扫式高光谱成像仪(Pushbroom Hyperspectral Imager, PHI)^[6]。

目前,高光谱成像技术已经从遥感应用落 地,并逐步实现商业化。以美国 Headwall 公 司和芬兰 Specim 公司为代表的多家光谱仪器 制造商推出了多款不同工作波段、不同成像方 式的高光谱成像仪,并占据了世界上绝大部分 的市场份额。与此同时,国内公司也逐步推出 了多款成像光谱仪及其配套的高光谱数据采集 系统整机。随着高光谱成像技术的进一步发 展,它已广泛应用于农林^[7-11]、地矿^[12-13]、防 伪[14-15]和环境保护[16-20]等领域。面对人们日益 增长的需求,除了高光谱成像仪硬件系统研发 技术之外,对高光谱数据处理技术也提出了非 常高的要求。不同于传统的二维图像数据或一 维光谱数据,高光谱数据是三维数据立方体, 包含了海量的数据。光谱数据可用于对不同物 质成分的鉴别,因此将成像与光谱相结合的高 光谱数据最常用于对裸眼不易区分的物质进行 分类与定位。

随着近年来越来越多的科研人员投身于高 光谱数据处理技术研究,越来越多的研究领域 对高光谱目标分类应用提出需求。本文围绕用 于目标分类的高光谱数据处理技术展开综述和 分析。

1 高光谱数据处理流程

高光谱数据处理的主要目的是分析、挖掘 和剥离有效数据,以便从中获取潜在的有用信 息。借助高光谱数据光谱维度的隐藏信息,并 结合空间维度的形态特征分布,可以实现不同 类别目标的区分和可视化表达。图1所示为高 光谱数据处理的通用流程。接下来将依次对数 据读取、图像预处理以及用于目标分类的数据 分析方法进行综述。



图 1 高光谱数据处理的通用流程

2 高光谱数据读取

类似于灰度图像,三维高光谱数据是将三 个维度(空间维 X、空间维 Y 和光谱维 λ)按照 一定的排列顺序进行存储的。每个数值代表该 三维坐标下的灰度值。高光谱数据通常由数据 文件和头文件组成。其中,数据文件包含数据 立方体内各个坐标下的灰度值信息;头文件则 用于修饰数据文件,包含了数据的位深、水平 和竖直分辨率、光谱通道数和存储类型等。

在读取高光谱数据时,通过头文件获取其 存储类型非常重要。若不匹配存储类型,读取 的数据将杂乱无章。高光谱数据主要有三种数 据存储格式:波段按行交叉(Band Interleaved by Line, BIL)、按波段顺序(Band Sequential, BSQ)和波段按像元交叉(Band Interleaved by Pixel, BIP)^[21-22]。其中,BSQ格式对应于滤光 片切换型高光谱成像仪。该成像仪采用面阵相 机获取图像,每次拍摄获取单个波段的整个空 间维信息,再通过切换滤光片或利用声光可调 谐滤光器(Acousto-Optic Tunable Filter,AO-TF)和液晶可调谐滤光器(Liquid Crystal Tunable Filter,LCTF)等器件可控地改变透光波长, 从而实现波长维度扫描。因此,存储 BSQ 格

式数据时,按照波段顺序先存储一个波段的二 维灰度信息,然后再存储下一波段的二维灰度 信息。BIL 格式对应于推扫型高光谱成像仪。 该成像仪采用面阵相机获取图像,每次拍摄线 视场及其所有波段信息,再通过垂直于线视场 方向的扫描获取数据立方体。因此,存储 BIL 格式数据时,按照空间行切换的顺序先存储一 行数据及其波长信息,然后再存储下一行数据 及其波长信息。BIP 格式对应于推帚式高光谱 成像仪。该成像仪采用线阵相机获取图像,每 次拍摄空间一点的光谱信息,然后通过二维扫 描获取数据立方体。因此,存储 BIP 格式数据 时,按照像元存储顺序先存储一个像元对应的 波长信息,然后再存储下一个像元对应的波长 信息。三种数据存储类型及其对应的数据获取 方式如图2所示。

3 高光谱数据预处理方法

完成高光谱数据读取后,首先需要对原始 数据进行预处理。由于数据采集过程受到外界 环境、相机噪声和照明条件等因素的影响,不 进行平滑和校准等预处理操作将导致处理结果 偏差,从而影响分析精度。此外,高光谱数据 的立方体数据量庞大。针对不同应用需求往往 只需若干不同波段或不同位置的数据。进行降 维等预处理操作能够消除数据冗余,保留有效 数据,从而极大地缩短数据分析时间,提升数 据分析精度。下面对目前主流的高光谱数据预 处理方法的原理进行综述。

