



中国科学院大学

University of Chinese Academy of Sciences

博士学位论文

高光谱遥感植被半监督精细分类算法研究

作者姓名: 尚 坤

指导教师: 童庆禧 院 士 中国科学院遥感与数字地球研究所

张 霞 研究员 中国科学院遥感与数字地球研究所

学位类别: 理学博士

学科专业: 地图学与地理信息系统

培养单位: 中国科学院遥感与数字地球研究所

2014 年 5 月

Research on Semi-Supervised Classification Algorithm of
Detailed Vegetation Classification using Hyperspectral
Remote Sensing Data

By
Kun Shang

A Dissertation Submitted to
University of Chinese Academy of Sciences

In partial fulfillment of the requirement

For the degree of
Doctor of Philosophy

In
Cartography and Geographical Information System

Institute of Remote Sensing and Digital Earth,
Chinese Academy of Sciences

May, 2014

中国科学院遥感与数字地球研究所
学位论文原创性声明和使用授权说明

原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。论文中凡引用他人已经发表或未发表的成果、数据、观点等，均已明确注明出处。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。

本声明的法律责任由本人承担。

论文作者签名： 日期： 年 月 日

学位论文使用授权说明

本人在导师指导下完成的论文，知识产权归属中国科学院遥感与数字地球研究所。

本人完全了解中国科学院遥感与数字地球研究所关于收集、保存、使用学位论文的规定，同意研究所保存或向国家有关部门及机构送交论文的印刷版和电子版，允许论文被查阅和借阅；同意授权研究所采用影印、缩印、数字化或其它复制手段保存和汇编本学位论文。

本人保证毕业后，以本论文数据和资料发表论文（著作）或使用论文工作成果时，第一署名单位为中国科学院遥感与数字地球研究所，并征得导师同意。

保密论文在解密后遵守此规定。

论文作者签名： 导师签名：

日期： 年 月

本文研究受以下项目资助

遥感与数字地球研究所“一三五”规划突破三：建立“新型对地观测模拟系统”。

摘要

高光谱遥感影像一直是植被精细分类研究中重要的数据源，其“图谱合一”的特点使很多在多光谱数据中无法区分的类别变得可以区分，在作物品种识别、入侵物种监测、精准农业等方面具有广泛的应用前景。但是，高光谱遥感数据在植被精细分类中主要存在以下问题：第一，随着类别精细度的不断提高，单纯利用光谱信息得到的分类结果已经无法满足应用需求；第二，其成百上千的光谱波段带来了巨大的运算量，对监督分类中训练样本的需求也大幅增加；第三，随着传感器空间分辨率的提高，分类结果中椒盐噪声严重影响分类结果的应用。

近几年发展起来的半监督分类算法，对高光谱遥感数据的小样本分类具有重要意义。半监督分类算法同时利用了少量已标记样本和大量未标记样本，综合对整个数据进行分类。而目前的高光谱数据半监督分类算法大多只是单纯利用地物光谱的统计特征，从机器学习的角度实现半监督学习过程。为此，本文重点开展植被特征库的构建与优化算法和针对小样本的支持向量机半监督分类改进算法研究，提出改进的直推支持向量机与主动学习支持向量机分类算法，并分别采用具有较高空间分辨率的地面成像光谱数据与航空高光谱数据，进行植被精细分类实验，验证所提算法在地面尺度与航空尺度植被精细分类中的有效性。论文的主要研究成果和结论如下：

1) 提出了基于植被特征库构建与优化的植被分类策略：在考虑原有光谱特征、空间纹理特征的基础上，重点加入了对植被叶绿素、类胡萝卜素、花青素、氮素等含量变化较为敏感的 50 个光谱指数特征，有效的提高了不同植被类别间的可分性；

2) 提出了基于类对可分性（CPS）的植被特征库光谱维优化算法：该算法重点考虑不同植被类别在不同特征波段中的可分性，即“类对可分性”，在特征库光谱维优化中，分别保留对各类对具有最大 Bhattacharyya 距离的原始波段、纹理特征和光谱指数特征，基于选取的特征计算各类对 Jeffries-Matusita 距离，通过迭代使各类别均达到较高的分类精度，并利用最优索引因子法进一步降低波段间相关性。实验结果表明，该算法可有效降低数据冗余，提高分类效率；

3) 提出了基于邻域光谱角距离（NPSAD）的植被特征库空间维优化算法：

该算法的提出主要是考虑到地物分布通常具有一定的空间连续性, 相邻像元属于同一地物类别概率较高。在有训练样本的情况下, 该算法可以根据训练样本自动设置阈值; 否则, 需由用户根据先验知识设置阈值。从与其他空间维优化算法的对比试验来看, 该算法能够在去除分类结果中椒盐噪声的同时, 避免边缘效应, 保留细节信息, 提高分类精度。

4) 提出了基于光谱角距离-欧式距离双重判定的渐进直推支持向量机 (SAD/ED-PTSVM) 分类算法: 该算法在传统渐进直推支持向量机基础上, 通过计算光谱角距离与欧式距离分别对未标记样本标签进行判定, 并结合其到分划边界距离, 实现对未标记样本的“自动”标注。该算法有效利用了高光谱数据的光谱维信息, 降低未标记样本的误标记概率, 进而降低后期标签重置带来的时间成本, 并有效简化了传统直推支持向量机的参数设置, 减少参数优化所需时间, 提高分类效率。

5) 提出了结合混合像元分解进行主动学习的支持向量机 (SUAL-SVM) 分类算法: 该算法将高光谱研究中的混合像元分解技术与机器学习领域中的主动学习策略进行有机结合, 充分利用了各像元的丰度信息来辅助分类决策; 同时, 算法中设置了权重系数, 用来综合考虑“丰度最均匀”与“最易误分”像元。在主动学习样本添加时, 该算法按各类对区分难易度由多到少添加样本, 使样本添加更具有针对性。该算法可有效提高在影像中分布较少的类别的识别精度, 利用更少的标记样本实现更高的总体分类精度, 有效减少样本标记工作量, 减少分类器样本训练所需时间。

6) 地面与航空尺度的植被精细分类实验结果表明: 基于植被特征库构建与优化的分类算法可以有效的提高地面/航空尺度的不同种类作物/杂草的分类精度, 提取出更完整的叶片/作物地块; 而 SAD/ED-PTSVM 与 SUAL-SVM 分类算法, 在各类别仅有 25 个标记样本的情况下, 便可得到较高的分类精度, 有效提高了分类精度和分类效率, 并降低样本标注所需的人力物力。实验结果表明本文所提算法在地面与航空尺度的植被精细分类研究中具有广阔的应用前景。

关键词: 高光谱遥感, 植被精细分类, 植被特征库, 直推支持向量机, 主动学习

Abstract

As one of the most important data source of detailed vegetation classification (DVC), hyperspectral remote sensing (HRS) data make some classes distinguishable that cannot be separated in multispectral images, due to its advantage of imaging spectrum. HRS data therefore have widespread application prospects in identification of crop species, monitoring of invasive species and precision agriculture. However, there are also some problems of HRS data in DVC: Firstly, classification results based on only spectral information cannot meet the application requirements, as the classes become more and more sophisticated. Secondly, the increasing amount of spectral bands requires not only a huge number of calculations but also the corresponding growth of training samples in supervised classification. Thirdly, as the spatial resolution of hyperspectral sensor improves, the applications of classification are seriously affected by the salt and pepper noise in the classification results.

The semi-supervised classification algorithms developed in recent years play a key role in HRS classification with limited samples. The semi-supervised algorithms could perform the classification process by using a small amount of labeled samples and a large number of unlabeled samples. However, most of the current semi-supervised classification algorithms of HRS data only take advantage of the statistical characteristics of ground feature spectrum. On the basis of full investigation of current overseas and domestic research status, this study focuses on both vegetation feature band set construction and optimization and semi-supervised classification method based on support vector machine (SVM) for limited labeled samples, to overcome the problem that labeled samples are difficult to be acquired in hyperspectral data classification. Some classification experiments of field imaging spectral data and airborne hyperspectral data with high spatial resolution are also performed, in order to verify the effectiveness of the proposed methods of sophisticated vegetation classification both on ground and airborne scales.

The main results and conclusions of this study are presented as follows.

(i) A DVC strategy based on vegetation feature band set (FBS) construction and optimization is proposed: besides the spectral and texture feature of original images, we add 50 spectral indices that are sensitive to chlorophyll, carotenoid, anthocyanin and nitrogen content to the vegetation FBS. Results show that this strategy is able to effectively improve the separability between different vegetation classes.

(ii) A spectral dimension optimization algorithm of FBS based on class-pair separability (CPS) is proposed. This method focuses on the separability of different vegetation classes at different feature bands, i.e. CPS. It preserves the original bands, texture features and spectral index features respectively that have the largest Bhattacharyya distance of each CP through the iteration, and calculates the Jeffries-Matusita distance to make sure that each CP maintains a good separability through the spectral dimension optimization of the FBS. Then Optimum Index Factor is employed to reduce the feature bands with high correlation. This proposed method can reduce the redundant data and improve the classification efficiency.

(iii) A spatial dimension optimization algorithm of FBS based on neighborhood pixels' spectral angle distance (NPSAD) is proposed, considering that in general the probability of adjacent pixels being the same class is relatively high. This method can set thresholds automatically according to training samples if there are, otherwise users have to set thresholds based on prior knowledge. From the comparison of classification results, the proposed method could remove the salt and pepper noise from the classification results while avoiding the “edge effect” and keeping details at the same time, which can help to increase the classification accuracy.

(iv) A progressive transductive support vector machine method based on the discrimination of both Spectral Angle Distance and Euclidian distance (SAD/ED-PTSVM) is proposed. On the basis of traditional PTSVM method, this method discriminates the unlabeled samples by respectively calculating SAD and ED, and implements “automatic label” for the unlabeled samples according to their

distances to the separating hyperplane borders. SAD/ED-PTSVM makes good use of spectral information of HRS data, reducing the risk of incorrectly labeling and thus the time cost of label reset later. This method also effectively simplifies the parameter set of traditional PTSVM, reducing the amount of time spent on parameter optimization and thus improving classifying efficiency.

(v) A support vector machine classification method based on active learning using spectral unmixing technology (SUAL-SVM) is proposed. This method combines spectral unmixing technology in hyperspectral study and active learning strategy in machine learning field and fully utilizes the abundance of each pixel for better classification. Meanwhile, weighting factors are set to adjust the ratio of the most homogeneously mixed pixels and the most easily misclassified pixels. During active learning samples are added according to the distinguishing complexity of class-pairs, which makes the newly added samples more targeted. This method greatly improves identifying accuracy of the classes with small distribution, acquiring higher overall accuracy with less labeled samples and reducing not only the workload of labeling samples but also the time of samples training.

(vi) The results of detailed vegetation classification experiments on ground and airborne scales show that the classification method based on vegetation FBS construction and optimization can increase the classification accuracy of different types of crops/weeds on ground/airborne scale and extract more complete blade/crop plots information. SAD/ED-PTSVM and SUAL-SVM can effectively improve the classification accuracy and efficiency, reaching relatively high accuracy when there are only 25 labeled samples for each class. The results of experiments illustrate that the proposed methods in this study have great potential and broad application prospects in DVC, both on ground and airborne scales.

Keyword: hyperspectral remote sensing, detailed vegetation classification, vegetation feature band set, transductive support vector machine, active learning

目录

摘要	I
ABSTRACT	I
目录	I
第 1 章 引言	1
1.1 研究背景与选题依据	1
1.2 国内外研究现状	2
1.2.1 高光谱遥感技术的发展	3
1.2.2 基于高光谱遥感数据的植被分类研究进展	6
1.2.3 高光谱数据的特征提取和特征选择算法	8
1.2.4 基于多特征的高光谱遥感数据分类算法	10
1.2.5 基于高光谱遥感数据的半监督分类算法	13
1.3 研究内容和技术路线	15
1.3.1 研究内容	15
1.3.2 技术路线	17
1.4 论文组织结构	18
第 2 章 高光谱遥感数据植被精细分类算法分析	19
2.1 高光谱遥感分类算法概述	19
2.1.1 高光谱遥感数据监督分类方法	20
2.1.2 高光谱遥感数据半监督分类方法	27
2.2 高光谱遥感数据植被精细分类方法适宜性分析	30
2.2.1 高光谱遥感分类器选择	30
2.2.2 高光谱植被精细分类实验	31
2.2.3 植被精细分类算法适宜性分析	37
2.3 本章小结	38
第 3 章 针对高光谱植被精细分类的特征库构建与优化	39
3.1 针对高光谱植被精细分类的特征库构建	39
3.1.1 植被原始光谱特征	40

3.1.2	植被空间纹理特征.....	40
3.1.3	植被生化参量敏感的光谱指数特征.....	45
3.2	高光谱数据波段质量评价优化与特征选择	49
3.2.1	特征波段质量评价与优化.....	50
3.2.2	特征波段选择算法.....	51
3.3	基于类对可分性与邻域 SAD 的植被特征库优化算法	53
3.3.1	基于类对可分性的植被特征库光谱维优化.....	错误!未定义书签。
3.3.2	基于邻域 SAD 的植被特征库空间维优化	54
3.4	基于植被特征库构建与优化的分类实验	56
3.5	本章小结	61
第 4 章	基于 SVM 的高光谱植被半监督分类算法优化	63
4.1	支持向量机基本原理	63
4.1.1	支持向量机基本原理.....	64
4.1.2	支持向量机核函数及参数介绍.....	67
4.1.3	支持向量机参数优化策略.....	68
4.2	基于 SAD-ED 双重判定的渐进直推支持向量机分类	68
4.2.1	直推支持向量机原理.....	69
4.2.2	基于 SAD-ED 双重判定的渐进直推支持向量机.....	71
4.2.3	实验与结果.....	74
4.2.4	本节小结与讨论.....	76
4.3	结合混合像元分解进行主动学习的支持向量机分类.....	77
4.3.1	混合像元分解与主动学习基本原理.....	77
4.3.2	结合混合像元分解进行主动学习的支持向量机分类.....	79
4.3.3	实验与结果.....	82
4.3.4	本节小结与讨论.....	87
4.4	本章小结	87
第 5 章	基于地面成像光谱数据的作物/杂草精细分类.....	89
5.1	地面成像光谱数据预处理	89
5.2	植被特征库构建与优化	92
5.2.1	植被特征库构建.....	92
5.2.2	植被特征库优化.....	93

5.2.3	基于植被特征库构建与优化的地面高光谱数据分类.....	94
5.2.4	本节小结.....	97
5.3	基于支持向量机的半监督分类	97
5.3.1	基于 SAD-ED 双重判定的渐进直推支持向量机分类.....	97
5.3.2	结合混合像元分解进行主动学习的支持向量机分类.....	99
5.4	本章小结	101
第 6 章	基于航空高光谱遥感数据的植被精细分类	103
6.1	航空高光谱遥感数据预处理	103
6.2	植被特征库构建与优化	105
6.2.1	植被特征库构建.....	105
6.2.2	植被特征库优化.....	107
6.2.3	基于植被特征库构建与优化的航空高光谱数据分类.....	108
6.2.4	本节小结.....	110
6.3	基于支持向量机的半监督分类	111
6.3.1	基于 SAD-ED 双重判定的渐进直推支持向量机分类.....	111
6.3.2	结合混合像元分解进行主动学习的支持向量机分类.....	113
6.4	本章小结	115
第 7 章	结论和展望	117
7.1	本文主要成果与结论	117
7.2	本文创新点	120
7.3	不足与展望	121
参考文献.....		123
个人情况简介.....		141
致谢		145

第1章 引言

1.1 研究背景与选题依据

植被是全球生物多样性的的重要组成部分，尤其是供给我们基本食物和纤维的植被，它们是保障人们生产和生活的基本资源，以及成千上万的野生植物，无论是在经济上还是文化上，都具有重要价值。植被分类，从小的层面上讲，包括田间杂草识别和分类，是精准农业中的一项基础工作。从大的层面来说，农作物的精细分类是一项关系到国家粮食安全的重要工作。另外，森林树种监测、草地类型监测等对于生态多样性保护具有重要意义。

相对于传统的地面调查方式，遥感探测具有以下特点：宏观观测，可以大范围的获取研究区影像；动态监测，可以实时的更新研究区的数据；多时间、空间、光谱尺度探测，可以获取多尺度海量数据。目前，遥感已经成为植被精细分类中不可或缺的技术手段，一直是世界各国学者的研究热点。然而，传统的多光谱遥感数据由于光谱分辨率和通道数量有限，可获取到的地物可识别特征受到一定局限，造成相似作物之间难以区分、无法准确的统计出各作物类型面积等。

20 世纪 80 年代初发展起来的高光谱遥感成像技术，是遥感界的一场革命，在植被精细分类方面具有重要作用。高光谱遥感数据通常具有成百上千个几乎连续的光谱波段，光谱分辨率可以达到甚至优于 10nm。高光谱数据强大的光谱分辨率使得很多在传统多光谱数据中无法识别的特征变得可以识别，不可区分的植被类别变得可以区分，也使得更加准确的参数反演成为可能。目前，高光谱遥感技术已经广泛应用在作物品种识别、入侵物种监测、精准农业，以及作物病虫害、植被重金属污染监测等方面。

高光谱遥感数据带来了大量的有效信息，但与此同时，也带来了很多问题：不同程度的数据冗余，造成运算速度降低；很多针对多光谱图像的分类方法变得不再适用，或不能得到较高的精度；高光谱数据中部分信噪比较低的波段在分类时会带来严重的噪声；由于 Hughes 现象，多种基于统计的监督分类算法对训练样本需求大幅度增加，而获取大量标记样本需要复杂的地面调查实验做支持，这一定程度上增加了地面实验的复杂度。

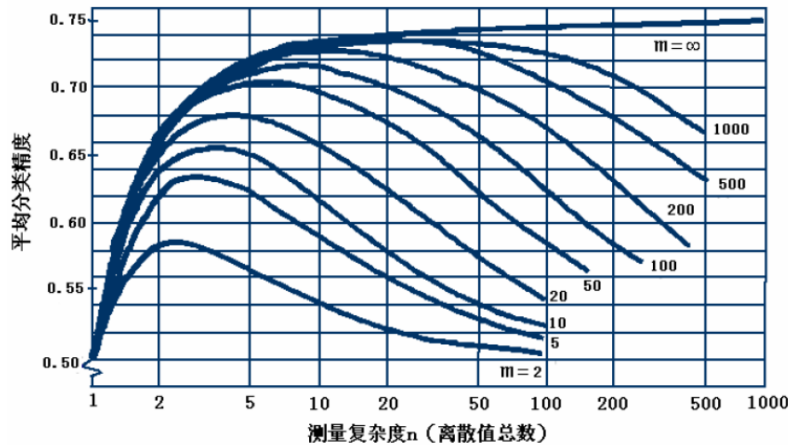


图 1.1.1 Hughes 现象(Hughes, 1968)

针对高光谱遥感植被精细分类中的主要问题,目前学者从不同角度尝试解决。首先是发展特征提取和特征选择算法,通过降低数据维度来减少数据冗余、降低分类器对训练样本需求,提高分类效率;其次,发展支持向量机等适应于高维数据小样本分类算法,或是基于专家知识的决策树分类算法,从统计学习或专家知识的角度对分类器进行改进,以提高分类精度和效率。近年来新发展起来的半监督分类算法,对于高光谱遥感数据的小样本分类有着重要意义。

半监督分类算法同时利用了少量的已标记样本和大量未标记样本的分布信息,综合对整个数据进行分类,半监督分类算法能够大幅度降低分类器对已标记样本的需求,每个类别只需要很少的标记样本,就可以得到很高的分类精度,在高光谱遥感分类中具有巨大的潜力。目前常用的高光谱数据半监督分类算法往往只是单纯利用地物光谱的统计特征,从机器学习的角度实现半监督学习过程,而没有从地物本身的特性出发,导致分类算法缺乏一定的机理性和针对性。

在航空和航天平台高光谱遥感技术迅速发展之际,地面成像光谱系统也取得长足发展。地面成像光谱遥感数据可以提供超高空间分辨率、受大气影响较小的成像光谱数据,得到较“纯”的像元信息,对于植被可以反映出更为精细的叶片空间、光谱特性。相对于航空、航天高光谱数据,地面成像光谱数据更适合进行机理性的算法研究。