3.1 数据校正

3.1.1 萨维茨基-戈莱平滑法

平滑是最常用的高光谱数据预处理方法, 其中萨维茨基-戈莱(Savitzky-Golay, S-G)平滑 法效果最优^[23-24]。其主要目的是消除由噪声引 起的光谱曲线高频抖动,保留光谱曲线有效峰 谷形貌,提升后续高光谱处理精度。该方法主 要包括窗口宽度 N 和拟合阶数 k 两个可调参 数。主要思路如下:对第 i 个参考点及其左右 各 N 个点(合计 2N+1 个点)的光谱数据进行 最小二乘拟合(拟合多项式系数为 k),求解多



图 2 高光谱数据存储类型及其对应的数据获取方式

项式系数 a_0 , a_1 , \dots , a_{k-1} 后, 将拟合值 F_i 代替 参考点原有的光强值, 即

$$F_{i} = a_{0} + a_{1}x + \dots + a_{k-1}x^{k-1}$$

(x = -N, -N+1, ..., N) (1)

用该方法遍历整个波段,即可得到平滑后 的光谱曲线(见图 3)。窗口宽度 N 会影响去噪 效果。N 越大则平滑后消除的噪声宽度越大, 也更容易将有用信息去除。拟合阶数 k 会影响 拟合精度。k 越大则拟合精度越高,但算法速 度也相应变慢。阶数过高将导致过拟合现象, 使精度反而变差。因此,S-G 平滑过程是通过 调整参数来评估平滑后的曲线能否保留原光谱 曲线趋势以及噪声是否被有效去除。调参评估 往往需要人为监督,常与后续数据处理方法联 合使用,通过最终效果进行评价。

3.1.2 多元散射校正

在对目标进行高光谱成像时,目标表面不 均匀会导致所测光谱的变化被目标表面散射掩 盖。多元散射校正(Multiple Scattering Correc-



tion, MSC) 是一种解决此类问题的方法^[25-26]。 MSC 能够有效消除由于散射水平不同造成的 光谱差异,增强光谱与数据之间的相关性。该 方法利用理想光谱/标准光谱来修正测试光谱 的基线平移和偏移。在实际操作过程中,一般 选择所有数据的平均值作为理想光谱/标准光 谱,即

$$\bar{S} = \frac{\sum_{i=1}^{N} S_i}{N} \tag{2}$$

http://journal.sitp.ac.cn/hw



图 4 MSC 校正效果图

对每个样本的光谱与标准光谱进行一元线 性回归,求解基线平移 k_i和偏移 b_i:

$$S_i = k_i \bar{S} + b_i \tag{3}$$

利用基线平移和偏移计算校正后的光谱。 MSC 校正的效果图如图 4 所示。

3.1.3 标准正态变量变换

标准正态变量(Standard Normal Variation, SNV)变换采用加权平均的方式来消除目标固体颗粒大小不一、表面散射等因素对测量光谱 的影响^[27-28]。SNV 变换假设标准光谱满足正 态分布。该方法将原光谱减去平均光谱后再除 以标准偏差,实质就是实现原光谱数据的标准 正态化,满足

$$Z_{ij} = \frac{x_i - \bar{x}_i}{S_{ij}} \tag{4}$$

$$\bar{x}_i = \frac{\sum_{i=1}^N x_{ij}}{N} \tag{5}$$

$$S_{ij} = \sqrt{\frac{\sum_{i=1}^{N} (x_{ij} - \bar{x}_i)^2}{N - 1}}$$
(6)

式中,*i*为样本个数,*j*为光谱通道数,*Z_{ij}为*SNV变换后的数据。

SNV 方法是目前应用较多的预处理方法, 主要针对噪声以及数据采集过程中环境、样品 和系统等对光谱数据产生的影响。此外,预处 理方法还包括一阶导数、二阶导数、连续小波 变换、去偏置和去偏移等^[29]。此类方法能够 最大限度地将样品内在因素之外的影响降到最 低,从而凸显样本本身因素对光谱的影响,为 (b) で正效果图

后续基于光谱进行目标分类打下基础。

3.2 降维/特征波长提取

3.2.1 连续投影算法

连续投影算法(Successive Projections Algorithm, SPA)是一种在矢量空间中利用共线性 最小化的前向迭代搜索算法,能够在全波段中 快速筛选出几个有效波长来解决高光谱数据冗 余问题^[30-31]。

记 *x*_{k(0)} 和 *M* 分别为初始迭代向量和需要 提取的变量个数,光谱矩阵为 *J* 列(即光谱通 道数为 *J*), SPA 的具体步骤如下:

(1)任选一个光谱通道(假设为第 j 列), 将建模集的第 j 列赋值给 x_j,并将该通道记为 x_{k(0)}。

(2)将未选入的光谱通道集合记为 s:

$$s = \{j, 1 \leqslant j \leqslant J, j \notin \{k(0), \}$$

$$k(1), \cdots, k(m-1)\}\}$$
(7)

(3)用 x_i 分别对剩余的各个光谱通道/向 量进行投影计算。

(4)选取向量投影最大的光谱通道,将该通道赋值给 x_i。

(5)令 *m*=*m*+1, 若 *m*<*M* 成立,转至步骤(2),否则 SPA 算法结束。

利用 SPA 算法筛选出的光谱通道基本能 够代表原始数据的信息。通常利用均方根误差 (Root Square Mean Error, RMSE)等参数对筛 选效果进行评价。随着光谱通道数的增加, RMSE 指标会降低。当达到一定的通道数量



图 5 SPA 效果验证

后,该指标将不再明显降低,从而验证通道数 选择的正确性(见图 5)。

3.2.2 主成分分析法

主成分分析(Principal Component Analysis, PCA)法的主要思想是将多维特征向量映 射到维度更低的特征向量上,即将高维特征转 换为少数几个主要成分,主成分之间互不相关 且包含最大的信息量^[32-35]。

记录高光谱数据为一个 m×n 的矩阵:

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{nm} \end{pmatrix} = (x_1, \cdots, x_n) \quad (8)$$

式中, m为样本个数, n为光谱通道数。

PCA 法的主要步骤如下:

(1)将矩阵 *X* 进行标准化变换,其均值为 *x*,标准差为 *S*_n,满足

$$X_{nm} = \frac{x_{nm} - \bar{x}}{S_n} \tag{9}$$

式中, 原始矩阵经过标准化得到矩阵 $X = (X_1, \dots, X_n)$ 。

(2)计算相关系数矩阵及其特征值和特征向量。

(3)计算方差贡献率并确定主成分,一般选择 90%作为贡献率阈值,即选择贡献率之和大于 90%的前 k 个主成分。

PCA 法是目前应用较多的降维预处理方法,其主要目的是将高维度且数据高度冗余的高光谱数据进行精简,并保留对后续数据分析

贡献最大的波段。此外,降维预处理方法还包括最小二乘回归分析、连续小波变换和分窗格 拉姆-施密特变换等^[36-38]。降维预处理方法能 够有效降低高光谱数据维数,解决数据冗余问 题,从而提高后续数据分析效率。

3.3 其他预处理方法

除了数据校正和降维这两类主要的高光谱 数据预处理方法之外,还有许多其他预处理方 法。根据特定的应用场景,预处理方法还包括 几何校正^[39-40]、辐射校正^[41-42]和数据拼接^[43] 等。所有预处理方法的目的都是在进行数据分 析前对原始数据进行预处理操作,以提升后续 处理的精度和速度。

4 用于目标分类的数据分析方法

高光谱数据分类技术是一种对未标记的像 素进行标号的过程,利用高光谱数据中的光谱 信息和空间分布信息对不同的像素进行分类。 按照训练样本是否有标签将分类算法分为监督 分类、无监督分类和半监督分类;按照是否利 用了空间信息将分类算法分为基于光谱信息的 分类和基于空-谱信息的分类^[44-45]。下面按照 第二种情况对目标分类方法进行综述。

4.1 基于光谱信息的目标分类方法

4.1.1 K最近邻算法

K最近邻(K-Nearest Neighbors, KNN)算 法是一种基于欧几里得距离推断事物类别的分 类方法^[46-47]。通过计算两个目标之间的距离, 完成高光谱分类。距离计算方法如下:

$$D(x,y) = \sqrt{\sum_{k=1}^{n} (x_k - y_k)^2} \qquad (10)$$

式中, *x* 和 *y* 分别为两个目标, *x_k* 和 *y_k* 代表 两个目标在第 *k* 个维度的坐标, *n* 为目标坐标 的维度。

KNN 算法的主要步骤如下:

(1)选择一个参考目标;

(2)依次计算其他目标相对于参考目标的 欧氏距离 D;

(3)对计算出的 D 进行升序排列,选择前 K 个目标并将其定义为同一类别。

http://journal.sitp.ac.cn/hw

KNN 算法能够用于非监督分类,然而 K 参数的取值将直接影响分类结果。K 选择太大 将导致分类模糊,K选择太小则容易受个例影 响。当 KNN 算法用于监督分类时,根据分类 标签进行 K 参数调节,选择合适的 K 值。

4.1.2 线性判别分析法

线性判别分析(Linear Discriminant Analysis, LDA)法是一种监督分类方法,即需要事 先知道训练集的样本分类结果[48-50]。该方法将 高维特征向量向低维特征向量作线性变换。此 变换使得属于同一种类的样本数据特征接近, 不同种类的样本特征差别很大。如图 6 所示, 以二维样本分为两个类别为例,设法构造一条 一维直线, 使得两个类别在该直线上的投影尽 可能远,而两个类别内部投影距离尽可能近。



图 6 LDA 法的示意图

LDA 法通过构造一个低维"分界线"来 将各样本向其投影,满足投影后类内方差最 小、类间方差最大。该分界线在监督分类样本 下获得,能够用于测试集和未知样本的分类。

4.1.3 支持向量机法 支持向量机(Support Vector Machine, SV-M)法是一种监督分类方法,将线性不可分问 题映射到高维空间,建立超平面从而实现分 类^[51-53]。SVM法的基本思想是将所有有标签 训练样本映射到一个高维空间中,并在此空间 内建立一个超平面; 该超平面既可对不同类别 的样本进行分类,又使其两侧的空白区域有最