本文利用航空和地面成像光谱遥感数据,充分挖掘植被光谱、空间、生理特征,研究综合多特征的高光谱遥感植被半监督分类算法,从机理上对已有算法进行改进,以更好的解决高光谱遥感植被精细分类问题,提高植被精细分类精度,

促进植被高光谱遥感分类技术的发展,推动高光谱遥感精细分类技术在农业、林业等各个领域的深入应用。

1.2 国内外研究现状

1.2.1 高光谱遥感技术的发展

1983年,世界第一台成像光谱仪 AIS-1 (Aero Imaging Spectrometer-1) 在美国喷气推进实验室 (JPL) 研制成功,并在地质、植被等众多研究领域取得了成功,显示了高光谱遥感技术的巨大潜力。其后,高光谱遥感技术得到了蓬勃发展,包括:美国的 AIS-2、GERIS、AVIRIS、MIVIS、DAIS-7915、HYDICE、Probe-1/2、TEEMS、SEBASS、PHILLS、WIS,加拿大的 FLI/PML、CASI、SASI、TASI、SFSI,澳大利亚的 GeosamMark II、Hymap,德国的 ROSIS-10 和 ROSIS-22,法国的 IMS,芬兰的 DAISA 等。其中以美国的 AVIRIS 影响最大,AVIRIS 是一台革命性的成像光谱仪,它极大地推动了成像光谱技术及应用的发展。加拿大的 CASI/SASI/TASI 分别在 $0.38\text{--}1.05\mu\text{m}$ 、 $0.95\text{--}2.45\mu\text{m}$ 和 $8\text{--}11.5\mu\text{m}$ 。澳大利亚的 Hymap 光谱范围为 $0.45\text{--}2.5\mu\text{m}$,具有 126 个波段,在 $3\text{--}5\mu\text{m}$ 和 $8\text{--}10\mu\text{m}$ 光谱范围内有两组共计 128 个可选波段,该传感器在很多方面均达到了较高的性能要求,如光谱定标、辐射定标和信噪比等。(童庆禧等, 2006)

航空成像光谱仪的飞速发展带动了航天成像光谱仪的发展。早期的航天成像光谱仪以美国的中分辨率成像光谱仪 MODIS 为代表。MODIS 成像光谱仪是 EOS-AM1 与 EOS-PM1 上的主要探测仪器,二者分别于 1999 年 12 月和 2002 年 5 月发射(张文杰, 2008)。美国的傅里叶变换高光谱 (FTHSI) 成像仪由 MightSat-II 搭载并于 2000 年 7 月升空,是干涉型成像光谱仪的代表。搭载在 Terra 卫星上的美日合作研制的 ASTER 成像光谱仪目前在反演陆面温度、反射率、比辐射率以及高程信息等方面应用较为广泛(张文杰, 2008)。美国宇航局发射的 EO-1 所搭载的 HYPERION 高光谱遥感器,是具有代表性的新一代航天成像光谱仪。欧空局 (European Space Agency) 于 2001 年成功发射了 PROBA 小卫星,其上搭载了多角度紧密型高分辨率成像光谱仪 CHRIS,可以提供五个角度的高光谱数据,共有 5 个成像模式,模式 1 主要是陆地与水成像模式,模式 2 为水成像,

3、4、5 则以陆地成像为主(张西雅等, 2010)。2005 年欧空局成功发射携带 PRISM 成像光谱仪的卫星, PRISM 有 223 个波段, 覆盖 0.45-2.35 μm 范围波长。近几年, 很多国家都在紧密筹划自己的高光谱传感器, 如德国的 EnMAP、加拿大的 HERO 和南非的多传感器小卫星成像仪 MSMI 均已有明确的发射计划(杨国鹏等, 2008)。

表 1.2.1 常见高光谱传感器参数 (Bioucas-Dias, 2013)

参数	HYDICE	AVIRIS	HYPERION	EnMAP	PRISMA	CHRIS	HysdIRI	IASI
轨道高度 /km	1.6	20	705	653	614	556	626	817
空间分辨率 /m	0.75	20	30	30	5-30	36	60	V: 1-2 H: 25
光谱分辨率 /nm	7-14	10	10	6.5-10	10	1.3-12	4-12	0.5 cm^{-1}
波段范围 / μm	0.4-2.5	0.4-2.5	0.4-2.5	0.4-2.5	0.4-2.5	0.4-1.0	0.38-2.5 7.5-12	3.62-15.5
波段数量	210	224	220	228	238	63	217	8461
数据立方体大小	200 \times 320	512 \times 614	660 \times 256	1000 \times 1000	400 \times 880	748 \times 748	620 \times 512	765 \times 120
sample \times lines \times bands	\times 210	\times 224	\times 220	\times 228	\times 238	\times 63	\times 210	\times 8461

我国时刻跟踪国际上高光谱技术的发展前沿, 在上世界 80 年代的中后期开始自行研发成像光谱系统。我国成像光谱系统的发展, 先后经历了从多波段到成像光谱扫描, 从光学机械扫描到面阵推扫的发展历程(童庆禧等, 2006)。考虑到海洋环境监测与森林火灾监测的应用需求, 研制了以紫外、红外及中长波红外为主的航空专用扫描仪(王楠楠, 2009)。80 年代中期, 结合地质勘探需求, 研制了覆盖 2.0-2.5 μm 波长范围与 8-12 μm 波长范围的 6-8 波段细分红外光谱扫描仪(FIMS)和航空热红外多光谱扫描仪(ATIMS)(雷彤, 2010)。80 年代后期我国研制成功了新型模块化航空成像光谱仪(MAIS), MAIS 在可见-近红外-短波红外具有 64 波段, 同时可以与热红外多光谱扫描仪进行集成, 使总波段数目达到 70-72 个(雷彤, 2010)。之后, 推扫式成像光谱仪(PHI)以及实用型模块化成像光谱仪(OMIS)也先后研制成功, 并多次在国内外实现成功应用。PHI 成像光谱仪具有 244 个波段, 光谱分辨率优于 5nm, 覆盖可见光到近红外光谱区; OMIS 分 OMIS-I 和 OMIS-II 两个型号, 前者具有 128 波段, 后者则具有 64 个波段(陈述彭等, 1998)。这些新的成像光谱系统除了在地质与固体地球等领域中继续发挥作用, 在城市地物及建筑材料分类、植被及农作物精细分类、生物地球化学效应研究中都显示除了巨大的应用潜力。

我国在航空成像光谱技术方面取得了一定成功,在此基础上,2002年3月我国载人航天计划中发射的第三艘试验飞船“神舟三号”上搭载了一台自行研制的中分辨率成像光谱仪,它具有34个波段覆盖可见光到热红外波长范围。2007年10月24日,在我国成功发射的嫦娥一号探月卫星上,搭载了我国首台基于傅里叶变换的航天干涉成像光谱仪,该成像光谱仪作为一种主要的载荷成功进入月球轨道。在2008年发射的环境与减灾小卫星(HJ-1)星座中,也搭载了一台高光谱成像仪,其具有128个波段、覆盖可见光至近红外光谱区,光谱分辨率优于5nm。2013年9月发射升空的风云3号气象卫星携带的中分辨率光谱成像仪(MERSI)波段范围0.40-12.5 μm ,通道数20,地面分辨率0.25-1km,用于探测海洋水色、气溶胶、水汽总量、云特性、地面特征、植被、冰雪、表面温度等。我国正在研制发展的晨昏轨道卫星的主要仪器载荷配置也包含红外高光谱大气探测仪。随着成像光谱系统在我国地质、植被、海洋、大气等各领域的广泛应用,标志着我国的成像光谱技术正在逐步走向成熟。

然而,航空、航天成像光谱数据空间分辨率通常超过1米,只能从宏观上监测和评价地物。为了深入地物目标的微观世界,增加目标解析率,许多发达国家的科研机构和公司正致力于地面成像光谱仪的研制,目前已成功推出了众多商业化产品,如Spectral Imaging Ltd公司的SisuCHEMA近红外高光谱分析仪、Resonon公司生产的Pika II和Pika NIR光谱仪和Surface Optics Corporation(SOC)公司生产的SOC700/720/730/740成像光谱仪。SisuCHEMA分为SisuCHEMA-SWIR和SisuCHEMA-NIR两种型号,光谱范围分别在0.97-2.5 μm 和0.9-1.7 μm ,最优光谱分辨率达6nm,空间分辨率在0.03-0.3mm。Pika II光谱仪的光谱范围在0.4-0.9 μm 有640个波段,光谱分辨率达2.1nm,Pika NIR光谱仪的光谱范围在0.9-1.7 μm ,有320个波段,光谱分辨率达5.5nm。SOC700/720/730/740成像光谱仪的光谱范围分别为0.43-0.89 μm 、0.9-1.7 μm 、0.4-0.9 μm 、0.8-2.5 μm ,波段数分别为120、120、240、120(刘波,2010)。

我国中科院遥感所和上海技物所在PHI的基础上,研制出了我国第一套地面成像光谱辐射测量系统(Field Imaging Spectrometer System, FISS),光谱范围覆盖可见光与近红外波段,有344个波段,光谱分辨率在4-7nm(童庆禧等,2010)。

相对于航天、航空成像光谱仪，地面成像光谱系统在入射光角度、光照强度、积分时间、地物类别选择、阴影遮盖、空间分辨率选择、混合象元模拟等方面具有更加灵活的可控性，也更适合进行机理性的算法研究。在我国，地面成像光谱数据已经开始逐步应用到多个领域。

1.2.2 基于高光谱遥感数据的植被分类研究进展

在植被的精细分类中，由于植物种类多且光谱相似，单纯依靠多光谱数据很多类别往往难以区分。而依靠高光谱数据的较高光谱分辨率和较多的波段信息，则能够更为全面、细致的获取不同种类植被的光谱特征及其差异性，可以大幅度提高植被遥感分类的类别精细度和准确度（李颖，2012）。航空/航天尺度的高光谱遥感数据可以用来进行植被识别，并对其时空变化进行动态监测；而地面尺度的高光谱数据则可以获得小块区域中详细的植被空间分布特征（刘波，2010）。

航空/航天高光谱遥感数据一直是植被精细分类的重要数据源。Nemani 等（1997）使用多时相 NOAA AVHRR 数据建立判别规则，将植被分为四种类型；Roberts 等（1998）利用 AVIRIS 高光谱数据，通过混合像元分解和光谱匹配的方法，对美国 Santa Monica 山区进行植被识别以及分类制图，取得了较高的精度；Hirano 等（2003）利用美国佛罗里达州 AVIRIS 高光谱遥感数据，对湿地植被进行了精细分类，研究显示高光谱遥感数据在入侵物种监测方面具有重要作用；Belluco 等（2006）比较了 ROSIS, CASI, MIVIS, IKONOS 和 QuickBird 数据在潮间带湿地植被分类中的作用，文中也比较了不同的分类算法，包括最大似然、光谱角度制图和 K 均值等；Zomer 等（2009）使用航空高光谱数据，建立湿地植被光谱库用于湿地植被的解译制图；Mitri 和 Gitas（2013）综合利用高光谱数据和高空间分辨率数据对火灾后的林地重建和林地恢复进行制图分析；Thorp 等（2013）基于 AVIRIS 数据利用多端元光谱混合分析方法，分析光谱和空间特性对半干旱区牧场植被制图的影响。

国内学者在高光谱数据植被精细分类中也进行了较多的研究。温兴平等（2008）使用 Hyperion 高光谱遥感数据，利用光谱角匹配的方法制作植被分布图；钱育蓉等（2011）使用点测量地面光谱仪获取的高光谱数据，利用 BP 神经

网络方法,对新疆阜康地区的7种典型荒漠草种进行了识别;韦玮等(2011)使用多角度高光谱 CHRIS 遥感影像数据,利用光谱角匹配方法,将湿地植被划分为小嵩草草甸、马先蒿沼泽、华扁穗沼泽化草甸、苔草草甸、华扁穗沼泽、针茅草甸、藏嵩草沼泽等7种类型。朱子先(2012)首先对高光谱和多光谱影像进行融合,然后采用一些监督、非监督分类方法(如 ISODATA 算法、K-means 算法、最大似然法、最小距离法以及光谱角填图法等)对研究区的植被覆盖进行划分,根据总体分类精度和 Kappa 系数比较分析了不同分类方法的分类精度。董连英(2013)利用 Hyperion 影像对森林植被进行分类方法和应用研究,研究结果表明,面向对象的决策树分类法在森林植被上的分类精度达到最高。

目前大部分针对植被精细分类的高光谱数据分类研究主要利用的地物光谱特征,或通过与其他高空间分辨率数据结合,综合利用空间特征和光谱特征来提高分类精度。

值得一提的是,贾坤等(2011)基于 Hyperion 高光谱影像,利用基于生化组分含量反演的决策树分类器,对罂粟、小麦、油葵三类植被进行区分,并与传统最大似然分类结果进行比较。研究表明,基于生化组分含量反演的分类方法在分类精度、运算速度均表现出良好的性能,该算法充分利用了植被生化特征,具有较强的机理性,在植被精细分类中有巨大的潜力。如何将植被生化组分含量与光谱、空间特征有效结合,是未来植被精细分类的一个重要研究方向。

相对于航天观测,航空高光谱遥感观测具有更加灵活的空间分辨率和成像时间,可以根据不同的目的对研究区进行按需观测,而地面成像光谱系统在精准农业中则有着广泛的应用,诸如作物杂草识别、变量施肥、作物长势监测、生化参量反演、作物病虫害监测等,因此成为农业中 site-specific 管理的重要技术手段和信息来源。大量美国、欧洲、日本的学者利用成像光谱技术进行作物杂草识别研究,或选择最佳工作波段,用以论证低成本的作物识别系统的开发。Borregaard 等(2000)在人工光照条件下,利用成像光谱仪对甜菜、马铃薯和三种杂草进行了识别。Feyaerts 和 Van Gool(2001)利用具有 35nm 光谱分辨率的多光谱成像系统,将杂草人工移植于机器视场范围内,获取各波段的影像进而实现甜菜与五种杂草的识别、分类。Virndtas 等(2002)利用成像光谱技术在室内和大田环境

下进行了杂草识别研究，用于除草剂的精准使用来减少环境污染和降低成本。Noble（2007）使用地面成像光谱仪数据，综合利用光谱信息与空间信息，在叶片分割的基础上进行植物识别。Mallahi（2008）则利用同一套系统实现了对土豆和土块的识别分类，指导开发机器视觉系统，用于研制农业自动化仪器对土豆进行自动收获。日本学者 Okamoto（2009）则利用成像光谱技术对未成熟的青桔子进行识别研究，该项技术可使水果估产在更早的阶段得以实现。这些研究表明，利用地面成像光谱数据进行的植被精细分类实验，具有可推广性，可以有效指导精准农业的实施，促进农业的智能化管理。

1.2.3 高光谱数据的特征提取和特征选择算法

高光谱遥感数据具有众多光谱波段，可以提供海量的遥感信息，这可以帮助我们实现更为精细的遥感地物分类以及目标识别，但是，遥感图像波段的增多也带来了不同程度的信息冗余、增加了数据处理复杂度（童庆禧等，2006）。特征提取和特征选择是高光谱遥感数据处理中最常用的降维手段，二者可以单独进行，也可以叠加进行，如先进行特征提取后进行特征选择，或先进行特征选择再进行特征提取等。提取出有效的地物特征，是后续地物精细识别和分类的重要前提。

目前常用的特征提取方法主要分光谱特征提取和空间特征提取两部分。

光谱特征提取的方法主要有基于主成分变换的特征提取、基于可分性准则的特征提取、判别分析特征提取、决策边界特征提取、和基于光谱重排的特征提取等（童庆禧等，2006）。Jia 和 Richards（1999）基于高光谱图像相邻波段间的相关性一般大于隔离开的波段间相关性的基本原理，发展了分块主成分分析方法，用来进行特征提取。该算法首先计算各波段间相关性，并按照相关性将原始波段分成一些高度相关的组，再对各组进行主成分变换。单波段可分离性分析，选择出各组中最重要的特征，再对这些特征进行后续的主成分分析与特征选择。Green 等（1988）发展了主成分分析方法，采用 MNF（Minimum Noise Fraction）变换使变换后各成份按照信噪比而不是方差从大到小的顺序来排列。直接利用原始波段的光谱信息来进行特征选择和特征提取，可以得到一些特征波段帮助识别不同类型地物。但有些时候，从地物的原始光谱很难发现可以帮助识别待分类地物的

特征。针对后一种情况，耿修瑞（2005）打破光谱按波长排列的次序，提出了光谱重排的方法，根据反射率或 DN 值重新排列各波段。其实验结果表明，不同地物的光谱在重排之后，总会出现一些显著的特征（不仅仅是幅度上的差别），而且对于不同地物类型，其特征出现的位置通常也不相同。

除了光谱特征变换，光谱特征参量化也是获取光谱特征的重要手段，典型的光谱特征参量包括光谱斜率和坡向、光谱二值编码、光谱吸收指数、植被倒高斯模型、光谱积分、光谱导数等，以及近年来新发展起来的光谱特征指数。钱育蓉等（2013）的研究显示不同草地类型和不同季相的荒漠植被，它们的红边、绿峰、红谷特征与植被指数均具有显著差异，但 NDVI 和 RVI 二者具有较高相关性。

空间特征包括地物的纹理特征和形状特征。目前常用的空间纹理特征计算方法有灰度共生矩阵（Grey Level Co-occurrence Matrix, GLCM）和 Anselin（1995）提出的局域指示空间分析（Local Indicators Spatial Association, LISA）。另外傅里叶变换、Gabor 变换、在特征提取研究中也具有重要作用。Bruce 等（2002）提出了二进离散小波变换方法来进行高光谱数据降维，并利用多种小波特征来解决地面植物的高光谱数据分类问题。

张连蓬（2003）建立了多方向投影寻踪算法，使原始数据的空间分布规律可以更精确地表现在低维空间中；并建立了主成分分析与投影寻踪相结合的特征组合方案，进而提高成像光谱数据的整体分类精度；同时，针对成像光谱数据的最优波段选择问题，提出了选元投影寻踪算法，并建立了一种快速选元投影寻踪算法。近年来新发展起来的流形学习算法（Roweis and Saul, 2000）在数据空间分布为非线性结构时具有优越性，Kim 和 Finkel 将流行学习方法与 K 均值聚类相结合，引入到高光谱数据降维中，取得了很好的结果。

目前研究中，对于高光谱数据特征提取的算法研究主要以统计学习为基础，从地物机理方面进行的研究较少，如何结合地物自身属性挖掘出其本质光谱、空间特征，对高光谱数据降维和分类具有重大意义。特征选择的目的是针对一个多种类别地物的组合，从光谱特征空间选择一个子集，简化原始的高维特征空间，并使该子集能够最大程度地区分该组合中的各类地物。特征选择的标准大体可分为四类：信息量准则、类间可分性准则、相关性准则和光谱学准则。其中，信息

量准则与类间可分性准则使用最为广泛。

信息量准则需确保选出的波段包含的信息丰富且相关性小,最优索引因子法 (Optimum Index Factor, OIF) 与协方差矩阵特征值法是其中有代表性的两种方法。根据类别可分性准则挑选特征一般应该包括三个步骤: 第一, 选择可分性准则; 第二, 确定使用可分性准则的策略; 第三, 确定选择特征的算法。类间可分性准则根据各待分类样本集在不同波段组合上的类间统计距离选择最优波段组合, 统计距离越大, 类间可分性越强。类间距离的多变量统计度量以离散度、B 距离 (Bhattacharyya distance) 和 J-M 距离 (Jeffries-Matusita distance) 为代表。

值得一提的是, Camps-Valls 等 (2010) 提出了一种基于核独立性估计的监督特征选择算法, 并分别以多光谱、高光谱和雷达数据进行实验, 结果表明, 该算法适用于多种类型遥感数据的特征选择。

1.2.4 基于多特征的高光谱遥感数据分类算法

传统的成像光谱数据分类算法主要分监督分类和非监督分类。经典的非监督分类方法有平行六面体法、聚类法、分裂法、动态聚类方法、K 均值算法和 ISODATA 算法等 (童庆禧等, 2006), 以及最近由 Tilton 发展的递归分层图像分割算法 (Recursive Hierarchical Image Segmentation, RHSEG)。但是利用这些分类方法无法获得各地物的实际类别, 而要根据地面实况调查和比较来决定, 通常被作为野外调查实验前的预分类过程。监督分类方法可以总结为两种思路: 一种是基于光谱相似性度量的分类方法, 主要是利用能够反映地物光学特性的光谱曲线来识别; 另外一种是基于图像统计特征或机器学习的分类方法, 主要是利用数据统计特性来创建一个分类模型。常用的监督分类方法主要有平行管道分类、最小距离分类、最大似然分类、神经网络分类、决策树分类、支持向量机分类、人工免疫系统分类、光谱匹配分类等 (张良培和张立福, 2011)。其中支持向量机分类方法由于其处理高维数据和小样本问题时的卓越性能, 是近年来高光谱分类研究的重点。Huang 等 (2002) 通过与最大似然、神经网络和决策树分类方法的对比试验来评价支持向量机分类算法的精度、稳定性和样本训练速度等, 同时也对核函数构成、训练样本选择以及其他输入变量对分类器的影响进行了分析。

然而，随着高光谱遥感数据在不同领域的广泛应用，用户对类别精细程度、精度的要求不断提高，传统的高光谱分类方法单纯依靠光谱信息，往往不能达到用户满意的分类精度（表 2）。Foody（2002）建议分类结果各类别精度应相对一致，且总体精度不低于 85%。越来越多的学者开始尝试综合光谱信息和空间信息对高光谱数据进行分类研究（Plaza 等，2009），研究表明综合利用光谱和空间信息可以显著提高分类器性能。Plaza 等（2004）提出了一种基于扩展形态学剖面的高光谱数据混合像元分类算法。Benediktsson 等（2005）提出了基于数学形态学剖面的高光谱数据分类算法，并成功应用到城市地区的航空高光谱数据实验中。Dalla Mura 等（2011）将扩展形态学属性剖面（Extended Multi-Attribute Profiles, EMAPs）与独立成分分析相结合，对高光谱影像进行分类。另一种利用空间信息的分类方法是在基于核的分类器中加入空间上下文信息，这也是一种重要的多特征分类方法。Camps-Valls 等（2006）提出了一种合成核（Composite Kernel, CK）高光谱图像分类框架，给出了构造核函数的方法与核函数特性，并通过实验验证了合成核分类方法在结合光谱信息和空间信息方面的有效性，同时可以灵活的调整光谱与空间信息所占比重，提高了计算效率。Velasco-Forero 和 Manian（2009）通过利用小波和偏微分方程的各向异性，同时结合合成核函数来进行空间预处理提高成像光谱数据分类精度，该算法有效降低了运算成本。Li 等（2013）给出了一种通用合成核（Generalized Composite Kernel, GCK）分类算法，通过利用多元线性回归（Multinomial Logistic Regression, MLR）和 EMAPs 来定义空间上下文信息，并给予 ROSIS 数据与多种分类算法进行比较，结果显示出通用合成核方法在结合空间与光谱信息方面具有巨大的潜力。马尔科夫随机场（Markov Random Fields, MRFs）是另一种经典的用来提取空间信息的模型。Tarabalka 等（2010）提出了一种基于支持向量机与 MRF 结合的高光谱数据精细分类方法，该方法主要分两步进行，首先利用基于象元的概率支持向量机分类方法对图像进行分类，然后利用空间上下文信息对分类结果进行改进。Li 等（2012）提出了一种基于子空间 MLR 与 MRF 结合的光谱-空间高光谱图像分类算法。其他光谱空间信息相结合的分类方法中有一些利用空间信息进行分类后处理改进分类精度，如 Kettig 和 Landgrebe（1976）提出的 ECHO（Extraction And

Classification Of Homogeneous Objects) 方法等。另外, Chen 等 (2011) 提出了一种针对高光谱数据的基于字典学习稀疏表达的分类方法, 其结果显示在很多情况下, 该方法可以得到优于传统支持向量机的分类结果。研究结果显示, 综合利用光谱和空间信息的分类结果基本可以满足应用需求 (表 1.2.3), 与只利用光谱信息的分类结果相比, 分类精度有大幅度提高 (表 1.2.2, 表 1.2.3)。

表 1.2.2 ROSIS 高光谱数据光谱分类精度 (%) (Bioucas-Dias, 2013)

METRIC	LDA	QDA	LogDA	SVM	MLR	MLR _{sub}
OA	77.95	77.95	78.41	80.99	80.11	67.08
AV	73.67	78.73	79.82	88.28	87.80	77.20
Kappa	0.606	0.770	0.720	0.761	0.750	0.703

表 1.2.3 ROSIS 高光谱数据光谱-空间结合分类精度 (%) (Bioucas-Dias, 2013)

METRIC	ECHO	LDA	QDA	LogDA	SVM	EMP	SVM	SVM	MLR	MLR _{sub}	MLR
		MLL	MLL	MLL	CK	SVM	W	RHSEG	MLL	MLL	GCK
OA	87.58	80.27	89.48	87.04	87.18	85.22	85.42	93.85	85.57	94.10	98.09
AV	92.16	78.05	91.91	83.32	90.47	90.76	91.31	97.07	92.54	93.45	97.76
kappa	0.839	0.739	0.864	0.872	0.871	0.808	0.813	0.918	0.818	0.922	0.974

近年来国内学者在综合空间和光谱信息的成像光谱数据分类方面也取得了一定进展。张钧萍和张晔 (2004) 提出了利用特征矢量来确定多分辨率融合权值的算法, 结果表明该算法可以有效降低原始数据维度并得到用于进行后续分类的特征波段。刘峰和龚健雅 (2009) 提出首先提取影像光谱、纹理及颜色特征, 获得多特征的高维特征空间, 并以互信息为评价指标进行特征选择, 然后在特征子集中应用支持向量机(SVM)实现图像监督分类的方法。苏红军等 (2008) 通过对同类曲线与不同类曲线特征分别求交, 得到识别某种地物类型的有效特征和区分不同类地物的有效特征, 然后根据提取出的特征构建地物识别分类决策树, 最终实现基于地物光谱特征的识别和分类。厉小润等 (2013) 提出一种利用合成核支持向量机 (SVM) 结合空间邻域信息和光谱信息的高光谱地物分类方法。

然而, 目前的多特征分类方法主要是结合光谱统计特征与空间特征, 而对于光谱曲线反映出的植被生理化学特征, 还没有很好的用于高光谱植被分类研究中。

1.2.5 基于高光谱遥感数据的半监督分类算法

高光谱数据监督分类时对训练样本的过高需求,给样本采集工作带来了很大的挑战。通常标记样本的获取费时费力,而未标记样本则可以很容易的获取到,因此,有必要发展高光谱数据的半监督分类算法。半监督分类算法认为新的训练样本可以利用已有的少量标记样本和大量未标记样本分布信息来获得。在进行半监督分类时应该注意以下几点:首先,新样本的获取方式应该尽可能简单;其次,获取的训练样本数目应该适当,太多会带来复杂的运算,太少会影响分类精度;另外,如果未标记样本选择不当,可能会带来误差降低分类精度。所以,如何选择信息量最大、对分类器最有帮助的样本,是半监督分类研究中的重要问题(Bioucas-Dias, 2013)。

近几年,高光谱遥感半监督分类算法得到了飞速发展。Bruzzone 等(2006)提出了基于直推支持向量机(Transductive Support Vector Machines, TSVMs)的遥感图像分类算法,通过综合利用已标记和未标记样本来搜索最优分划面。Camps-Valls(2007)提出了基于图论的高光谱影像半监督分类方法,综合利用了合成核函数和上下文信息。Gómez-Chova 等(2008)提出了一种基于核函数和图论的拉普拉斯支持向量机(Laplacian Support Vector Machine, LapSVM)半监督分类方法。Velasco-Forero 和 Manian(2009)首先对原始高光谱数据进行空间滤波,然后在基于图论的分类器中利用合成核函数,来提高高光谱数据的分类精度。Marconcini 等(2009)提出了一种综合利用光谱和空间信息的合成核支持向量机半监督分类方法,分类结果与传统监督支持向量机、单核半监督支持向量机、合成核监督支持向量机相比有显著提高。Tuia 和 Camps-Valls(2009)发展了利用集群内核的支持向量机遥感图像半监督分类算法。Ratle 等(2010)提出了基于神经网络的遥感图像半监督分类框架。Bruzzone 和 Persello(2009)给出了一种对误标记样本具有一定鲁棒性的新型上下文相关半监督支持向量机分类方法。Tuia 和 Camps-Valls(2011)给出了一种有效利用已标记样本和未标记样本的半监督分类方法,在训练支持向量机分类器时,利用2个核函数分别对已标记样本和未标记样本进行训练,再对两个核函数进行线性合成,并成功将该算法应用在高光谱数据城市分类中。高光谱图像半监督分类中常用的样本增量方法还有主动

学习策略。Li 等（2010, 2011）基于主动学习策略，分别利用多元逻辑回归和新型贝叶斯方法对高光谱影像进行半监督分类，得到了较高的精度。Tuia 等（2011）整理和测试了目前常用的主动学习策略，包括基于委员会的策略、基于最大间隔的策略和基于后验概率的策略，并讨论了各主动学习策略在遥感应用领域最新的研究进展。Di 和 Crawford（2011）给出了一种新型针对高光谱影像分类的主动学习协同正则化框架。Patra 和 Bruzzone（2012）针对一对多支持向量机多类分类时的不确定性问题，提出了一种新型的批处理模式主动学习策略。

除了典型的半监督分类算法，还有一些分类算法通过对监督分类和非监督聚类结果进行决策级融合来提高分类精度。如 Tarabalka 等（2009）提出基于像元的支持向量机分类与分割式聚类相结合的分类算法，在聚类时利用 ISODATA 与高斯混合模型，该方法得到的分类结果不仅能有效提高分类精度，还可以减少分类噪声。Yang 等（2010）提出了基于监督与非监督分类结果决策级融合的分类方法，利用支持向量机分类算法进行类别划分，利用 K 均值聚类方法同质区域降低细微的光谱差异，然后提出一种权重多数表决方法进行决策级融合。

国内对于高光谱遥感数据半监督分类方法的研究起步较晚，研究也相对较少。高恒振等（2011）根据聚类假设，首先利用基于光谱角度量的 K 均值聚类算法对样本集进行聚类，然后比较多次聚类的结果，进行包袋核函数的构造，再利用加法和乘法运算组合包袋核函数与 RBF 核函数得到新的核函数，从而把未标记样本的信息融入到分类器中。其高光谱实测数据实验结果表明，该算法在高光谱数据分类中具有一定优越性。高恒振等（2011）采用凹凸过程规划对直推支持向量机的非凸目标函数进行分解，得到凸函数和凹函数的组合，并将非凸问题转化为凸优化问题进行求解，考虑到不同波段对不同地物类别鉴别能力的差异，对不同的波段赋予不同的权重，从而改进直推支持向量机中的核函数，提高分类精度。姚伏天等（2012）提出用于高光谱图像分类的空间约束半监督高斯过程方法，通过对高斯过程模型中的核函数施加空间近邻约束，建立未标记样本与已标记样本之间的空间联系，提高仅有少量已标记训练样本情况下高光谱影像的分类精度。

1.3 研究内容和技术路线

1.3.1 研究内容

本研究以具有较高空间分辨率的航空/地面成像光谱数据为实验数据，主要针对在高光谱遥感植被精细分类中波段数量多、数据冗余大、训练样本需求大、很多植被特征未被有效利用等问题，构建适用于高光谱遥感数据植被精细分类的原始光谱、空间纹理和反映植被生化参量的光谱指数特征集，并提出植被特征库的光谱维与空间维优化算法。同时，考虑到大多数情况下可获得的标记样本较少，本研究基于目前新发展起来的半监督分类算法，从高光谱数据“图谱合一”特点出发，着重研究基于支持向量机的直推学习、主动学习分类算法，力求在训练样本有限的情况下，尽可能提高分类结果精度，为成像光谱数据在农作物精细分类、林地/草地精细识别、作物/杂草识别等方面的广泛应用提供技术支撑。

（1）针对高光谱遥感植被精细分类的特征库构建

目前成像光谱数据植被分类常用的特征主要有光谱特征和空间特征，其中光谱特征包括诊断性波段、波段计算和波段变换得到的特征波段等，空间特征主要有纹理特征和几何形状特征等。本文利用具有高空间分辨率的成像光谱数据，在挖掘植被的空间、光谱特征的基础上，重点研究对植被叶绿素、胡萝卜素、花青素、氮素、木质素和水分含量敏感的光谱指数特征，初步构建适用于高光谱数据植被精细分类的特征库，以提高不同植被类别间的可分性。

（2）针对高光谱遥感植被精细分类的特征库优化

对于植被特征库的优化，本文主要从降低数据冗余、提高分类效率与减少分类结果中的椒盐噪声、提高分类图像平滑度两个方面出发，重点植被特征库的光谱维优化与空间维优化。目前常用的特征选择算法通常是选择出信息量最大的波段组合，并不考虑实际要区分的植被类别，为此，本文主要以植被的精细分类识别为目的，研究最有利于提高类别可分性植被特征库光谱维优化算法，在提高总体分类精度的同时，使各个类别均能实现较好的区分。对于分类结果中的椒盐噪声，传统的去除方法是通过窗口滤波算法，但是这类算法通常会造成分类结果中的边缘信息丢失，对于某些重要的细节信息也无法保持，为此，本文将利用高光

谱数据的光谱维信息来帮助空间维平滑,进而提出能够避免边缘效应、保留一定细节信息的植被特征库空间维优化算法。

(3) 基于直推支持向量机的高光谱遥感数据半监督分类

传统的直推支持向量机主要是利用有限的标记样本,同时结合大量未标记样本分布情况,来对成像光谱数据进行分类。该算法可以有效的解决成像光谱遥感数据分类中常见的 Hughes 现象,但其分类结果精度受多种因素影响,如对总体样本类别分布的估计情况、未标记样本的标签添加算法等。

针对以上问题,本文在深入理解直推支持向量机原理基础上,充分挖掘高光谱数据的光谱信息,利用光谱信息来提高未标记样本的标签添加的准确度,以提高分类的精度和后期样本重置所需时间,并分析不同数量的初始标记样本和未标记样本对分类精度的影响,并与其他算法进行比较,最终提出一种改进的针对高光谱遥感数据的直推支持向量机分类算法。

(4) 基于主动学习的高光谱遥感数据半监督分类

成像光谱遥感数据分类通常需要用户选择大量的训练样本,而普通用户根据目视判读选择出的训练样本中存在很大程度的数据冗余,这些样本中大部分对分类模型的建立贡献度很小,而对大量样本的标注不仅费时费力,在分类器模型进行训练时也会大幅度增加模型运算时间。如何能够通过对少量样本的标注得到较高的分类精度,是成像光谱遥感数据分类的一个重要问题。而通过主动学习,分类器主动选择信息量最大的样本点,而不是随机选择,然后交由用户进行类别标注,再把这些样本点添加到训练样本中重新训练分类器。

在高光谱遥感数据处理中,混合像元分解与图像分类一直是重要的研究内容,混合像元分解技术可以给出各个像元的各组分含量,这对于改善图像分类效果具有重要意义。而目前的研究中,结合混合像元分解技术进行高光谱图像分类还不多见。为此,本文将重点研究结合混合像元分解进行主动学习的支持向量机分类算法,并分析初始标记样本数量、后续添加标记样本数量、迭代次数等对分类精度的影响。

1.3.2技术路线

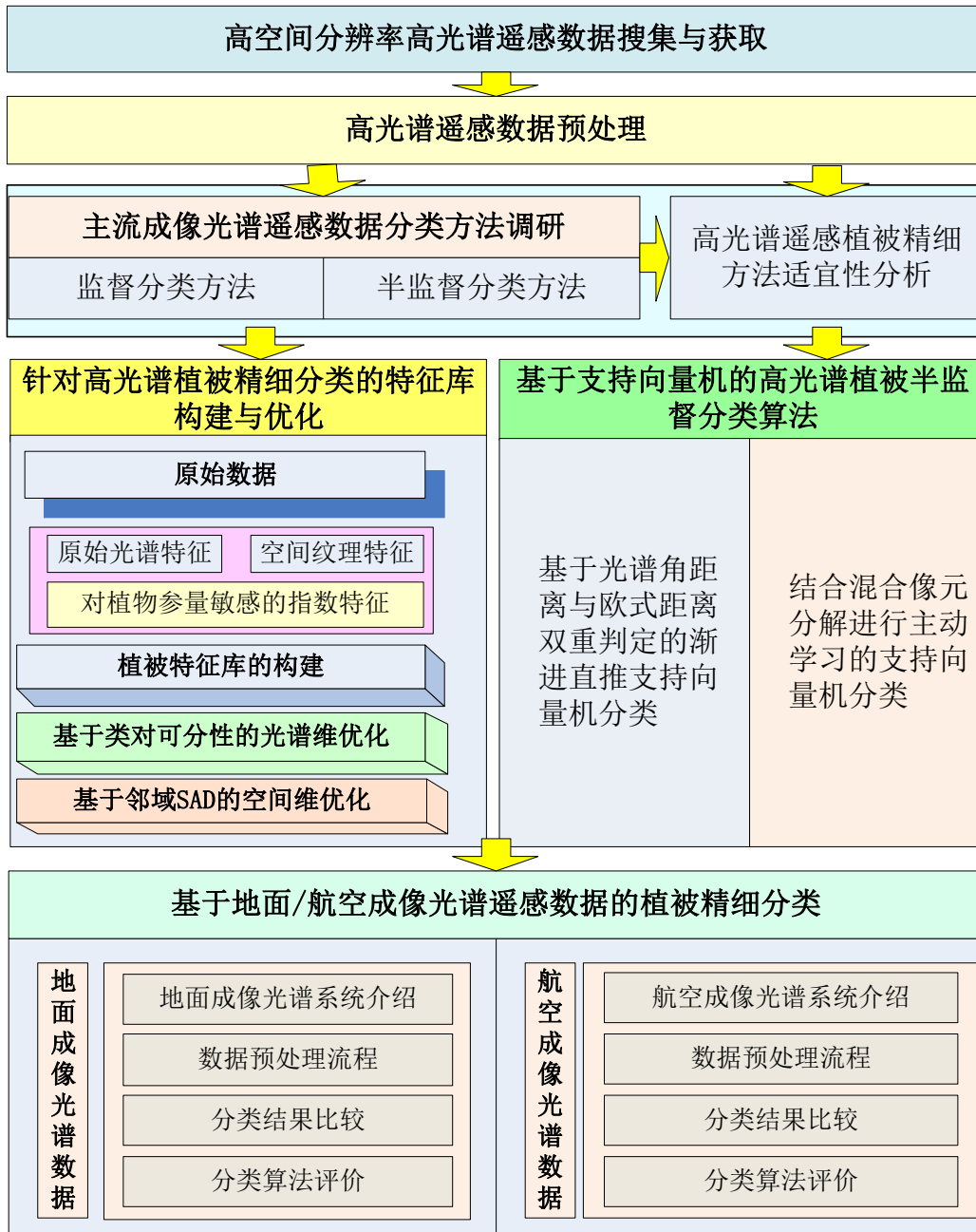


图 1.3.1 总体技术路线

根据本文的研究目标和研究内容,针对研究内容中存在的科学问题,本文的总体实验方案设计如图 1.3.1。首先,进行地面/航空高光谱遥感数据的搜集和获取,针对数据特点对数据进行预处理操作;其次,对成像光谱遥感数据的常用分类方法进行回顾和总结,并对各方法在植被精细分类方面的适宜性进行分析;接

着,根据原始光谱数据进行空间纹理信息的提取和光谱指数计算,完成植被特征库的构建,并采用基于类对可分性的光谱维优化算法和基于邻域 SAD 的空间维优化算法对植被特征库进行优化;然后,研究基于支持向量机的高光谱植被半监督分类算法,重点研究基于 SAD-ED 双重判定的渐进直推支持向量机分类算法和结合混合像元分解进行主动学习的支持向量机分类算法,并对两种算法分类结果的影响因素进行分析;最后分别利用地面和航空成像光谱遥感数据进行植被精细分类实验,验证所提方法的有效性。

1.4 论文组织结构

本论文共分七章,结构组织如下:

第一章,主要介绍本文的研究背景与意义,以及研究领域的现状,并由此引出本文的研究内容。

第二章,从监督分类、非监督聚类 and 半监督分类三方面,对目前高光谱遥感数据分类算法进行介绍,并在此基础上,对植被精细分类进行算法适宜性分析。

第三章,提出针对高光谱遥感植被精细分类的特征库构建方法,包括原始光谱特征、空间纹理特征和光谱指数特征三方面,同时在介绍常用波段质量评价与优化、特征选择算法基础上,提出针对光谱维和空间维的特征库优化策略,并以测试数据进行实验,验证所提方法在降维、去噪、提高分类精度方面的有效性。

第四章,在对支持向量机和半监督算法基本原理进行介绍的基础上,提出两套改进分类算法:基于 SAD-ED 双重判定的渐进直推支持向量机分类算法与结合混合像元分解进行主动学习的支持向量机分类算法,并以测试数据进行实验,分析所提算法的性能。

第五章和第六章,分别利用地面和航空成像光谱数据进行作物/杂草分类与作物精细分类研究,通过本文所提算法与其他算法的横向比较,验证植被特征库构建与优化分类方法、以及在此基础上进行基于 SAD-ED 双重判定的渐进直推支持向量机分类算法与结合混合像元分解进行主动学习的支持向量机分类算法在地面和航空尺度植被精细分类中的有效性。

第七章,对全文进行总结,并展望未来的发展趋势。

第2章 高光谱遥感数据植被精细分类算法分析

高光谱遥感数据植被精细分类研究,是指针对植被进行不同类别、不同品种的精细区分,其研究核心在于高光谱遥感分类算法的研究。本章在对高光谱遥感分类算法进行梳理分析的基础上,评价不同类型高光谱分类算法在植被精细分类研究中的适用性。鉴于非监督聚类算法并不能确定聚类后各类别对应的真实地物,通常将其作为监督或半监督分类的辅助,故不在本章分析范围之内。

2.1 高光谱遥感分类算法概述

传统的机器学习算法主要分为有监督学习和无监督学习,有监督学习只使用标记的样本,无监督学习只利用未标记数据。而新发展起来的半监督学习,则综合利用了少量的已标记数据和大量的未标记数据。这三种机器学习算法对应于三种高光谱遥感分类算法。有监督学习对应监督分类,监督分类通过对已标记数据进行训练,得到分类器,再用分类器对新数据进行分类。无监督学习对应非监督聚类,通过最小化类内距离和最大化类间距离,实现相似个体聚集为一类。而半监督学习则对应于半监督分类。半监督分类首先在少量标记数据上训练,进而开拓大量未标记的数据,通过不断迭代,最终得到一个更准确的分类器,再利用这个分类器对新数据进行分类。高光谱遥感数据分类的一般流程可以概括如下图。

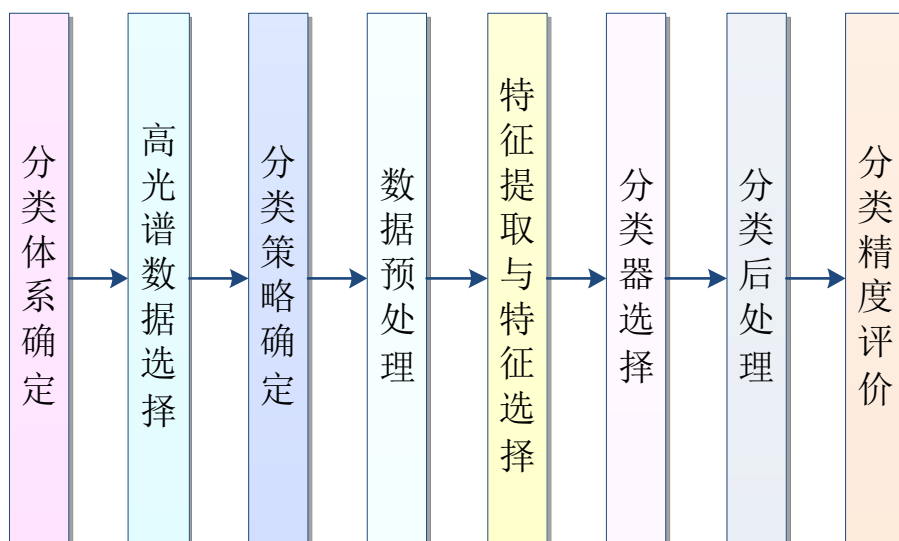


图 2.1.1 高光谱遥感数据分类流程图

2.1.1 高光谱遥感数据监督分类方法

监督分类方法常用于高光谱遥感数据的定量分析，它需要一定数量的训练样本，并根据训练样本对分类器进行训练，获取图像上各个类别像元的分类特征，再根据选择合适的分类判据、训练得到的分类准则对整景影像进行分类。传统的遥感监督分类算法主要有最大似然分类、最小距离分类、马氏距离分类等。传统的监督分类算法在应用到高光谱数据时，可以通过特征提取或特征选择降低高光谱数据维度，然后在低维空间直接利用这些传统的分类方法进行分类。

目前，应用在高光谱遥感数据中的分类算法大致可以分三类：一种是基于光谱相似性度量的分类方法，主要是利用反映地物反射、辐射光学性质的光谱曲线进行分类与识别，如光谱角度填图方法、光谱特征参量化方法等；一种是基于像元统计特征的分类方法，该类算法主要是建立在随机变量统计分析或机器学习的基础上，如最大似然分类、支持向量机、人工神经网络等；另一种则是基于规则的分类方法，如决策树，该类方法的核心是如何制定合适的分类规则。

2.1.1.1 基于光谱相似性度量的分类方法

基于光谱相似性度量的分类方法是针对高光谱数据最具特色的分类方法，主要是基于光谱特征匹配。该类算法可以利用整个光谱的波形特征对地物进行匹配分类，也可以只是利用对地物类型最具有诊断性的光谱特征或光谱特征参量（如吸收光谱的位置和吸收深度等）进行匹配分类。该类算法中比较常用的包括光谱角度填图、光谱波形匹配、二值编码匹配、基于包络线去除的图像分类等。

（1）光谱角度填图

光谱角度填图（Spectral Angle Mapper, SAM）将像元在 N 个波段上的光谱曲线作为 N 维特征空间中的矢量，通过计算该矢量与参考光谱在相应 N 维空间中的广义夹角来度量二者的相似程度，光谱角度距离越小，二者光谱特征越接近，是相同地物类别的可能性越大（Kruse 等，1993）。光谱角度距离（Spectral Angle Distance, SAD）定义为：

$$SAD(T, R) = \cos^{-1} \left(\frac{\sum_{i=1}^{nb} T_i R_i}{\left(\sum_{i=1}^{nb} T_i^2 \right)^{\frac{1}{2}} \left(\sum_{i=1}^{nb} R_i^2 \right)^{\frac{1}{2}}} \right) \quad (2-1-1)$$

上式中, nb 表示波段数量, T 与 R 分别是待分类像元光谱与参考光谱。当参考光谱与实验测量光谱或光谱库中光谱不匹配时, 需要将参考光谱按照高光谱图像中的波长进行重采样, 使参考光谱与像元光谱具有相同的维度。

可以看出, 光谱角度距离只与波形有关, 与相对数值关系不大, 因而对光照条件或反照率变化不敏感, 可以较好的应用在有阴影的影像中。但另一方面, 当高光谱数据的光谱维存在噪声时, 光谱角度的计算会受到一定影响, 且当光谱维数过高时, 特征波段的信息反而被减弱, 即地物的吸收谷、反射峰强度信息及位置信息被削弱 (李颖, 2012)。

(2) 光谱波形匹配

光谱波形匹配主要有两种: 一种是将参考光谱的全部或者部分进行光谱曲线的特征函数拟合, 通过计算像元光谱与参考光谱特征函数之间的拟合度来计算像元属于该地物类别的概率, 如植被倒高斯模型、光谱吸收谷的函数模拟等; 另外一种是直接计算参考光谱与各像元光谱之间的相似度 (张良培和张立福, 2011)。

光谱特征拟合是较常用的一种波形匹配算法, 该算法主要是基于地物吸收特征, 采用最小二乘法来计算各像元光谱吸收特征与参考光谱吸收特征之间的相似度。相似度可利用皮尔森相关系数进行衡量:

$$R_{x,y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2-1-2)$$

其中, x 和 y 分别为像元光谱与参考光谱, \bar{x} 和 \bar{y} 分别为 x 和 y 的平均光谱。

(3) 二值编码匹配

Goetz (1990) 针对高光谱数据存在的巨大冗余, 提出了一系列对光谱进行二进制编码的建议, 使得光谱可用简单用 0 和 1 进行二值表述。该方法可以在光谱库中对待分类像元进行快速的查找和匹配。最简单的二值编码规则如下:

$$\begin{cases} h(n) = 0, & \text{如果 } x(n) \leq T \\ h(n) = 1, & \text{如果 } x(n) > T \end{cases}, n=1,2,\dots,N \quad (2-1-3)$$

式中, $x(n)$ 是像元第 n 波段的亮度值, $h(n)$ 是其二值编码, T 是设置的阈值, 一般取光谱的平均亮度值。完成编码后, 便可利用基于最小距离等算法进行后续分类识别 (Jia 和 Richards, 1993)。二值编码匹配的分类方法可以有效的提高高光谱影像的分析处理效率, 但是该方法在处理编码的过程中会损失大量的光谱细节信息, 因而更适合对地物进行粗略的分类和识别。

(4) 基于包络线去除的图像分类

包络线去除法可以有效地突出光谱曲线中的反射与吸收特征, 并归一化到相同光谱背景上, 更利于不同光谱曲线之间进行特征位置比较, 进而提取出不同地物的特征波段供后续的分类识别。该算法通常只是应用在像元光谱分析、特征波段选择这一步骤中。

基于包络线去除的分类方法基本流程如下 (白继伟等, 2003): 首先利用包络线去除算法对原始高光谱影像进行处理, 得到新的图像文件, 突出光谱维的特征; 然后利用光谱分析的方法提取得到不同地物类型相应的特征波段; 最后利用光谱角度制图等方法, 对新的图像文件进行分类和识别。

基于包络线去除的图像分类方法主要有两个特点: 第一, 原始高光谱影像经过包络线去除后, 图像噪声得到了有效抑制, 地物光谱中的特征信息得到增强, 更有利于图像光谱的匹配, 进而提高图像分类运算的精度和效率; 第二, 对于某些原始光谱曲线相似、平缓的地物类型, 进行包络线去除后光谱特征更加相似, 而原始图像的灰度值又被忽略, 会导致分类精度的降低。

2.1.1.2 基于像元统计特征的分类方法

基于像元统计特征的分类方法在遥感图像分类研究中应用的最为广泛, 主要包括统计分析类的算法和机器学习类的算法。由于高光谱数据波段众多, 基于统计分析的分类算法在应用于高光谱数据时需要先进行数据降维, 最常用的是最大似然分类方法; 而基于机器学习的高光谱数据分类算法主要有人工神经网络、基于统计学习的决策树、支持向量机分类等方法。

(1) 最大似然分类

最大似然分类 (Maximum Likelihood Classification, MLC) 利用统计的方法建立一个判别函数集, 然后通过该判别函数集计算各个待分类像元对于各个类别的归属概率, 将像元划分到归属概率最大的一类中。最大似然分类方法假设遥感图像中各类别地物的概率密度服从多维正态分布。遥感图像中待分类像元属于第 i 类的概率可以表示为

$$p(x|w_i) = (2\pi)^{-N/2} |C_i|^{-1/2} \exp \left[-\frac{1}{2} (x - m_i)^T C_i^{-1} (x - m_i) \right], i = 1, 2, \dots, K \quad (2-1-4)$$

式中, x 为像元光谱, N 为波段数目, m_i 和 C_i 分别为第 i 类的均值向量和协方差矩阵, K 为类别数。该式的判别准则为: 对于 j ($j=1, 2, \dots, K; j \neq i$), 如果 $p(x|w_i) \cdot p(w_i) > p(x|w_j) \cdot p(w_j)$, 则 $x \in w_i$ 。式中 w_i 为第 i 个类别, $p(w_i)$ 和 $p(w_j)$ 分别为类别 i 与类别 j 的先验概率。

最大似然分类方法目前已经广泛的应用在多光谱遥感数据分类中, 对于高光谱遥感图像分类, 主要存在两个问题: 一是需要大量的训练样本, Richards(2013) 建议训练样本数目应该是波段数的 10~100 倍, 这对于某些分布较少的类别, 可能无法获取到足够的训练样本; 二是由于波段数的增多, 分类所花费的时间呈平方增长。

为此, Jia 和 Richards (1994) 提出了一种针对高光谱遥感数据的高效最大似然分类算法。由于高光谱数据相邻波段相关性较高, 该算法根据波段间的相关性, 将高光谱数据分成几个组, 由不同的波段组构成各类别的协方差矩阵, 然后从每个波段组计算出一个判别函数值, 再对所有波段组产生的函数值进行求和, 对每个像元进行分类, 即简化最大似然分类法。这种优化的最大似然方法, 各波段组内对每一类的训练样本数目要求比全部波段最大似然分类所需样本数目大大减少, 减少分类计算时间, 是一种比较有效的高光谱数据分类算法。

(2) 人工神经网络

人工智能神经网络技术 (Artificial Neural Network, ANN) 是利用计算机模拟人类学习过程, 建立起输入数据和输出数据之间联系的方法, 是模仿人脑学习的过程, 通过重复输入和输出训练来增强和修改输入和输出数据之间的联系 (赵英

时, 2003)。单个的人工神经元输入与输出之间的关系可以表示为:

$$y = f\left(\sum_{i=0}^{n-1} w_i x_i - \theta\right) \quad (2-1-5)$$

上式中, y 是处理单元的输出, f 是传递函数, x_i 是第 i 个输入元素, w_i 是第 i 个输入元素与处理单元间的连接权重, θ 是处理单元的内部阈值。传递函数一般是非线性的, 常用的有阈值型函数、阈值逻辑型函数和 S 型函数等。

人工神经网络是一个由大量人工神经元连接而形成的一个非线性的动力学系统, 具有自组织性、高维性、模糊性和冗余性等特点, 它一般包括一个节点数目与参与分类特征数相同的输入层、一个节点数目与类别数目相同的输出层和一个或多个的隐含层, 隐含层的节点数目由实验确定。目前常用于遥感图像分类的人工神经网络模型主要有: 自组织特征映射神经网络 (SOFM)、多层前馈型神经网络、对向传播神经网络、Hopfield 神经网络和自适应共振神经网络等 (Brandt 等, 2009)。

与传统分类器相比, 人工神经网络主要具有以下优点: 首先, 它是一种无参数分类器, 不需要预先对原始类别的概率分布进行假设; 其次, 它的输入与输出节点之间通过隐含层来链接, 多源数据可以通过节点设置权重来有效的融合到分类中, 提高分类能力; 另外, 由于人工神经网络是一个非线性系统, 可以使一些在原始空间中线性不可分的类别变得可以区分。

然而, 人工神经网络在高光谱遥感数据分类应用时也存在一定的问题, 如: 高光谱数据波段数众多造成的输入层维度高, 导致运行速度缓慢和难以收敛, 严重降低分类精度; 要达到一定的训练精度, 通常需要大量的迭代次数, 造成运行时间久, 分类效率低。针对这些问题, 通常的解决方法是在应用人工神经网络进行分类之前, 先进行高光谱数据降维, 以提高分类效率。

(3) 基于统计学习的决策树

决策树学习是机器学习领域应用最广的归纳推理算法之一, 对于噪声数据有很好的鲁棒性。它的原理是利用一种逼近离散值目标函数的方法学习得到最终判别函数, 并以决策树的形式表现出来。在决策树算法研究中, 如何构造精度高、规模小的决策树是其核心内容。

决策树的构造可以分两步进行：第一步是决策树的生成，指由训练样本生成决策树的过程；第二步是决策树的剪枝，这是在生成决策树的基础上进一步检验和修正决策树的过程，主要是利用新的样本集作为测试数据来校验决策树生成过程中构建的初步规则，将影响分类准确性的规则剪除。

与其他机器学习算法相比，决策树方法易于理解和实现，不需要使用者有很多的背景知识，可以很清晰的显示出哪些特征对分类决策更有帮助，同时数据准备简单，虽然决策树构建过程复杂，但一旦构建好决策树，可以在相对较短的时间内对大量数据进行分类，得到较好的分类结果。决策树方法也存在两个缺点：类别较多时错误增速可能较快；决策树分类算法是一种“贪心”搜索算法。因此，决策树算法生成的分类模型往往只能达到某种意义上的局部最优，而没有达到全局最优（杜培军等，2012）。

在遥感图像分类中，常用的决策树方法按照分裂规则的不同，可以分为基于信息论的方法和基于最小距离 GINI 指标方法，前者主要有 ID3、C4.5、C5.0 等（Quinlan, 1993），后者主要有 CART、PUBLIC、SLIC 和 SPRINT 等方法（Breiman 等，1984）。

（4）支持向量机

支持向量机（Support Vector Machines, SVM）作为新一代的统计学习分类方法，是上世纪 90 年代逐渐发展起来的，并被广泛应用于文本分类、手写字符识别、图像分类、生物序列分析等领域。近年来，SVM 被广泛应用于遥感图像的分类。

在遥感领域，SVM 主要应用于高光谱数据分类，并有很多成功的应用实例。与常规的监督分类方法如最大似然分类法不同，SVM 不需要数据一定要符合高斯分布，并且对休斯现象（Hughes Phenomenon）敏感程度也较低。另外，在多维数据分类中，SVM 能够用较少的训练样本得到比常规方法更高的分类精度。多数研究表明，支持向量机可获得了比传统的最大似然方法更高的精度。

支持向量机作为本文核心算法，其技术原理将在第四章第一节详细介绍。

2.1.1.3 基于规则的分类方法

基于规则的分类方法种类繁多,形式多样,但其核心均是依靠专家的先验知识,通过规则及其阈值的制定,实现地物类别的划分。常用的光谱特征主要包括:原始数据特征波段、原始光谱波形特征、对原始数据经过数学运算得到的特征波段、利用原始波段计算得到的特征指数、特征变换得到的特征波段等。常用的空间特征有面积、周长、形状指数、空间纹理特征(如灰度共生矩阵、局域指示空间分析、地统计学特征)等。

(1) 人工决策树分类

人工决策树分类(Decision Tree Classifier, DTC)是最简单也是最典型的基于规则的分类算法,它以设定的像元灰度值为基准,通过对图像中各像元按照设定的规则进行分层逐次比较,最终确定各像元的类别归属。决策树是一种可以循环的把数据集划分到子数据集的分类过程。决策树由根节点、中间节点与末端节点组成,每个中间节点有一个父节点和两个以上子节点。在划分过程中,通过中间节点比较总能分割成两组的叫做二叉决策树(Binary Decision Tree)。决策树分类方法对于数据分布没有要求,不需要分类器参数的设置(Hasen 等, 1996)。对于各中间节点的判别,可以采用阈值法也可以采用最大似然等算法(Borak, 1999; Tu 等, 1998)。

常用的决策树主要有单变决策树(Univariate Decision Trees)、多变决策树(Multivariate Decision Trees)、混合决策树(Hybrid Decision Tree)三种。单变决策树中的各中间节点都采用相同的特征和算法;多变决策树中的各中间节点虽然采用的算法相同,但是采用的特征不同;而混合决策树中的各个中间节点采用的算法和特征均不同,是三类决策树中分类能力最强也最有效的。

(2) 专家系统分类

专家系统(Expert System)分类是指把某一领域的专家知识输入到计算机中来辅助解决问题的系统,利用专家系统可以有效的把判读专家的经验知识进行综合以辅助分类(童庆禧等, 2006)。专家系统分类对于专家知识要求较高,在以分类为目的的遥感图像处理分析中,需要具备以下两方面的知识:一方面是关于

图像分析方法的知识,为了获得较好的结果,分类时需要把遥感图像处理分析方法通过适当的组合来加以利用,建立具有适用性分析方法的知识,并在适当的时候提示,提供图像处理的最佳步骤;另一方面是关于目标物的知识,在进行遥感图像的分类和目标识别时,对于目标物的各种知识是不可或缺的,这部分知识可以通过丰富的现场实践获取和修改。