SVM法的超平面多维特性使其在高维模 型、非线性模型、小样本数据模型识别中有很 大优势,并能与其他分类方法结合,产生新的 算法模型^[54]。

4.1.4 光谱角制图法

光谱角制图(Spectral Angle Mapper, SA-M)法是一种常用的光谱相似性计算方法^[55-56]。 通过计算参考光谱与目标光谱矢量夹角的大小 来度量二者之间的相似性。光谱角越小说明目 标光谱与参考光谱是同一类的可能性越大。光 谱角的计算公式如下:

$$\theta = \arccos\left[\frac{\sum_{i=1}^{n} A^{\mathrm{T}}B}{\sqrt{\sum_{i=1}^{n} A^{\mathrm{T}}A} \sqrt{\sum_{i=1}^{n} B^{\mathrm{T}}B}}\right] \quad (11)$$

式中, A 和 B 分别代表参考光谱数组和目标光 谱数组,*i*表示波段的序号,*n*为总波段数。

4.2 基于空-谱信息的目标分类方法

4.2.1 基于组合核的分类算法

基于组合核的分类算法是利用组合核函数 策略将目标的空域和谱域特征相结合的方 法[57-58]。在讨论该算法前需要先介绍核函数。

低维空间线性不可分的数据通过非线性映 射到高维特征空间则可能实现线性可分,但是 直接确定非线性映射函数的形式和参数、特征 空间维数非常困难。而核函数就可以解决该问 题。设 $x,z \in \mathbb{R}^n$, $F \in \mathbb{R}^m$ ($n \ll m$), 即x, z为低 维空间,F为高维空间。非线性函数 φ 是低维 空间到高维空间的映射,核函数 K(x, y)满足

 $K(x,y) = \langle \varphi(x), \varphi(y) \rangle$ (12)式中, <>为内积。不难看出, 核函数运算解 决了低维空间到高维空间的映射问题。

基于组合核的分类方法将原本仅包含谱域 信息的核函数加入至空域信息, 使分类过程中 引入空域特征。结合核函数的概念可以知道, 该方法需要依托带有核函数的光谱信息分类方 法,而且它具有低维到高维映射的特征。比较 常见的有前面提到的 SVM 法。

4.2.2 马尔可夫随机场法

马尔可夫随机场(Markov Random Field,

7

http://journal.sitp.ac.cn/hw

大的间隔。

MRF)法利用空间信息的马尔可夫性来描述数 据背景并建立异常检测算法^[59-60]。MRF 理论 利用地物空间连续分布的特性并基于邻域模型 建立各个像元的空间相关性模型,且在像元类 条件概率计算中综合考虑光谱信息,能够有效 实现光谱-空间信息联合的地物分类与识别。

马尔可夫性指的是图像中一个像素的统计 特性只与其附近的一个小领域中的像素有关, 而与其他位置的像素无关,即

 $P(y_i | y_j, j \neq i) = P(y_i | y_j, j \in \delta_i)$ (13) 式中, δ_i 指的是与像素 *i* 相邻的像素。

4.3 其他方法

除了上述方法之外,最近比较常见的高光 谱图像分类方法还有决策树法^[61-62]、随机森 林法^[63-64]和深度学习方法^[65-70]等等。

决策树分类法基于树结构原则,按照一定 的分类依据将原始数据分为特征更为均质的子 集,属于多元统计分类法的一种。这些子集在 数据结构中称为节点,其基本思想是利用一组 自变量来预测每个样本最可能对应的类型(即 因变量)。通过光谱曲线寻找特征波段,并利 用决策树分类法进行波段运算,从而实现目标 分类。

随机森林是一个包含多个决策树的分类 器,其输出类别由个别树输出类别的众数来确 定。该方法善于处理大量输入参数的情况,并 且能够平衡不对称数据资料带来的误差。

深度学习是机器学习领域的一个宽泛概 念,是学习样本数据内在规律,最终实现类似 人一样具有分析学习能力的技术。深度学习用 于高光谱数据分类时具有极佳的表现,在大量 标记样本训练加持下可以达到极高的分类 精度。

5 分类评价标准

分类评价标准用于对高光谱分类结果进行 定量评价,是衡量分类方法优劣的重要准则。 利用建立的分类模型/分类器对测试样本进行 分类,统计测试样本中每类像素分类正确的样 本数量,建立分类误差矩阵,然后计算分类评

INFRARED (MONTHLY)/VOL.44, NO.8, AUG 2023

价指标。分类评价标准包括整体精度(Overall Accuracy, OA)、类内精度(Class Accuracy, CA)、平均精度(Average Accuracy, AA)、卡帕(Kappa)系数等^[71-72]。

误差矩阵又称为混淆矩阵(Confusion Matrix, CM),用于描述真实类别和分类器预测的 分类是否相同。假设数据需分类类别的总数为 *k*,则混淆矩阵 *X* 可以表示为