人工智能技术与专家系统已经广泛的应用在高光谱遥感分类和自动识别研究中,如地质解译专家系统、矿物类型识别系统、植被光谱识别专家系统等。

2.1.2 高光谱遥感数据半监督分类方法

由于高光谱遥感数据对训练样本的需求大幅增加,而获取足够的训练样本往往需要较多的时间和精力,尤其是对于完全未知的研究区,地面调查需要消耗大量的人力物力。所以,如何能够在仅有少量标记样本的情况下,通过利用大量未标记样本来改善分类器性能,成为了高光谱数据分类的重要问题之一。高光谱半监督分类算法就是在这一背景下提出的。

在高光谱遥感半监督分类中,最常用的利用未标记的学习技术主要有直推学习(Transductive Learning)和主动学习(Active Learning)。这两类技术的基本思想有显著不同:前者主要是学习器自行使用未标记样本,它假设未标记样本是测试样本,而直推学习的目的就是在未标记样本上获得最佳泛化能力;而后者则是假定学习器对环境具有一定控制力,可以“主动”向学习器之外的某个“神谕”(Oracle)查询以此获得对训练样本的标记,如何能够利用较少的查询来获得更强的泛化能力是主动学习研究的技术难点。可以看出,直推学习是完全依靠学习器自身,而主动学习则需要与外界进行一定的交互。

2.1.2.1 直推学习策略

直推支持向量机(Transductive Support Vector Machine, TSVM)最早由Joachims(1999)提出并引入到文本分类中,有效的改善了分类性能。传统的支持向量机只是单纯利用标记样本进行分类,分类结果主要取决于这些标记样本的数量和质量。标记样本的获取通常费时费力,而未标记样本的获取则相对简单。

直推学习是机器学习领域处理未标记样本的一种有效手段,它通过对未标记样本的挖掘,将其与已标记样本综合,来构造一个更优的分类器。与主动学习相比,直推学习在利用未标记样本的过程中需要较少的人工干预。考虑到遥感图像覆盖区域广、标记样本获取困难,同时未标记样本丰富、相同植被类别具有一定相关性等特点,直推学习无论是理论上还是实践中,对于高光谱遥感植被精细分类都具有重要价值。

直推支持向量机是基于直推学习的一种半监督支持向量机,它通过综合已标记样本与未标记样本在核空间内寻求一个可靠的分划面。直推学习主要就是通过未标记样本的进行标注,并结合原有标记样本,利用最大间隔法找到一个高维空间的最优分划面。利用直推学习,可以有效提高支持向量机的分类性能,尤其是在训练样本很少或不足的情况下。

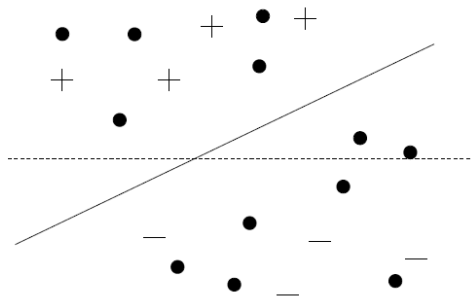


图 2.1.2 直推支持向量机分划线示意图

注: 图中‘+’为正类样本, ‘-’为负类样本, ‘•’为未标记样本。通过最大化间隔, 传统支持向量机得到虚线分划线, 而直推支持向量机可以得到实线分划线

目前,对于直推支持向量机的研究主要集中于如何得到一个经过核空间低密度区域的决策分划面 (Joachims, 1999; Chen 等, 2003; Bruzzone 等, 2006)。Joachims (1999) 解决了直推支持向量机中的二次优化问题, 实现了算法并在文本分类中进行应用。该算法在预先掌握样本中正负类比例的情况下, 具有较好的分类结果。Chapelle 和 Zien (2004) 通过利用梯度下降法来寻找核空间低密度区域中的决策边界, 进而优化直推支持向量机目标函数。Sindhwani 和 Keerthi (2006) 提出了一种适合大规模文本分类的快速线性直推支持向量机。Adankon 和 Cheriet (2007) 在标准直推支持向量机目标函数的基础上额外添加了规则, 并利用遗传算法来优化目标函数。

值得一提的是, Chen 等 (2003) 提出了渐进直推支持向量机 (Progressive Transductive Support Vector Machine, PTSVM)。该算法每次选择未标记样本集中最靠近决策边界的正负两个样本作为直推样本添加标签, 与其他添加过标签的直推样本和初始已标记样本一起构建最优分划面。该算法通过动态调整样本标签来降低前期误标记对分类精度的影响。Bruzzone 等 (2006) 修改了该算法, 实现批量添加正负类未标记样本。Singla 等 (2014) 提出了一种利用 KNN 与非监督来辅助未标记样本标注的渐进直推支持向量机分类算法。

2.1.2.2 主动学习策略

在主动学习中, 学习器可以主动要求用户对那些对于当前分类器模型最有价值的样本进行类别标记, 并把这些已添加标注的样本加入到训练样本集中, 对分类器进行重新训练, 通过迭代的方式对分类器进行更新。理论上, 在获得相似的分类精度的情况下, 相对于随机选择, 利用主动学习策略进行样本选择可以显著减少所需要的样本数 (Dasgupta, 2005)。

目前常用的主动学习策略主要有: 基于委员会的策略 ((Query by Committee)、基于最大间隔的策略 (Margin Sampling) 和基于后验概率的策略 (Posterior Probability-Based) (Tuia 等, 2011)。

主动学习的基本流程如下图所示。

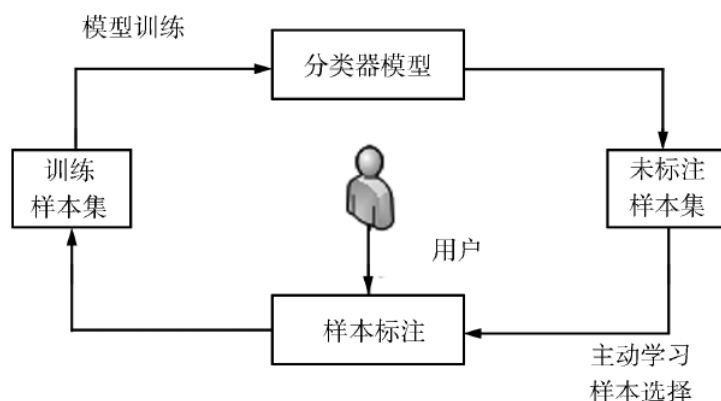


图 2.1.3 基于主动学习策略的分类框架 (陈荣等, 2011)

近几年, 基于主动学习的分类算法在遥感图像处理领域迅速发展。Rajan 等 (2008) 提出了一种主动学习策略, 可以利用较少的标记样本有效的提高现有分

类器性能,同时该方法对于已标记样本和未标记样本具有显著光谱差异的数据也有很好的学习和调整能力。Tuia 等(2009)提出了两套用于遥感图像分类训练样本半自动标记的主动学习算法,基于预先设置的递归循环,分类器可以对未标记样本进行排序并自动选择出最有利于分类器优化的样本。仅仅依靠初始时较少的非最优化的训练样本集,分类模型便可自行构造出使分类精度最高的最优样本集。Li 等(2010, 2011)基于主动学习策略,分别利用多元逻辑回归和新型贝叶斯方法对高光谱影像进行半监督分类,得到了较高的精度。Di 和 Crawford(2011)给出了一种新型针对高光谱影像分类的主动学习协同正则化框架。Pasolli 等(2011)提出了一种针对支持向量机的主动学习策略,它根据显著性理念将原始分类问题转化为一个需要判别样本显著与否的问题,通过构建一个显著性空间来适当指导潜在样本的选择,多光谱与高光谱图像的分类实验结果显示了该方法在收敛速度、稳定性、稀疏性等方面的优势。Demir 等(2011)研究了不同的批处理主动学习策略在支持向量机遥感图像分类中的作用,并提出了一种新型的查询函数,该函数基于一种新的用于评价样本差异的核聚类技术和一种用于筛选最优信息量和代表性样本的样本选择策略。Patra 和 Bruzzone(2012)针对一对多支持向量机多类分类时的不确定性问题,提出了一种新型的批处理模式主动学习策略。

2.2 高光谱遥感数据植被精细分类方法适宜性分析

2.2.1 高光谱遥感分类器选择

目前,高光谱遥感数据的分类算法数目繁多、种类多样,如何选择合适的分类器一直是高光谱分类中的重要问题之一。不同的分类器有不同的适应性,没有普适的、最优的分类器,针对不同的应用条件有不同的分类策略。

在实际进行植被精细分类应用时按照已标记样本数量可以按如下方式进行分类器选择:

(1) 已标记样本充足

针对原始波段空间:在已标记样本可以达到波段数目 10-100 倍的情况下、且各类别地物的概率密度服从多维正态分布时,可以直接采用最大似然分类算法,

该算法稳定、速度较快,但对噪声较敏感,更适合图像质量较高的遥感数据分类;同时也可以采用支持向量机等适应于高维空间的分类器,相对最大似然分类,支持向量机对样本分布要求低,且具有一定的抗噪性能。

针对降维后的特征空间:降维后波段数量一般会大幅度减少,对样本需求分类器性能要求都随之降低,通常可以直接采用传统基于统计学习的遥感图像监督分类算法,如最大似然分类算法和支持向量机算法,以及最小距离分类、人工神经网络等;同时也可以利用已标记样本提取各地物类别参考光谱,进而利用基于光谱相似性度量的分类算法,如光谱角度制图等。降维后由于波段众多带来的光谱度量不确定性随之减少,该类方法也可以得到较好的分类结果。

(2) 已标记样本有限

在已知地物参考光谱的情况下,可以根据已有光谱采用基于光谱相似性度量的分类算法,该类方法不仅可以给出各个像元的地物类别,还可以给出像元与各参考光谱的相似性程度,便于进行后续的分析处理。

在先验知识掌握充分的情况下,则可以采用基于规则的分类算法,通过人工设置一系列的判定得到最终分类结果,该类方法利用“If Then”模式可以对大数据进行简单快速的类别判别。

在无其他参考数据和先验知识的情况下,则可以充分利用半监督分类算法进行分类。半监督分类可以根据已有的少量标记样本来挖掘未标记样本的潜在标记,并将其加入到分类器来辅助分类。利用直推学习,分类器可以充分利用少量已标记样本和大量未标记样本分布信息来综合分类;而利用主动学习策略,分类器在仅有少量初始标记样本的情况下,自动搜索到对分类决策最有帮助的样本寻求标记,可以有效提高样本标记的质量,同时降低样本标记数量。研究表明,半监督分类算法可以有效的节省高光谱图像分类所需的人力物力,在高光谱遥感图像分类应用中具有巨大的潜力。

2.2.2 高光谱植被精细分类实验

为了更直观的比较不同情况下高光谱植被精细分类中各监督分类器的适应性,本文采用示例高光谱数据(如图 2.2.1)进行不同分类器在不同已标记样本

数量情况下的分类实验。

该数据为截取的 PHI 影像（真实影像及详细参数见第六章），具有 80 个波段，覆盖可见光与近红外波长，该数据是根据真实地物分布、对于 8 个植被类别分别随机生成 500 个样本点，通过重新组合得到的示例数据。示例数据总共包含 4000 个像元。

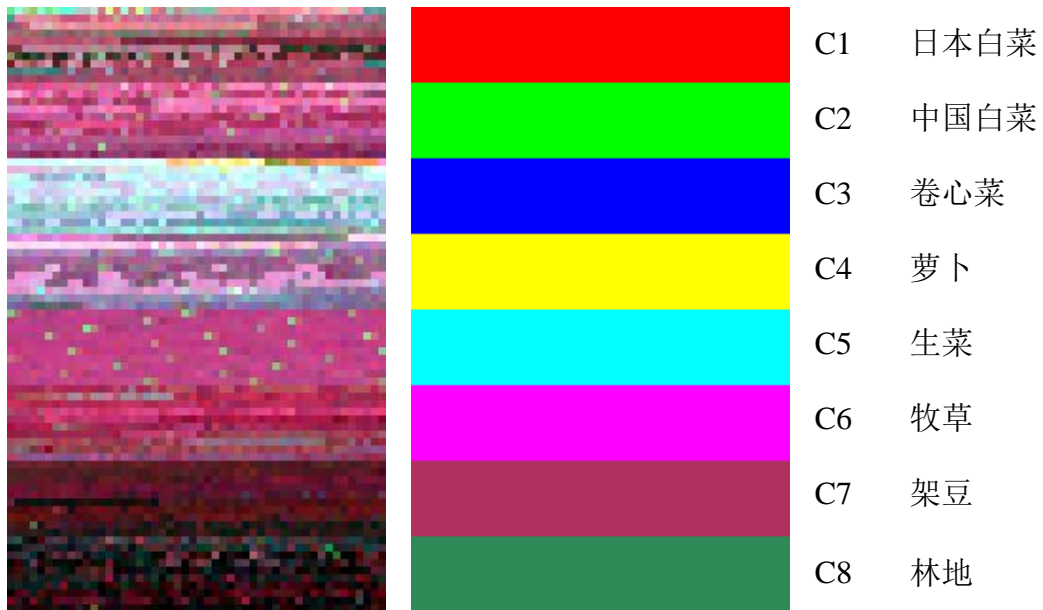


图 2.2.1 示例高光谱数据（R/G/B:832.8nm/650.8nm/553.7nm）与实际地物分布

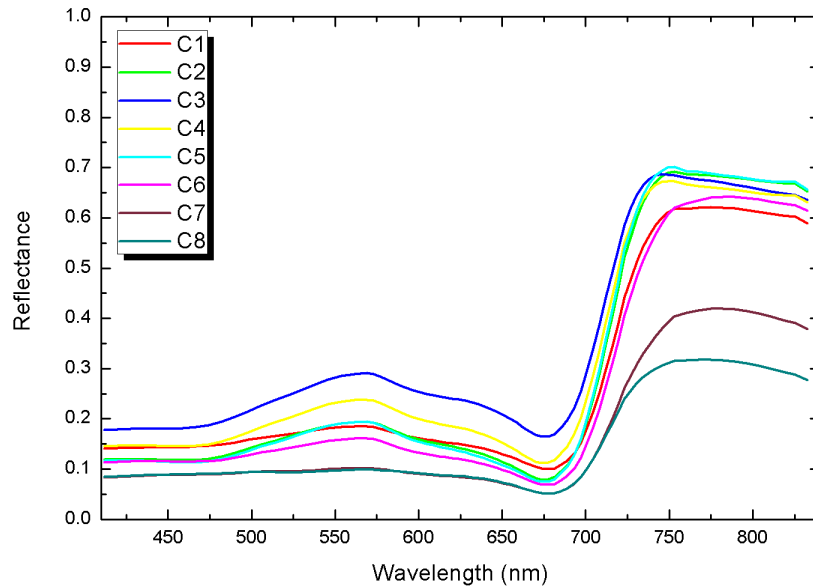


图 2.2.2 各植被类别光谱反射率曲线

注：为突出各植被类别光谱特征，本图光谱曲线已经 Savitzky-Golay 滤波处理，但后续实验中所用数据均为原始数据（Savitzky 和 Golay, 1964）。

在分类器选择方面,针对基于光谱相似性度量、基于像元统计特征和基于规则的三类分类器,本节分别选择光谱角度制图(SAM)、最大似然分类(MLC)和决策树(DTC)作为代表,同本文重点研究的支持向量机(SVM)分类算法进行比较试验。同时,为了测试各分类器对不同训练样本数量和特征空间维度的适应性,研究分别利用实际地物分布影像随机生成160个(原始波段数量2倍)和50个(降维后特征波段数量10倍)训练样本,基于ENVI5.0软件进行分类实验,所用参数均为软件默认参数。

- (1) 利用每类160个训练样本,在原始光谱空间直接进行分类,测试分类器在训练样本较充足的情况下针对高维空间的分类性能;
- (2) 利用每类50个训练样本,分别在原始光谱空间、降维后特征空间(主成分变换后前5个主成分组成的特征空间)进行分类,测试分类器在训练样本不足的情况下针对高维空间与降维后特征空间的分类性能。

分类结果的比较主要通过对分类图像的目视比较和精度统计比较两方面,其中精度比较包括总体精度(Overall Accuracy, OAA)和Kappa系数(Kappa Coefficient, KC)两部分。

2.2.2.1 训练样本较充足（每类 160 个训练样本）

(1) 原始空间

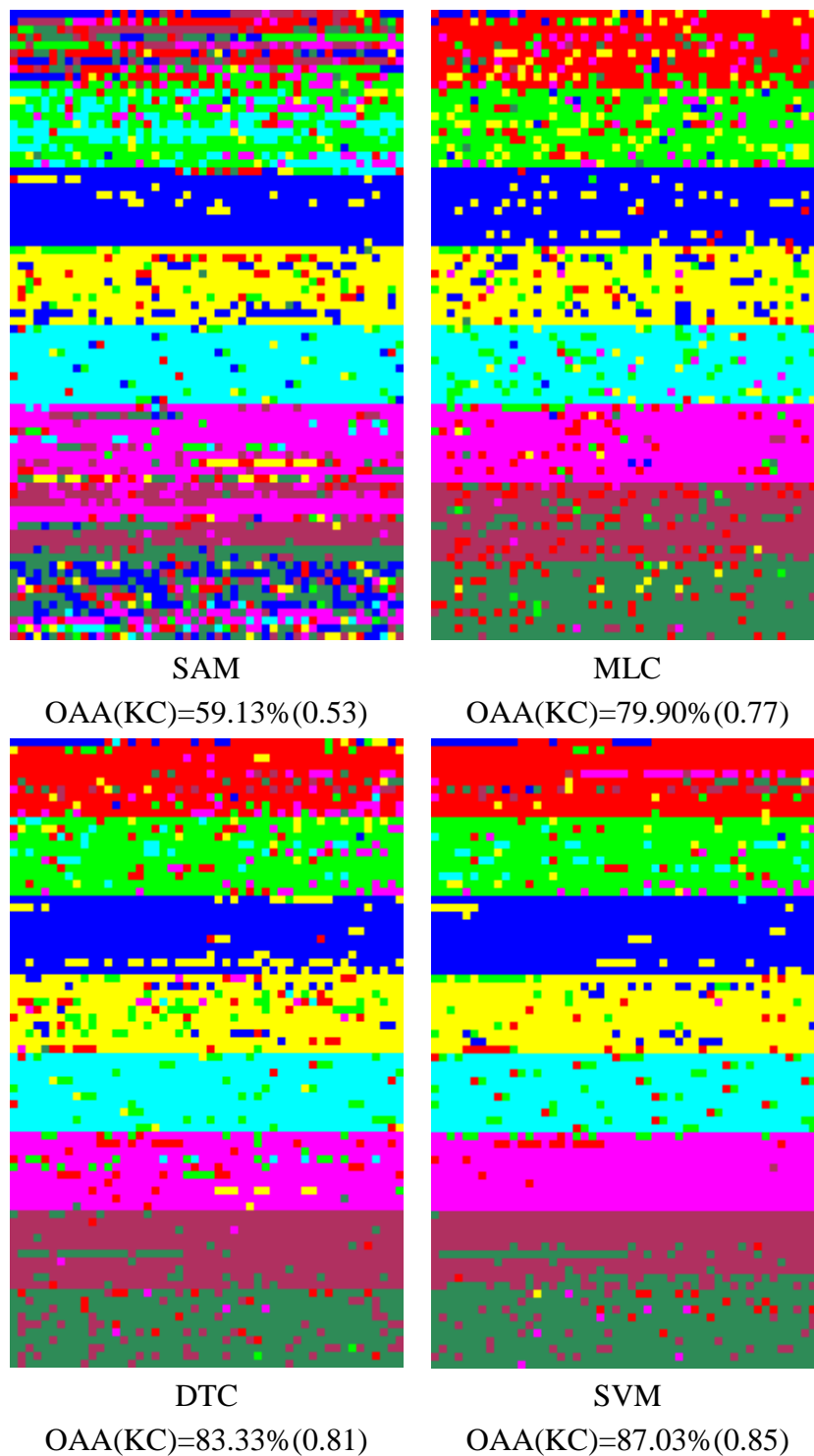


图 2.2.3 训练样本较充足情况下各分类器在原始空间分类结果

从上图可以看出，在训练样本较充足的情况下，SVM 分类器性能卓越，分

类准确度明显高于其他三种方法，噪声也相对较少；其次是 DTC 分类方法，OAA 和 KC 也均高于 80%；对于 MLC，虽然训练样本数量可以满足分类器运算需求，但依然未达到 10-100 倍，所以分类效果不尽人意；SAM 的分类精度很低，远不能满足应用需求。

2.2.2.2 训练样本不足（每类 50 个训练样本）

（1）原始空间

由于 MLC 要求训练样本数目一定大于待分类数据波段数目，所以本节只比较其他三种分类算法。

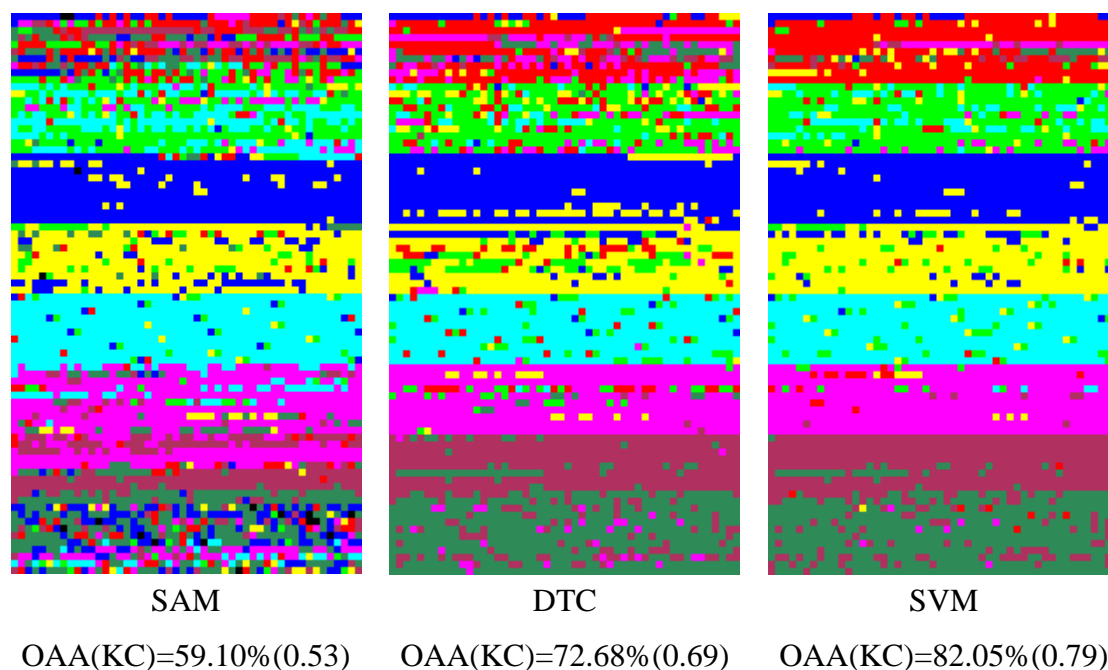


图 2.2.4 训练样本不足情况下各分类器在原始空间分类结果

从上图可以看出，在训练样本不足的情况下，与图 2.2.3 中相比，SAM 精度变化不大；DTC 精度大幅度下降，下降超过 10%，可见 DTC 对于高维空间的适应性仅限于训练样本充足的情况下；而 SVM 分类精度虽然也有所下降，但 OAA 依然达到 80% 以上，相对于其他分类算法具有很大优势。

(2) 降维后特征空间

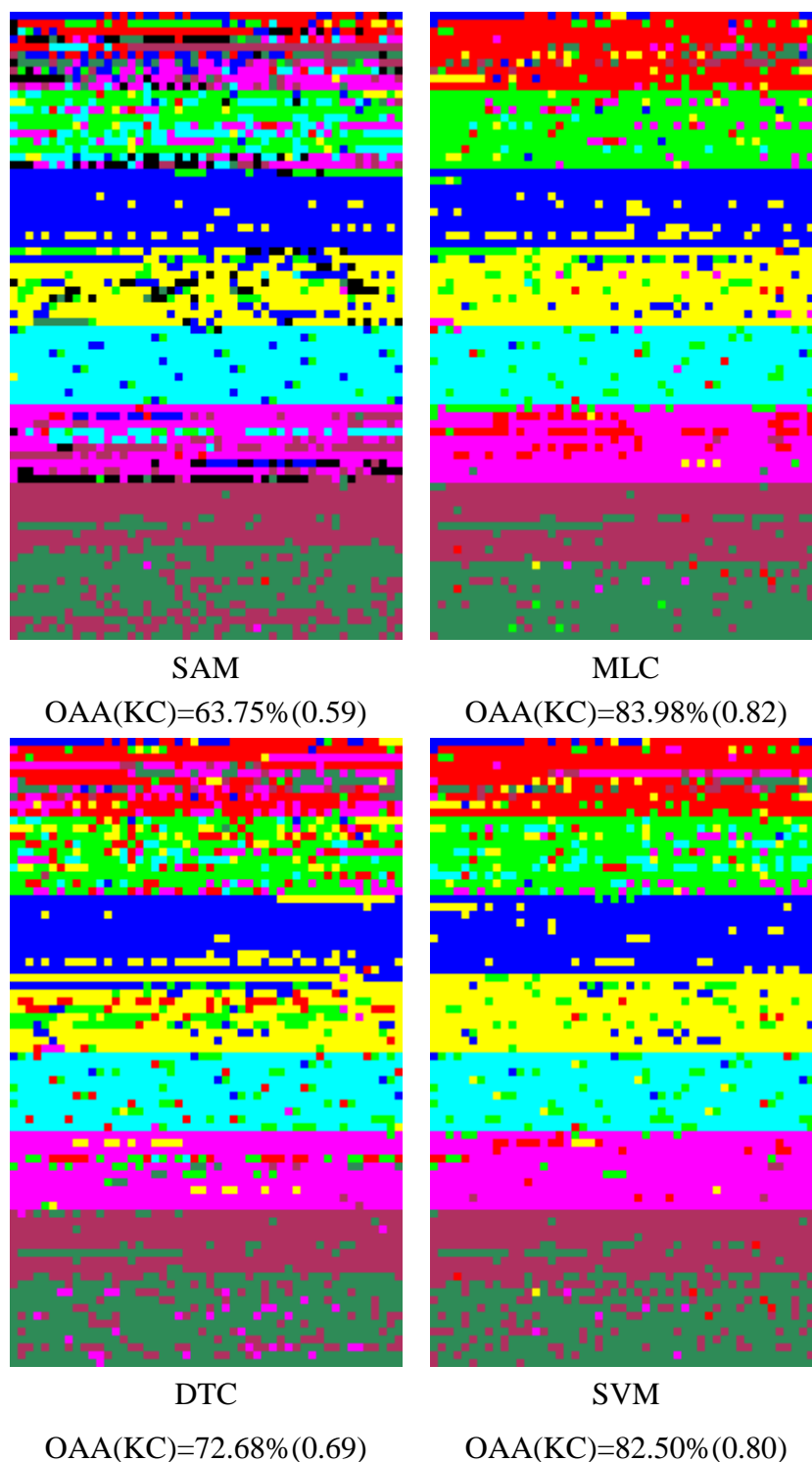


图 2.2.5 训练样本不足情况下各分类器在降维后特征空间分类结果

结合图 2.2.4 与图 2.2.5 可以看出,在训练样本不足的情况下,通过数据降维, SAM 精度稍有所提升; MLC 分类器则得以运行并能得到较好的分类结果,分类

精度高于图 2.2.3 中样本充足但空间维度过高情况下的精度；DTC 与 SVM 的分类精度变化不大。

2.2.3 植被精细分类算法适宜性分析

针对不同类型的监督分类器，通过 2.2.1 部分的理论分析与 2.2.2 部分的实验结果，我们可以得出以下结论：

- (1) SAM 分类器在进行高光谱数据分类时，通过降维可以一定程度提高分类精度，但由于相近植被类型光谱波形较为相似，而 SAM 只利用波形并未有效利用相对数值大小关系，导致其分类精度难以满足植被精细分类的应用需求。在实际应用中，可以考虑将其与基于统计特征的分类算法进行有机结合；
- (2) 对于 MLC 分类器，其分类结果对于样本数目与波段数比要求较高，在对高光谱数据进行分类时，针对原始空间通常难以保证各类别训练样本数目要求，所以分类效果并不好；但通过特征降维，在样本数目要求得以满足的情况下，MLC 可以有效的提高分类精度，并得到较好的分类结果；
- (3) DTC 在样本充足的情况下，可以进行高光谱数据分类，并得到较好的结果；在样本数目不足的情况下，分类精度大幅度下降，对于原始高维空间和降维后特征空间得到的分类结果接近，均无法满足应用需求。可见 DTC 更适合在样本充足情况下进行高维空间分类；
- (4) SVM 分类器无论是在样本充足的情况下，还是在样本不足的情况下，无论是针对原始光谱空间，还是针对降维后特征空间，总体分类精度始终在 80% 以上，Kappa 系数均在 0.79 以上，分类结果较为稳定，可以利用较少的样本直接对原始高光谱数据进行分类，对于大部分植被类别都具有较好的识别结果。

综上，在利用高光谱数据进行植被精细分类而无法满足样本数目与波段比的情况下，利用 SVM 分类器可以得到相对最优的分类结果。

2.3 本章小结

本章首先对常用的高光谱遥感数据分类方法进行了分析,按照对已标记训练样本和未标记训练样本的不同处理方法,将现有算法分为非监督聚类方法、监督分类方法以及半监督分类方法三类,并对后两类算法进行了详细介绍。对于监督分类算法,介绍了典型的基于光谱相似性度量分类方法、基于像元统计特征分类方法和基于规则的分类方法;对于半监督分类方法,主要介绍了直推支持向量机和主动学习策略两类经典算法。

由于不同的分类方法具有不同的适应性,本章从已标记训练样本数量、先验知识掌握情况、数据分类的特征空间维度等几方面,对不同分类器用于高光谱遥感植被精细分类的适应性进行了分析,给出了不同条件下的分类策略;同时以包含 8 个植被类别、具有 80 个光谱波段的 PHI 示例数据进行了实验,测试在不同训练样本数量、待分类数据维度情况下各分类器的性能;最后给出了不同分类器在高光谱数据植被精细分类中的适应性情况,并得出以下结论:

- (1) **SAM** 在高光谱植被精细分类时最好与基于统计特征的分类器结合,以弥补其未有效利用相对数值大小关系的劣势;
- (2) **MLC** 分类器对于样本数目与波段数比要求较高,在满足要求的情况下,可以得到较好的分类结果;
- (3) **DTC** 更适合样本充足情况下针对原始光谱空间直接分类,在样本数目不足时,无论降维与否,分类结果均无法满足应用需求;
- (4) **SVM** 分类器性能稳定,在不同训练样本数量和特征空间维度的情况下,均能得到较好的分类结果,对于大部分的植被类别也都有较好的识别能力,适用于高光谱数据的植被精细分类。

第3章 针对高光谱植被精细分类的特征库构建与优化

高光谱遥感数据具有纳米级的光谱分辨率、较为连续的光谱曲线,其中蕴含着丰富的光谱特征。同时,对于航空或地面等具有高空间分辨率的高光谱遥感数据,植被的空间纹理特征也可以有效应用在植被的分类和识别研究中。值得注意的是,通过光谱波段运算我们可以得到一些对植被生化参量敏感的指数特征,如果能够将光谱指数特征加入到植被精细分类中,不仅可以有效的提高植被识别分类精度,也能使分类模型更加具有机理性。

为此,本章旨在充分挖掘高光谱遥感数据中可用于植被精细识别和分类的有效特征,初步构建针对高光谱遥感数据分类的植被特征库,并在研究现有高光谱数据波段质量评价优化与特征选择算法的基础上,提出一套改进的高光谱遥感植被特征库优化策略。

3.1 针对高光谱植被精细分类的特征库构建

目前,高光谱遥感数据中常用于分类的特征主要有原始光谱特征和空间纹理特征。常用的空间纹理特征计算方法有灰度共生矩阵(Grey Level Co-occurrence Matrix, GLCM)和Anselin(1995)提出的局域指示空间分析(Local Indicators Spatial Association, LISA)。

另外,由于不同植被类型的生化组分含量差别较大,而同种植被类型在相近的水热条件及生长发育阶段中生化组分含量通常相近。因此,在植被的精细识别与分类时可以充分利用不同的生化组分信息。然而,植被的绝对生化组分含量反演过程通常较为复杂,而且在针对高光谱遥感影像进行植被分类时,往往并不需要绝对生化组分含量的精确反演,只需要了解不同植被间的相对差异即可。所以本章在构建高光谱植被精细分类的特征库时,并未选择植被的生化组分特征,而是选择了对植被生化参量敏感的光谱指数特征,有效降低了特征获取的复杂度。

本章针对高光谱植被精细分类构建的特征库主要包含三类特征:第一类是从原始高光谱数据可以直接获得的光谱波段特征;第二类是从高光谱图像上可以计算得到的植被空间纹理特征;第三类则是通过不同光谱波段间计算得到的、对植被生化参量敏感的光谱指数特征。

3.1.1 植被原始光谱特征

对于高光谱遥感数据，植被原始光谱特征是获取到的第一手资料，后续指数计算、纹理计算均是在原始光谱特征基础上进行的。

从模式识别的角度，在植被精细分类中利用灰度值、辐亮度值和光谱反射率数值差别不大；但是从高光谱遥感数据的角度，推荐用经过辐射校正、大气校正后得到的植被光谱反射率值进行分类。本节以图 2.2.1 中示例 PHI 数据进行实验，从图 3.1.1 可以看出，左图辐亮度曲线中，由于大气影响，存在过多的反射峰与吸收谷，并不利于植被有效特征的提取；而右图经过大气校正的反射率曲线中，植被的红边效应更加明显，反映植被特征的反射峰和吸收谷也更加突出。

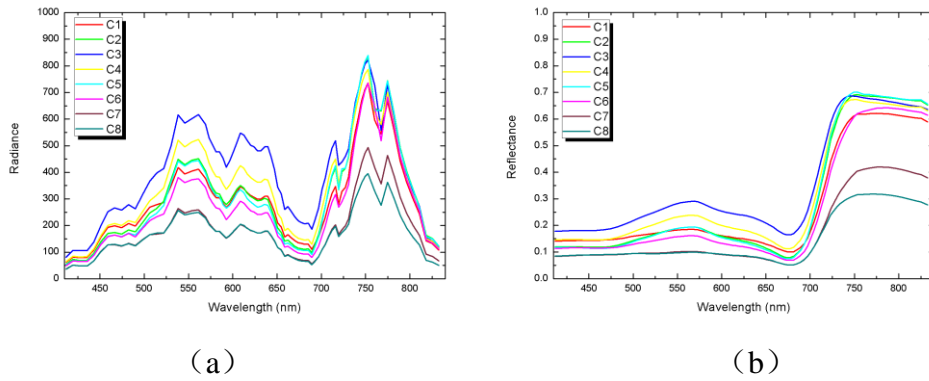


图 3.1.1 大气校正前后植被光谱辐亮度曲线 (a) 与光谱反射率曲线 (b) 对比。

为突出光谱曲线中植被特征，(b) 中光谱曲线已经 Savitzky-Golay 滤波处理。

3.1.2 植被空间纹理特征

高光谱遥感图像中地物光谱的不确定性可以通过加入空间和结构特征来抑制（承继成等，2004）。与其他地物类型相比，植被具有较为鲜明的空间纹理特征，无论是叶片尺度还是农田地块尺度，在空间分辨率适宜的遥感影像中均具有一定的纹理特征。因此在构建高光谱植被特征库时，加入基于灰度共生矩阵（Grey Level Co-occurrence Matrix, GLCM）和局域指示空间分析（Local Indicators Spatial Association, LISA）的纹理特征，有利于植被信息的准确提取。

灰度共生矩阵是遥感图像中最常用的纹理分析方法，它通过研究两种灰度级象元的联合分布情况得到图像灰度级的空间相关性规律（Haralick 等，1973）。

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。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，

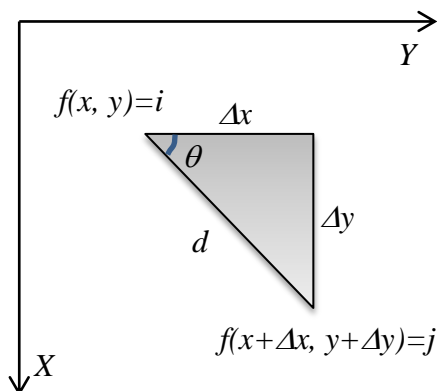


图 3.1.2 灰度共生矩阵中像元对

本章在构建植被特征库时重点考虑了均值 (Mean)、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：

$$f_{mean} = \frac{1}{L^2} \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P(i, j, d, \theta) \quad (3-1-1)$$

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：

$$f_{homo} = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \frac{1}{1 + (i - j)^2} P(i, j, d, \theta) \quad (3-1-2)$$

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：

$$f_{cont} = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i - j)^2 P(i, j, d, \theta) \quad (3-1-3)$$

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。计算公式如下：

$$f_{diss} = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} |i - j| P(i, j, d, \theta) \quad (3-1-4)$$

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：

$$f_{ent} = - \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P(i, j, d, \theta) \log P(i, j, d, \theta) \quad (3-1-5)$$

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。具体计算公式如下：

$$f_{sm} = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P(i, j, d, \theta)^2 \quad (3-1-6)$$

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：

$$f_{corr} = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \frac{ijP(i, j, d, \theta) - u_1 u_2}{\sigma_1^2 \sigma_2^2} \quad (3-1-7)$$

其中， $u_1 = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} iP(i, j, d, \theta)$ ， $u_2 = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} jP(i, j, d, \theta)$ ， $\sigma_1 = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i - u_1)^2 P(i, j, d, \theta)$ ，

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

除了基于灰度共生矩阵计算的纹理特征，局域指示空间分析(Local Indicators Spatial Association, LISA) 纹理特征在高空间分辨率遥感图像植被分类中也具有重要作用(Mallinis 等, 2008)。LISA 中最常用的有 Moran 指数 I (Anselin Local Moran's I)、Geary 系数 C (Anselin Local Geary's C) 和 G 统计量 (Getis-Ord Local Gi) 等 (Anselin, 1995; Getis and Ord, 1992)。空间自相关统计指标包括全局指标和局域指标。全局指标主要用于判断某属性值在空间上的整体分布情况，对于其聚集位置及区域相关程度则难以探测出；而局域指标主要用于反映一个局域小单元中某属性值与其相邻局域小单元中同一属性值的相关程度。所以，本章在进行空间自相关纹理特征提取时主要利用其局域指标。

Moran 指数 I 用来进行聚类 and 异常值分析，常用来指示像元间的相似情况，正值意味着两个单元中的属性相似度较高，负值则意味着两个单元间具有较高的差异性。具体计算公式如下：

$$I_i = \frac{(x_i - \bar{x})}{S^2} \sum_j w_{ij} (x_j - \bar{x}) \quad (3-1-8)$$

其中, $S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$, $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, w_{ij} 为空间权重矩阵。

Geary 系数 C 用于检测像元与其临近像元之间的差异性, 在边缘检测方面具有重要作用, 其计算公式如下:

$$C_i = \frac{(n-1) \sum_j w_{ij} (x_i - x_j)^2}{2(x_i - \bar{x})^2} \quad (3-1-9)$$

G 统计量用于进行热点分析, 可以识别出具有统计显著性的高值 (热点) 和低值 (冷点) 的空间聚类, 计算公式如下:

$$G_i = \sum_i w_{ij} x_j / \sum_j x_j \quad (3-1-10)$$

通常纹理特征是通过原始光谱波段计算得到的, 每一个原始波段可以得到一组相应的纹理特征波段。本章基于 GLCM 与 LISA 共选用了 10 个纹理特征, 如果针对原始高光谱数据的每一个光谱波段均计算其相应纹理特征, 则有 $10 \times N$ 个纹理特征 (N 为原始高光谱数据波段数)。对于 PHI 数据, $N=80$, 则 $10 \times N=800$ 个纹理特征; 而对于 FISS 数据, $N=344$, 则可得到 $10 \times N=3440$ 个纹理特征。如此多的纹理特征无论是前期的获取过程, 还是后续的分类过程, 运算量都是巨大的, 带来的时间成本也是可观的。