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{k1} & \cdots & x_{kk} \end{pmatrix}$$
(14)

式中,矩阵元素 x_{ij}表示测试样本实际为第 i 个 类别而分类器预测为第 j 个类别的样本数目, 即混淆矩阵中的对角线元素为被正确分类的样 本点个数。

OA 用于评价测试样本的整体分类效果, 为分类正确的总样本数 *n* 与测试样本总数 *N* 之比:

$$OA = \frac{n}{N} = \frac{\sum_{i=1}^{K} x_{ii}}{\sum_{i=1}^{K} \sum_{j=1}^{K} x_{ij}}$$
(15)

CA 是每一类数据的准确率,为该类别分 类正确样本总数与样本总数之比:

$$CA^{k} = \frac{n^{k}}{N^{k}} \tag{16}$$

AA 是所有类内精度的平均值:

$$AA = \frac{\sum_{i=1}^{K} CA^{i}}{K}$$
(17)

Kappa 系数是一种综合性的评价指标,其 表达式为

$$Kappa = \frac{OA - \frac{\sum_{i=1}^{K} n^{i} \times N^{i}}{N \times N}}{1 - \frac{\sum_{i=1}^{K} n^{i} \times N^{i}}{N \times N}}$$
(18)

相较于总体准确率和类内准确率,Kappa 系数 更能同时利用正确分类样本和错误分类样本的 数量来评价一种算法的效果。Kappa 系数一般 在 0 到 1 之间,当其小于 0.4 时说明预测一致 性很差,当其大于 0.8 时说明一致性极好。

http://journal.sitp.ac.cn/hw

6 目标分类应用实例及分析

以2019年刘爽等人针对小麦赤霉病籽粒 的识别算法为例,简要阐述高光谱目标分类的 实际处理方法[73]。在数据读取过程中,成像 光谱范围两端部分波段的机器噪声较大,对后 续处理有较大影响。因此对该部分成像波段进 行剔除,并利用图像分割算法选择合适的掩膜 来提取小麦种子区域。在图像预处理过程中, 分别采用 S-G 平滑和 SNV 两种算法进行数据 校正操作。其中,S-G 平滑算法可以有效降低 由成像系统机械噪声引起的光谱曲线抖动噪 声, SNV 算法可以减小样本表面散射及样本 固体颗粒形状导致的光谱误差。此外,分别使 用 PCA 和 SPA 两种算法进行降维操作。其 中, PCA 算法选取前 4 个主成分(总贡献率为 99.3533%), SPA 算法筛选出 RMSE 最小时 的8个波长。在数据分析(即建立分类模型)的 过程中,首先用 LDA 算法对图像预处理后的 数据集进行测试,发现 S-G 平滑后的图像检测 效果明显优于 SNV 算法。因此分别采用 LDA、KNN和SVM算法对经过S-G预处理及 PCA、SPA 降维的图像进行建模检测。基于数 据校正、降维和分类模型三个过程中不同算法 的排列组合,采用 OA 和 Kappa 系数对分类一 致性进行评价,最终选择 S-G-SPA-SVM 为 最佳算法组合。其OA达到97%, Kappa系数 达到 0.94 (见表 1)。为了充分发挥高光谱成像 技术的优势, 文献[73]还对分类后的结果进行 了可视化处理(见图 7)。

从文献[73]的应用实例不难看出,在处理 实际的高光谱图像目标分类检测问题时,可在



图 7 可视化分类识别图像[73]

数据校正、特征波长提取、分类模型构建等几 个步骤中选择合适的处理方法,并通过排列组 合找到最佳的分类方案。面对不同的应用场 景,选择的处理方法和评价方式往往不同,所 以需要有针对性的设计。除此之外,结合算法 调参和引入机器学习方法能够进一步提升分类 精确度,使高光谱技术在目标分类方面具有极 高的优势,特别是在农林业生产、工业产品检 测和军事侦察等领域拥有广阔的应用前景。

7 结束语

本文将高光谱数据目标分类处理流程划分 为数据读取、图像预处理和数据分析三个阶 段,并针对高光谱目标分类应用依次对三个阶 段用到的方法及其原理进行了综述,最后总结 了常用的目标分类评价指标,并举例说明了在 实际的高光谱数据目标分类应用中各处理环节 对处理方法的选择和组合优化方式。在具体应 用中,并非简单地将各步骤的常规方法叠加, 而是在了解各算法优缺点和适用场景后对各算 法进行排列组合与参数调节,并利用分类评价 指标对不同算法组合和关键参数条件下的分类

分类模型	建模集		测试集		Kanna 玄 粉
	OA/ %	RMSE	$OA/\frac{0}{0}$	RMSE	Kappa 东奴
S-G-PCA-LDA	86.17	0.2019	91.5	0.2915	0.83
S-G-PCA-KNN	86.67	0.1946	90.5	0.3082	0.81
S-G-PCA-SVM	95.83	0.2041	91.5	0.2915	0.83
S-G-SPA-LDA	85.83	0.2038	93	0.2646	0.86
S-G-SPA-KNN	87.83	0.1876	92.5	0.2739	0.85
S-G-SPA-SVM	95.33	0.2160	97	0.1732	0.94