因此, 在进行纹理特征提取时, 首先对原始高光谱数据进行主成分变换, 并保留前 n 个使累积贡献率达到某一值 (例如 90%, 95% 或 98%) 以上的主成分, 之后对这 n 个主成分分别基于 GLCM 和 LISA 进行纹理波段计算。在实际分类应用中, 为节省运算量, 可先对训练样本所在窗口计算纹理特征, 分析各纹理特征在地物识别分类中的贡献, 再挑选出对分类最有帮助的纹理特征。对整幅影像只需要计算这些最优纹理特征进行后续分类实验即可。

为了更直观的比较各纹理特征, 本节分别利用图 2.2.1 包含 8 个植被类别 80 个波段的示例 PHI 数据, 和直接截取的真实 PHI 数据进行实验, 主成分变换后保留第一主成分, 计算得到上述各纹理波段 (如图 3.1.3 和图 3.1.4)。

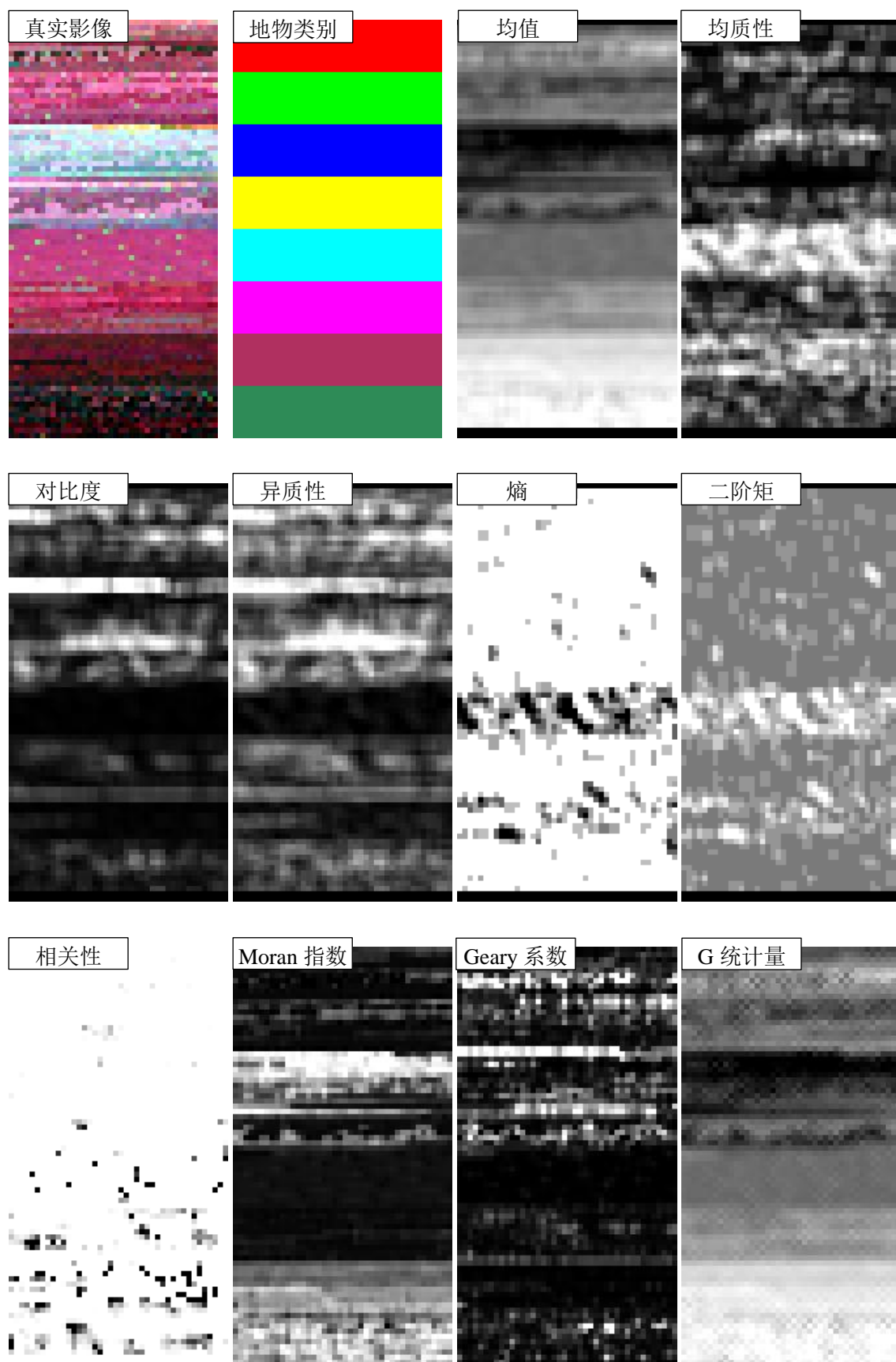


图 3.1.3 基于 GLCM 与 LISA 计算的高光谱遥感影像纹理特征

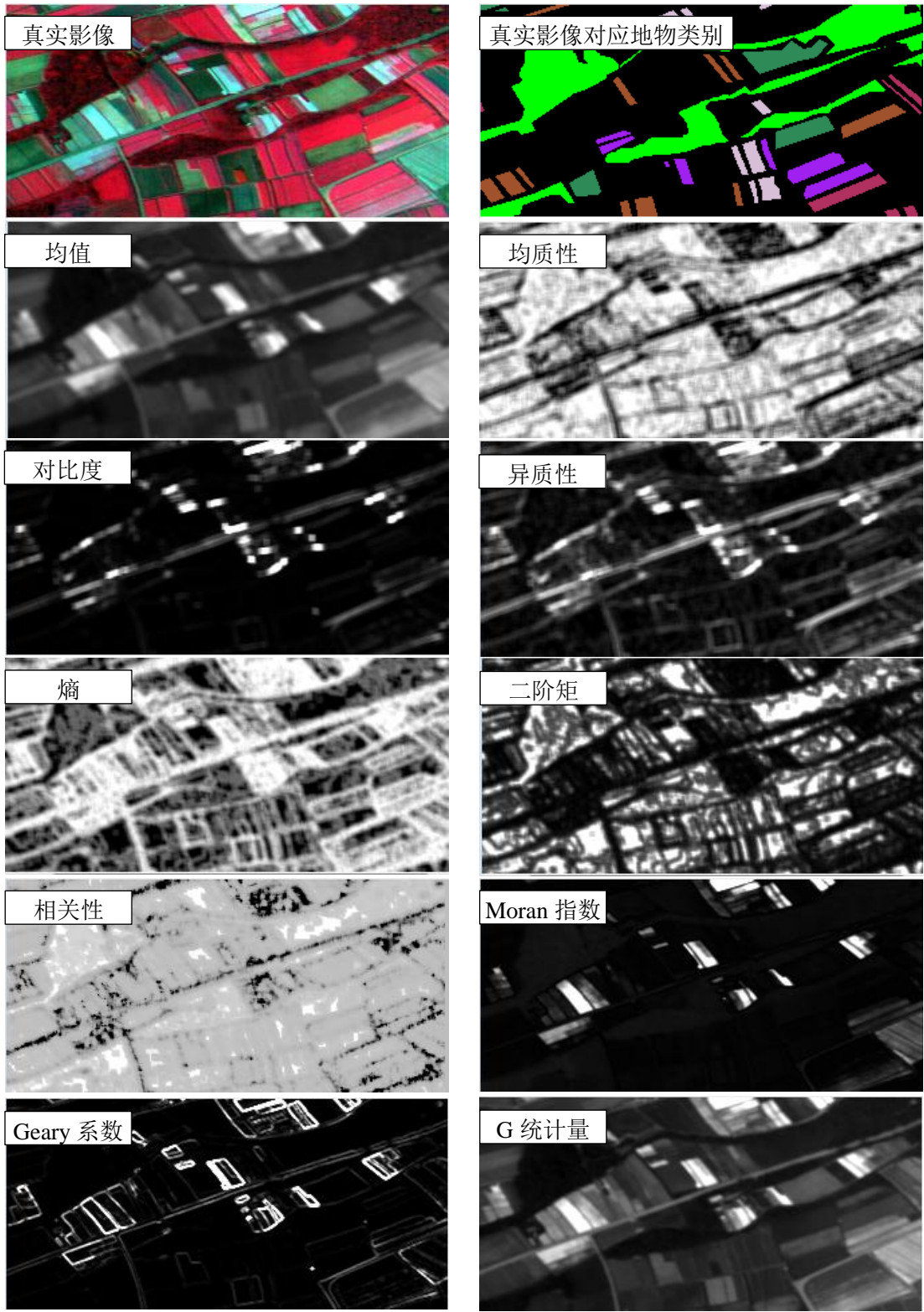


图 3.1.4 基于 GLCM 与 LISA 计算的高光谱遥感影像纹理特征

3.1.3 植被生化参量敏感的光谱指数特征

随着高光谱遥感传感器的飞速发展，近年来植被指数得到了蓬勃发展，其中

部分植被指数与植被中的某些生化组分含量具有较强的相关关系。本章在构建植被特征库时，主要考虑到不同植被类型间叶绿素、类胡萝卜素、花青素和氮素含量的不同，重点选择了一些对这些生化组分含量变化敏感的植被指数，以提高不同植被的识别和分类精度。

通过大量文献调研，本章在构建植被特征库时选取了 50 种光谱指数特征，具体见表 3.1.1。

表 3.1.1 植被生化参量敏感的光谱指数特征

NO.	缩写	表达式	来源文献	生物参数
1	NDVI	$(\rho_{800} - \rho_{670}) / (\rho_{800} + \rho_{670})$	Rouse 等 (1974)	绿色生物量
2	PSDNa	$(\rho_{800} - \rho_{680}) / (\rho_{800} + \rho_{680})$	Blackburn (1998)	叶绿素 a
3	PSDNb	$(\rho_{800} - \rho_{635}) / (\rho_{800} + \rho_{635})$	Blackburn (1998)	叶绿素 b
4	PSDNc	$(\rho_{800} - \rho_{470}) / (\rho_{800} + \rho_{470})$	Blackburn (1998)	类胡萝卜素
5	PI1	$(\rho_{750} - \rho_{705}) / (\rho_{750} + \rho_{705})$	Gitelson (1994)	叶绿素
6	PI2	$(\rho_{780} - \rho_{550}) / (\rho_{780} + \rho_{550})$	Gitelson (1994)	叶绿素
7	NPCI	$(\rho_{680} - \rho_{430}) / (\rho_{680} + \rho_{430})$	Peñuelas 等 (1993)	新黄素与叶绿素 a 比率
8	NPQI	$(\rho_{415} - \rho_{435}) / (\rho_{415} + \rho_{435})$	Barnes 等 (1992)	叶绿素降解
9	PRI1	$(\rho_{570} - \rho_{531}) / (\rho_{570} + \rho_{531})$	Gamon (1992)	叶黄素 氮素
10	PRI2	$(\rho_{515} - \rho_{531}) / (\rho_{515} + \rho_{531})$	Hernández-Clemente 等 (2011)	叶黄素
11	SI	$(\rho_{774} - \rho_{677}) / (\rho_{774} + \rho_{677})$	Zarco-Tejada 等 (2009)	叶绿素
12	PPR	$(\rho_{550} - \rho_{450}) / (\rho_{550} + \rho_{450})$	Jacquemoud (1996)	叶绿素
13	Green NDVI	$(\rho_{750} - \rho_{550}) / (\rho_{750} + \rho_{550})$	Gitelson (1994)	叶绿素

14	NRI	$(\rho_{560} - \rho_{670}) / (\rho_{560} + \rho_{670})$	Schleicher (2001)	氮素
15	SIPI	$(\rho_{445} - \rho_{800}) / (\rho_{445} + \rho_{800})$	Peñuelas 和 Filella (1998)	类胡萝卜素 叶绿素 a 比率
16	LCI	$(\rho_{850} - \rho_{710}) / (\rho_{850} + \rho_{680})$	Le Maire 等 (2004)	叶绿素
17	TCARI	$3[(\rho_{700} - \rho_{670}) - 0.2(\rho_{700} - \rho_{550})(\rho_{700} / \rho_{670})]$	Haboudane (2002)	叶绿素
18	TVI	$0.5[120(\rho_{750} - \rho_{550}) - 200(\rho_{670} - \rho_{550})]$	Broge 和 Leblanc (2001)	叶绿素
19	MCARI	$[(\rho_{700} - \rho_{670}) - 0.2(\rho_{700} - \rho_{550})](\rho_{700} / \rho_{670})$	Daughtry (2000)	叶绿素
20	MCARI ₂	$\frac{1.5[2.5(\rho_{800} - \rho_{670}) - 1.3(\rho_{800} - \rho_{550})]}{\sqrt{(2\rho_{800} + 1)^2 - (6\rho_{800} - 5\sqrt{\rho_{670}}) - 0.5}}$	Haboudane 等(2004)	叶面积指数
21	SR1	ρ_{750} / ρ_{550}	Gitelson (1994)	叶绿素
22	SR2	ρ_{750} / ρ_{710}	Zarco-Tejada 等(2004)	叶绿素
23	Vog4	$\rho'_{715} / \rho'_{705}$	Zarco-Tejada 和 Miller(1999)	叶绿素
24	CI _{rededge}	$\rho_{750} / \rho_{710} - 1$	Gitelson 等 (2005)	叶绿素
25	CI _{green}	$\rho_{800} / \rho_{750} - 1$	Gitelson 等 (2005)	叶绿素
26	MSAVI	$\frac{0.5\rho_{800} + 1 - \sqrt{(2\rho_{800} + 1)^2 - 8(\rho_{800} - \rho_{670})}}{2}$	Qi 等(1994)	绿色生物量
27	OSAVI	$\frac{(1 + 0.16) \cdot (\rho_{800} - \rho_{670})}{(\rho_{800} + \rho_{670} + 0.16)}$	Rondeaux 等 (1996)	绿色生物量
28	TCARI/ OSAVI	TCARI/OSAVI		叶绿素 ab
29	MCARI / OSAVI	MCARI/OSAVI		叶绿素 ab
30	PRI3	$\rho_{675} \cdot \rho_{690} / \rho_{683}^2$	Meroni 等 (2009)	叶绿素

31	PRI4	ρ_{750} / ρ_{800}	Meroni 等 (2009)	叶绿素
32	PRI5	ρ_{685} / ρ_{655}	Meroni 等 (2009)	叶绿素
33	PRI6	ρ_{680} / ρ_{630}	Meroni 等 (2009)	叶绿素
34	PRI7	$\rho_{685}^2 / \rho_{675} \cdot \rho_{690}$	Meroni 等 (2009)	叶绿素
35	PRI8	$\rho'_{688} \cdot \rho'_{710} / \rho'^2_{697}$	Meroni 等 (2009)	叶绿素
36	PRI9	$\rho'_{705} / \rho'_{722}$	Meroni 等 (2009)	叶绿素
37	PRI10	$\rho'_{730} / \rho'_{706}$	Meroni 等 (2009)	叶绿素
38	PRI11	ρ_{690} / ρ_{600}	Meroni 等 (2009)	叶绿素
39	RDVI	$(\rho_{800} - \rho_{670}) / \sqrt{\rho_{800} + \rho_{670}}$	Roujean 和 Breon (1995)	绿色生物量
40	MTVI	$1.2[1.2(\rho_{800} - \rho_{550}) - 2.5(\rho_{670} - \rho_{550})]$	Haboudane 等(2004)	叶绿素
41	PSSRa	ρ_{800} / ρ_{675}	Blackburn (1998)	叶绿素 a
42	PSSRb	ρ_{800} / ρ_{650}	Blackburn (1998)	叶绿素 b
43	BGI1	ρ_{400} / ρ_{550}	Zarco-Tejada 等(2005)	叶绿素 ab
44	BGI2	ρ_{450} / ρ_{550}	Zarco-Tejada 等(2005)	叶绿素 ab
45	ARI	$1 / \rho_{550} - 1 / \rho_{700}$	Gitelson 等 (2001)	花青素
46	mARI	$(1 / \rho_{530-570} - 1 / \rho_{690-710}) \times \rho_{NIR}$	Gitelson 等 (2006)	花青素
47	R520/ R500	ρ_{520} / ρ_{550}	Zarco-Tejada 等(2012)	类胡萝卜素
48	R515/ R570	ρ_{515} / ρ_{570}	Zarco-Tejada 等(2012)	类胡萝卜素
49	R515/ R670	ρ_{515} / ρ_{670}	Zarco-Tejada 等(2012)	类胡萝卜素

50	HI	$\frac{\rho_{534} - \rho_{698}}{\rho_{534} + \rho_{698}} - \frac{\rho_{704}}{2}$	Mahlein 等 (2013)	健康度
----	----	--	---------------------	-----

由于不同的光谱指数特征能反映不同植被生化组分含量的差异，在分类应用中，可以根据先验知识直接从特征库中选取最有利于研究区地物类别区分的特征波段，加入到分类中；在缺少研究区植被类型生化组分方面先验知识的情况下，可以先对各类别选取适量训练样本，根据训练样本计算各类别在不同光谱特征波段中的特征值，判断不同特征对于各类别的区分度，选取对分类器最有帮助的光谱指数特征；然后对整幅影像仅计算最优的光谱指数特征，并与原始光谱特征和空间特征结合，进行后续的分类处理。

下图是图 2.2.1 包含 8 个植被类别 80 个波段的示例 PHI 数据中各植被类型光谱指数特征平均值分布图。

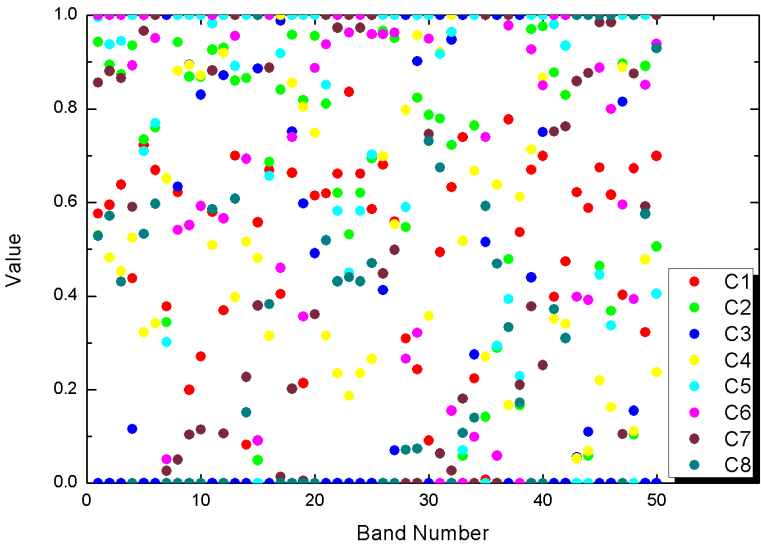


图 3.1.5 各植被类别的光谱指数特征平均值分布

3.2 高光谱数据波段质量评价优化与特征选择

高光谱遥感数据波段数多、波长覆盖范围广，但是部分波段图像质量较差，在进行分类时会带来严重的误分。因此，在实际应用中往往需要对高光谱遥感数据各波段进行质量评价，对质量较差的波段进行一定优化处理后再加以利用。另一方面，高光谱遥感数据光谱维度高，波段间相关性较强，数据本身存在一定程度的信息冗余，而为突出地物特征构建的特征波段库进一步增加了数据的维度和

冗余量，这使得特征降维成为了高光谱遥感分类中的非常重要的一个步骤。

3.2.1 特征波段质量评价与优化

目前用于高光谱遥感图像波段质量评价的指标有很多，如图像信噪比、清晰度、辐射精度、反差等（张霞和张兵，2002），不同指标用于反映图像的不同特性。在高光谱遥感图像分类研究中，为评价图像中随机噪声含量，通常选用信噪比（Signal-Noise Ratio, SNR）这一指标。然而，研究表明，传统的信噪比评价方法只适用于土地覆盖单一的区域，对于土地覆盖类型复杂的区域，该指标难以得到客观的评价结果。针对这一问题，很多学者尝试对该指标进行改进（Gao, 1993; Gao 等, 2008; Martin-Herrero, 2008; Acito 等, 2011）。