表1 文献[73]中不同分类模型的识别结果对比

http://journal.sitp.ac.cn/hw

结果进行对比,以选择最佳的组合与参数。

随着高光谱成像仪硬件性能的不断提升, 尤其是国产高光谱仪器的崛起,越来越多的科 研工作者、各领域应用方都开展了大量的研究 工作, 也推动着高光谱数据处理方法的迭代升 级。高光谱数据处理精度和速度的提升,同样 意味着新领域大门的开启。比如在工业在线检 测领域, 高光谱成像技术凭借非接触式无损快 速检测的优势并结合在线实时分析算法,能够 在生产线上完成大批量、高精度、无人化的检 测作业,在"机器替人"大趋势下工业领域的 良莠筛选、识别分类、品质控制等方面都具有 非常广阔的应用前景。在发光照明领域,高光 谱成像技术凭借高通量和高精度的优势,通过 机器学习方法完成数据建模,能够很好地满足 检测发光材料的发光效率、评估发光色差、评 价发光均匀性等应用需求,对于我国超大规模 发光材料和器件的检测工作是一个利好消息。 可以乐观地推断,在未来工业、农业、科技、 军事、环境等诸多领域需求的驱动下,高光谱 技术将会得到长足的发展,并逐步从基础研究 和实验研究转向实际的生产应用。届时将对高 光谱图像处理算法特别是目标分类方法提出更 大的需求量,并且会对分类精度、处理速度以 及半监督/非监督分类识别方面提出更高的要 求。这也将是未来高光谱图像目标分类技术发 展的方向。

参考文献

- [1] 张若岚,陈洁.从单波段到超光谱——面向多 维信息感知的红外光谱成像技术[J].红外技 术,2014,36(4):257-264.
- [2] 王建宇, 李春来. 高光谱遥感成像技术的发展 与展望[J]. 空间科学学报, 2021, 41(1): 22-33.
- [3] Goetz A F H, Vane G, Solomon J E, et al. Imaging spectrometry for earth remote sensing [J]. Science, 1985, 228(4704): 1147–1153.
- [4] Vane G, Goetz A F H, Wellman J B. Airborne imaging spectrometer: a new tool for remote sens-

ing [J]. IEEE Transactions on Geoscience and Remote Sensing, 1984, **22**(6): 546–549.

- [5] 刘银年,薛永祺,王建宇,等. 实用型模块化成像光谱仪[J]. 红外与毫米波学报,2002,21
 (1):9-13.
- [6] 董广军,张永生,范永弘. PHI 高光谱数据和高空间分辨率遥感图像融合技术研究[J]. 红外 与毫米波学报,2006,25(2):123-126.
- [7] 谭洋. 高光谱土壤有机质含量反演精度影响因 素分析及模型优化 [D]. 长春: 吉林大学, 2022.
- [8] 李睿妍. 基于高光谱的果园土壤主要速效养分 估测研究 [D]. 泰安: 山东农业大学, 2022.
- [9] 苍圣. 基于压缩感知与稀疏表示的高光谱遥感 影像森林分类方法研究 [D]. 哈尔滨:东北林 业大学,2021.
- [10] 杨荣超. 基于叶片高光谱图像的树种识别研究 [D]. 北京: 北京林业大学, 2020.
- [11] 冯宝坤. 高光谱数据与 LiDAR 数据协同的树种 分类研究 [D]. 昆明:西南林业大学, 2020.
- [12] 叶发旺,张川,邱骏挺,等.铀矿勘查岩心高光 谱技术研究及应用十年进展[J].世界核地质 科学,2022,39(4):623-636.
- [13] 肖树群.基于深度学习的高光谱遥感重金属污染识别研究 [D].石家庄:河北地质大学, 2022.
- [14] 杨戈. 基于高光谱的钞票与印刷品的图像分析 方法的研究 [D]. 北京:北京印刷学院, 2020.
- [15] 赵静远,张俊芹,孙梅,等.基于高光谱成像的 羊肉掺假可视化无损定量检测[J].食品与机 械,2022,38(10):61-68.
- [16] 肖文. 基于高光谱视觉的建筑垃圾多特征检测 方法及实验研究 [D]. 厦门: 华侨大学, 2021.
- [17] 张金诺. 水稻白叶枯病抗性表型的获取与分析 方法研究 [D]. 杭州:浙江大学, 2022.
- [18] 丁俐文. 基于高光谱的风化煤矿源腐植酸含量 速测模型研究 [D]. 泰安:山东农业大学, 2021.
- [19] 柴红. 谷子田土壤颗粒有机碳氮的高光谱监测 研究 [D]. 太原:山西农业大学,2019.
- [20] 胡兴堂. 高光谱水质遥感监测系统关键技术研究 [D]. 北京:中国科学院遥感应用研究所, 2006.

- [21] 曲海成.面向光谱解混的高光谱图像快速处理 技术研究 [D].哈尔滨:哈尔滨工业大学, 2016.
- [22] 钱嘉伟. 基于双目异质机器视觉的检测方法及 应用研究 [D]. 杭州: 浙江大学, 2014.
- [23] 白铁成,王涛,陈佑启,等.南疆沙尘区骏枣叶 片水分含量检测的近红外光谱预处理方法对比
 [J].光谱学与光谱分析,2019,39(4):1323-1328.
- [24] 谢树刚. 基于高光谱的黄河三角洲土壤有机质 含量估测模型研究 [D]. 泰安:山东农业大学, 2021.
- [25] Chu Y W, Tang S S, Ma S X, et al. Accuracy and stability improvement for meat species identification using multiplicative scatter correction and laser-induced breakdown spectroscopy [J]. Optical Express, 2018, 26(8): 10119–10127.
- [26] 张东辉,赵英俊,秦凯,等.光谱变换方法对黑 土养分含量高光谱遥感反演精度的影响[J]. 农业工程学报,2018,34(20):141-147.
- [27] 刘莉,陶红燕,方静,等.基于近红外高光谱的 梨叶片炭疽病与黑斑病识别[J].农业机械学 报,2022,53(2):221-230.
- [28] 张磊. 基于高光谱技术的喀斯特地区耕地土壤 含水量反演[D]. 贵阳:贵州大学, 2022.
- [29] 李杰,李尚科,蒋立文,等.基于近红外光谱技术与化学计量学的绿茶无损鉴别方法研究[J]. 分析测试学报,2020,39(11):1344-1350.
- [30] 牛芳鹏,李新国,麦麦提吐尔逊·艾则孜,等. 基于连续投影算法的博斯腾湖西岸湖滨绿洲土 壤有机碳含量的高光谱估算[J].浙江大学学 报(农业与生命科学版),2021,47(5):673-682.
- [31] 张楠楠,张晓,王城坤,等.基于高光谱和连续 投影算法的棉花叶面积指数估测[J].农业机 械学报,2022,53(S1):257-262.
- [32] 王丹丹,李燕,张庆银,等. 基于主成分分析的 黄瓜新品种引进筛选综合评价 [J]. 北方园艺, 2022,46(23):21-28.
- [33] 杜俊鹏,吕军,吴计生,等.基于主成分分析法 和系统聚类分析法的河流水质评价研究[J]. 水利技术监督,2022,30(12):216-220.
- [34] 武斌, 沈嘉棋, 汪鑫, 等. 主成分分析排序和模

http://journal.sitp.ac.cn/hw

糊线性判别分析的生菜近红外光谱分类 [J]. 光谱学与光谱分析, 2022, **42**(10): 3079-3083.

- [35] 曹晓兰,邓梦洁,崔国贤.高光谱结合主成分分 析的苎麻品种识别[J].光谱学与光谱分析, 2019,**39**(6):1905-1908.
- [36] 高永强. 连续小波变换高光谱数据降维挖掘系 统设计 [J]. 激光杂志, 2020, **41**(7): 137-140.
- [37] 曹英丽,肖文,刘亚帝,等.高光谱数据降维与 水稻氮素含量解析方法 [J].沈阳农业大学学 报,2021,52(1):109-115.
- [38] 李世波,林辉,葛森.东洞庭湖湿地植被高光谱 数据降维与分类 [J].中南林业科技大学学报, 2019,39(11):36-41.
- [39] 翟伟林,刘建霞,李金富,等. 机载高光谱 CA-SI/SASI 数据几何校正及精度评定 [J]. 地质找 矿论丛, 2022, **37**(2): 244-248.
- [40] 胡海斌,张浩,李聪,等.惯导约束下的高光谱 图像行相关几何校正算法[J].测绘科学, 2018,43(7):1-6.
- [41] 张迷霞,孙继鑫,贺希敏,等.无人机载高光谱 大气辐射校正方法比较与精度评价[J].测绘 与空间地理信息,2022,45(11):98-101.
- [42] 许凯秋. 基于光谱角约束的遥感影像辐射校正 研究 [D]. 武汉:武汉大学, 2020.
- [43] 李赛, 尹球, 胡勇, 等, 基于 SPHP 的推扫式高 光谱航空影像拼接 [J]. 红外与毫米波学报, 2021, **40**(1): 64-73.
- [44] 杨国鹏. 基于机器学习方法的高光谱影像分类 研究 [D]. 郑州: 解放军信息工程大学, 2010.
- [45] 黄良. 高光谱遥感图像高精度分类方法研究 [D]. 长春:中国科学院长春光学精密机械与物 理研究所, 2021.
- [46] 赵晋陵, 胡磊, 严豪, 等. 联合局部二值模式与 K-最近邻算法的高光谱图像分类方法 [J]. 红 外与毫米波学报, 2021, 40(3): 400-412.
- [47] 李婵, 王俊杰, 邬国锋, 等. 基于叶片光谱特征 的农业区域植物分类 [J]. 深圳大学学报(理工 版), 2018, 35(3): 307-315.
- [48] 朱梦远. 基于高光谱成像技术的水稻纹枯病病 害早期检测研究 [D]. 南京:南京农业大学, 2018.
- [49] Hou Q, Wang Y, Li C. Semisupervised linear discriminant analysis based on pairwise constraint

propagation for hyperspectral images [J]. IEEE Geoscience and Remote Sensing Letters, 2020, 18 (10): 1801–1805.

- [50] 朱梦远,杨红兵,李志伟,高光谱图像和叶绿素 含量的水稻纹枯病早期检测识别[J].光谱学 与光谱分析,2019,39(6):1898-1904.
- [51] 董元. 基于高光谱的小米产地溯源模型研究 [D]. 哈尔滨: 东北农业大学, 2021.
- [52] 张红涛,张亮,谭联,等.基于近红外高光谱成 像的单籽粒小麦品种分类研究[J]. 粮食与油 脂,2022,35(12):59-62.
- [53] 张立秀,张淑娟,孙海霞,等.高光谱技术结合网格搜索优化支持向量机的桃缺陷检测[J/OL].
 食品与发酵工业:1-10[2022-11-10].https://doi.org/10.13995/j.cnki.11-1802/ts.033557.
- [54] 刘姝妍. 基于高光谱成像的迷彩伪装识别研究 [D]. 太原:中北大学,2022.
- [55] 宗婧,黄志轩,陈晓宇,等.基于改进光谱角制图算法的乳粉安全非定向筛查新方法 [J].食品科学,2020,41(2):315-320.
- [56] 张子慧,李新国,李勇. 基于 CR-SAM 的博斯 腾湖西岸湖滨带典型植被分类 [J].中山大学 学报(自然科学版), 2022, 61(6): 36-43.
- [57] 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.
- [58] 高恒振,万建伟,王力宝,等.基于谱域-空域 组合核函数的高光谱图像分类技术研究[J]. 信号处理,2011,27(5):648-652.
- [59] 李娜,李咏洁,赵慧洁,等.基于光谱与空间 特征结合的改进高光谱数据分类算法[J].光 谱学与光谱分析,2014,34(2):526-531.
- [60] 王晓真. 基于级联马尔可夫随机场和自监督学 习的高光谱图像分类 [D]. 西安: 西安电子科 技大学, 2020.
- [61] 徐隆鑫,孙永华,吴文欢,等.基于无人机高 光谱影像的建筑垃圾分类研究[J].光谱学与 光谱分析,2022,42(12):3927-3934.
- [62] 杨阳, 韩芬, 杨娅婷, 等. 基于高光谱数据的

黄河流域水体提取方法研究 [J]. 绿色科技, 2022, **24**(22): 253-256.

- [63] 谢鹏,王正海,肖蓓,等. 基于海鸥算法优化随机森林的土壤硒含量高光谱反演 [J/OL]. 激光与光电子学进展: 1-16 [2022-11-01]. https://kns.cnki.net/kcms/detail/31.1690.TN.20221031.
 1650.152.html.
- [64] 肖树群,汪海城,袁兆宪,等.基于随机森林 与神经网络的高光谱土壤重金属 Zn 含量反演
 [J].环境监测管理与技术,2022,34(5): 22-26.
- [65] 于航,谭炳香,沈明潭,等. 基于机器学习算法的机载高光谱图像优势树种识别 [J/OL].
 自然资源遥感:1-10[2023-02-07]. https://kns.cnki.net/kcms/detail/10.1759.P.20230207.
 1408.006.html.
- [66] 司维. 基于机器学习框架的无人机高光谱遥感 水质参数反演对比研究 [D]. 武汉: 湖北大学, 2022.
- [67] 王春玲, 史锴源, 明星, 等. 基于机器学习的水体化学需氧量高光谱反演模型对比研究
 [J]. 光谱学与光谱分析, 2022, 42(8): 2353-2358.
- [68] 周少康. 基于无监督领域适应的高光谱图像分 类[D]. 北京:中国矿业大学, 2022.
- [69] 王倩. 基于深度神经网络的病理多样性显微高 光谱影像分析 [D]. 上海:华东师范大学, 2021.
- [70] 庞磊. 基于高光谱与深度学习的林木种子活力 检测及预测方法研究 [D]. 北京:北京林业大 学,2021.
- [71] 解玮. 基于注意力机制和胶囊网络的高光谱图 像分类 [D]. 西安: 西安电子科技大学, 2021.
- [72] 丁一民. 基于像素信息与协同表示高光谱图像 分类算法研究 [D]. 大连: 辽宁师范大学, 2022.
- [73] 刘爽, 谭鑫, 刘成玉, 等. 高光谱数据处理算 法的小麦赤霉病籽粒识别 [J]. 光谱学与光谱 分析, 2019, 39(11): 3540-3546.