在 Gao（1993）改进的信噪比计算方法中，用局域均值与局域标准差代替全图的均值与标准差。该方法主要基于以下假设：遥感图像通常是由许多均质性较高的小块组合而成，同时图像中的噪声主要是加性噪声。该方法首先将图像分割为许多小块，在空间分辨率足够高且地表覆盖足够均匀时，认为这些小块是均匀的。与传统信噪比计算方法相比，该方法可以自动筛选出最均匀的区域。通过计算各小块的标准差作为局域噪声，并以包含块数最多区间的局域标准差作为对整幅影像的最佳噪声估计。具体计算方法如下：

- （1） 首先将图像分为 4×4 、 5×5 、...等众多小块，计算各子块的局域均值（Local Means, LM）与局域标准差（Local Standard Deviations, LSD），如下：

$$LM = \frac{1}{N} \sum_{i=1}^N S_i \quad (3-2-1)$$

$$LSD = \left[\frac{1}{(N-1)} \sum_{i=1}^N (S_i - LM)^2 \right]^{1/2} \quad (3-2-2)$$

其中 S_i 为图像子块中第 i 个像元的灰度值， N 为图像子块的像元总数。

图像子块越均匀，LSD 越小。计算出整幅图像所有子块的 LSD 后，找出其中最大和最小的 LSD；

- （2） 在最大和最小 LSD 值之间建立一些等间隔区间，并按照各子块 LSD

值将其纳入相应区间，将包含块数最多的区间 **LSD** 平均值作为整幅图像的噪声值；

(3) 取整幅图像均值与估算的平均噪声值，将二者之比作为整幅图像 **SNR**。

该算法克服了传统信噪比评价算法中难以自动寻找均匀区域这一问题，但是在应用高光谱图像时仍然易受到地物覆盖类型的影响。为此，高连如（2007）提出了一种基于均匀区域性划分和光谱维高相关性的高光谱图像噪声评估方法（Homogeneous Regions Division and Spectral De-Correlation method, **HRDSDC**）。**HRDSDC** 方法利用高光谱遥感图像中地物均质性和光谱维高相关性的特点，根据地物空间分布的连续性首先对图像进行自动分块，然后在每个均质子块内部利用光谱维的多元线性回归进行去相关获得残差，再将残差标准差作为该子块的噪声标准差，最终取全部子块的噪声标准差的平均值作为图像的信噪比。

通过波段质量评价可以得到高光谱遥感影像中各波段的信噪比分布情况，一般可以通过设置一定的阈值，对信噪比小于该阈值的波段进行一定处理。在处理时可以结合其他图像质量评价指标的评价结果，如辐射精度、清晰度、反差等，对图像中不同类型的噪声进行相应处理。

目前常用的滤波方法主要分为空间滤波和频率滤波。空间滤波方法有均值滤波、统计排序滤波和自适应滤波三类，主要用于消减加性噪声；频率滤波方法主要分为带阻、带通和陷波滤波器，通常用于消减周期性噪声（Gonzalez 和 Woods, 2004）。针对遥感图像分类中影响较大的随机噪声，通常选用均值滤波或中值滤波来提高图像质量，但是该类滤波算法通常会造成部分图像细节信息的丢失。因此，在遥感图像分类中，通常利用分类后处理来抑制分类结果中的椒盐噪声，通过对分类结果进行孤立点筛选，并对这些孤立点进行众数滤波（Majority Filter），来提高图像分类精度。

3.2.2 特征波段选择算法

高光谱遥感数据特征波段选择算法中，最常用的是基于信息量准则与基于类间可分性准则的算法（童庆禧等，2006）。

信息量准则需确保选出的波段包含的信息丰富且相关性小，最优索引因子法

(Optimum Index Factor, OIF) 与协方差矩阵特征值法是其中有代表性的两种方法。最优索引因子定义为:

$$\text{OIF} = \frac{\sum_{i=1}^n S_i}{\sum_{i=1}^n \sum_{j=1}^n |R_{ij}|} \quad (3-2-3)$$

式中, i 和 j 为波段号, S_i 为第 i 个波段的标准差, R_{ij} 为第 i 个波段与第 j 个波段的相关系数, n 为从总的波段 N 中选择的波段数, $1 \leq n \leq N$ 。

协方差矩阵特征值法通过计算波段子集的熵值来优选波段组合, 熵值 S 被定义为:

$$S = -\sum_{i=1}^M P_i(x) \ln P_i(x) \quad (3-2-4)$$

其中, M 为样区总像元数, $P_i(x)$ 为变量的概率密度函数。熵值 S 越大, 代表该波段组合包含的信息量越大, 信息量最大的波段组合即为最佳波段组合。

根据类别可分性准则挑选特征一般应该包括三个步骤: 第一, 选择可分性准则; 第二, 确定使用可分性准则的策略; 第三, 确定选择特征的算法。类间可分性准则根据各待分类样本集在不同波段组合上的类间统计距离选择最优波段组合, 统计距离越大, 类间可分性越强。类间距离的多变量统计度量以离散度、Bhattacharyya 距离 (Bhattacharyya Distance, BD) 和 J-M 距离 (Jeffries-Matusita Distance) 为代表。

类别 i 与类别 j 之间的离散度定义为:

$$D_{ij} = \frac{1}{2} t_r \left[(C_i - C_j)(C_i^{-1} - C_j^{-1}) \right] + \frac{1}{2} t_r \left[(C_i^{-1} - C_j^{-1})(u_i - u_j)(u_i - u_j)^T \right] \quad (3-2-5)$$

式中, u_i 和 u_j 分别是类别 i 与类别 j 图像变量的均值矢量, C_i 和 C_j 分别是类别 i 与类别 j 在波段子集上的协方差矩阵, $t_r[\dots]$ 为矩阵 $[\dots]$ 对角线元素之和。

Bhattacharyya 距离定义为:

$$BD_{ij} = \frac{1}{8} (u_i - u_j)^T \left(\frac{C_i + C_j}{2} \right)^{-1} (u_i - u_j) + \frac{1}{2} \ln \left[\frac{\left| \frac{C_i + C_j}{2} \right|}{|C_i|^{\frac{1}{2}} |C_j|^{\frac{1}{2}}} \right] \quad (3-2-6)$$

式中, u_i 和 u_j 分别是类别 i 与类别 j 图像变量的均值矢量, C_i 和 C_j 分别是类别 i 与类别 j 在波段子集上的协方差矩阵。

通常认为 J-M 距离更适合表达不同类别间可分性，它是基于条件概率理论的光谱可分性指标，具体的计算公式如下：

$$JMD_{ij} = \left\{ \int_x \left[\sqrt{p(x/W_i)} - \sqrt{p(x/W_j)} \right]^2 dx \right\}^{\frac{1}{2}} \quad (3-2-7)$$

其中， $p(x/W_i)$ 为条件概率密度，也就是第 i 个象元属于第 W_i 个类别的概率。 JMD_{ij} 的值通常在 0 到 2 之间，其值越大代表类间可分度越大。J-M 距离位于 0 到 1 时，样本之间一般不具有可分性；位于 1 到 1.9 时，样本之间具有一定的可分性，但存在较高程度的重叠；位于 1.9 到 2 时，样本之间具有较好的可分性。但同时，J-M 距离对样本代表性具有较高的要求，如果单纯追求较高的 J-M 距离，可能会导致选择的特征对样本外数据不具有普适性。

3.3 基于类对可分性与邻域 SAD 的植被特征库优化算法

为了更有效的在高光谱植被精细分类中利用所构建的特征库，本节在上述对已有高光谱波段质量评价优化与特征选择算法进行分析的基础上，提出了针对基于类对可分性（Class Pair Separability, CPS）的植被特征库光谱维优化和基于邻域 SAD（Neighborhood Pixel Spectral Angle Distance, NPSAD）的植被特征库空间维优化算法，一方面用来降低所构建的植被特征库光谱维度、提高运算速度；另一方面用来抑制随机噪声，提高分类精度以及分类图像的平滑度。

3.3.1 基于类对可分性的植被特征库光谱维优化

由于本章所构建的植被特征库主要是用于进行植被精细识别与分类，故在对植被特征库进行光谱维优化中，重点考虑不同植被类别在不同特征中的可分性，即“类对可分性”（Class Pair Separability, CPS），在特征库光谱维优化中，保留对各类对具有最高识别能力的特征波段，增加类别可分度。

在基于 CPS 的植被特征库光谱维优化算法中，每个类对都有自己的最佳波段组，计算过程主要是在类对层面进行的。当类别数为 K 时，类对数量 $NCP = K(K-1)/2$ 。具体计算过程如下。

- (1) 针对各类对，直接对构建的植被特征库按原始光谱特征、空间特征和

光谱指数特征分组计算 B 距离,选出各组中对区分各类对具有最大 B 距离的波段;

- (2) 对筛选出的波段进行去重处理,最终保留 U 个波段;
- (3) 计算各类对在 U 个波段构成的最优波段组中的 $J-M$ 距离;
- (4) 对于 $J-M$ 距离小于 1.95 的类对,选择剩余波段中具有最大 B 距离的波段,加入到最优波段组中,重新计算 $J-M$ 距离;
- (5) 重复步骤(4),直到各类对在最优波段中 $J-M$ 距离均大于 1.95 为止。
- (6) 对步骤(5)得到的最优波段进行相关性分析,保留与其他波段相关性均小于 0.95 的波段,并将彼此间相关性大于或等于 0.95 的各波段归为一组,得到一些相关性较高的波段组;
- (7) 对于各波段组,利用最优索引因子法 OIF (见公式 3-2-3) 进行筛选,各组分别保留 1 个波段,以降低数据冗余。

3.3.2 基于邻域 SAD 的植被特征库空间维优化

在自然界中,地物的分布通常具有一定的空间连续性,反映在遥感影像中,即相邻像元是同种地物类型的概率最大(耿修瑞等,2004)。基于这一理论,本文提出了一种基于 $NPSAD$ 的植被特征库空间维优化算法,该算法在对空间维进行滤波优化的同时,充分利用了植被光谱维特点。在有训练样本的情况下,该算法可基于训练样本自动选定阈值进行计算;在无训练样本的情况下,也可以根据用户手动设置阈值进行计算,算法流程图如下:

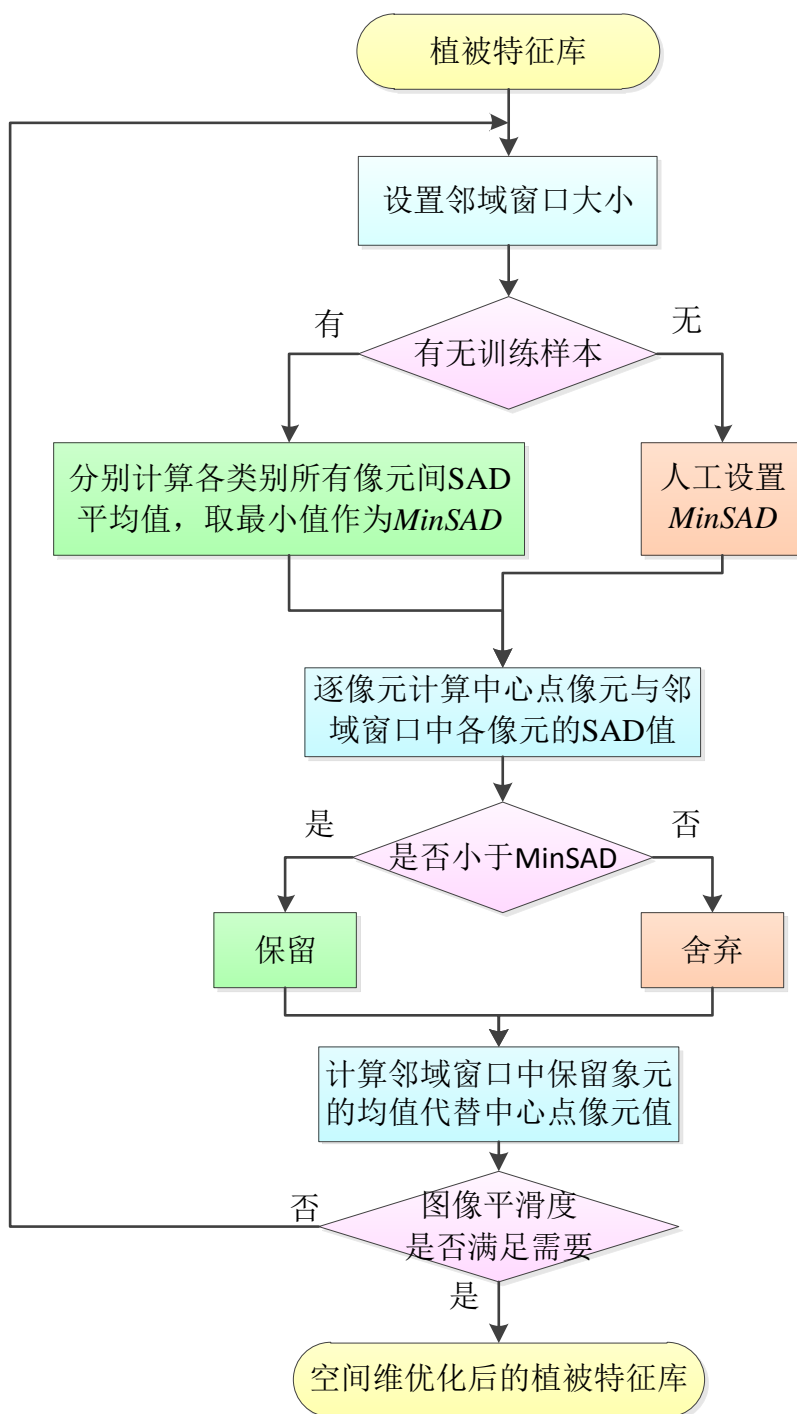


图 3.3.3 基于 NPSAD 的植被特征库空间维优化算法流程图

基于 NPSAD 的植被特征库空间维优化算法的具体计算过程如下：

- (1) 设置邻域窗口大小，如 3×3 , 5×5 , \dots , 9×9 等；
- (2) 设置 SAD 阈值 $MinSAD$:
 - a) 有训练样本情况下，分别计算各类别所有像元间 SAD 平均值，取最

小值作为 $MinSAD$ ，公式如下：

$$MeanSAD_k = \frac{\sum_{i=1}^{NT_k} \sum_{j=i+1}^{NT_k} SAD(TS_{ki}, TS_{kj})}{NT_k(1 + NT_k)/2} \quad (3-3-1)$$

$$MinSAD = \min(MeanSAD_k) \quad (3-3-2)$$

其中， $MeanSAD_k$ 为第 k 类各训练样本间 SAD 平均值， $MinSAD$ 为所有类别的 $MeanSAD_k$ 中最小值， NT_k 为第 k 类训练样本像元个数， TS_{ki} 与 TS_{kj} 表示训练样本 i, j 的特征波段曲线。

- b) 无训练样本情况下，人工设置 $MinSAD$ 阈值。数值越大，图像越平滑；数值越小，保留的细节信息越丰富。
- (3) 逐像元计算中心点像元与邻域窗口中各像元的 SAD 值；
- (4) 计算窗口中与中心点像元 SAD 值小于 $MinSAD$ 的像元平均特征值，用以代替中心点特征值；
- (5) 重复步骤 (1) 到步骤 (4)，直到图像的平滑度可以满足分类需求。

3.4 基于植被特征库构建与优化的分类实验

为验证本章所构建的植被特征库在植被精细识别分类中的有效性，以及本文所提的基于 CPS 与 NPSAD 的植被特征库优化算法在降低数据冗余、消减噪声方面的性能，本节利用图 2.2.1 包含 8 个植被类别 80 个波段的示例 PHI 数据进行分类实验。基于植被特征库构建与优化的植被精细分类实验流程如图 3.4.1。

本实验中主要比较了单纯利用原始波段分类结果与植被特征库全部特征分类结果，原始波段集与植被特征库分别进行光谱维优化前后的分类结果，原始波段集光谱维优化前、原始波段集光谱维优化后、植被特征库光谱维优化前、植被特征库光谱维优化后进行空间维优化前后的分类结果。实验所用训练样本与第二章第三节所用训练样本相同，均为各类别 50 个样本点。验证数据为真实影像。实验中各方法所用数据及方法如表 3.4.1。

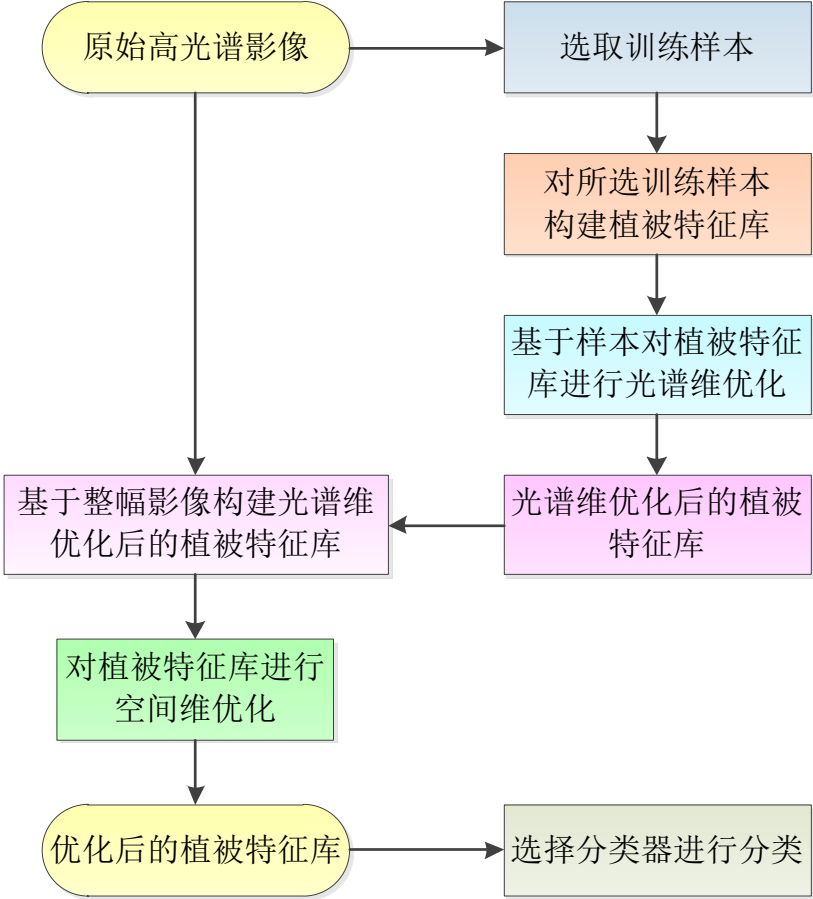


图 3.4.1 基于植被特征库构建与优化的植被精细分类实验流程图

表 3.4.1 实验中各方法所用数据及方法

方法名称	原始波段(O)/ 植被特征库(FT)	光谱维优化(CPS)	空间维优化 (NPSAD)
O	O	×	×
FT	FT	×	×
O-CPS	O	√	×
FT-CPS	FT	√	×
O-NPSAD	O	×	√
FT-NPSAD	FT	×	√
O-CPS-NPSAD	O	√	√
FT-CPS-NPSAD	FT	√	√

植被特征库构建：

本节所构建的植被特征库中，原始波段为反射率数据；空间纹理特征窗口为

3×3，方向是 45°，步长 (1,1)；光谱指数特征为表 3.1.1 中全部指数特征。

光谱维优化：

实验中针对原始波段集，利用基于 CPS 的光谱维优化算法，最终保留 6 个光谱波段，对应的波长如下（单位：nm）：522.1, 600.6, 658.5, 700.7, 723.3, 796.4。

针对综合原始光谱特征、空间特征与光谱指数特征构建的特征库，通过基于 CPS 的光谱维优化算法，最终保留了 18 个特征波段，具体波段信息如下：

- （1）原始光谱特征 5 个，与针对原始波段集计算保留的前 5 个波段相同；
- （2）空间特征 7 个，对应纹理特征为：均值、均质性、对比度、相关性、Moran 指数、Geary 系数、G 统计量；
- （3）光谱指数特征 6 个，对应表 3.1.1 中序号分别为：5、18、19、21、29、36。

空间维优化：

实验中进行空间维优化时，利用训练样本自动计算 MinSAD，窗口大小设置为 5×5。空间维优化结果如下图。

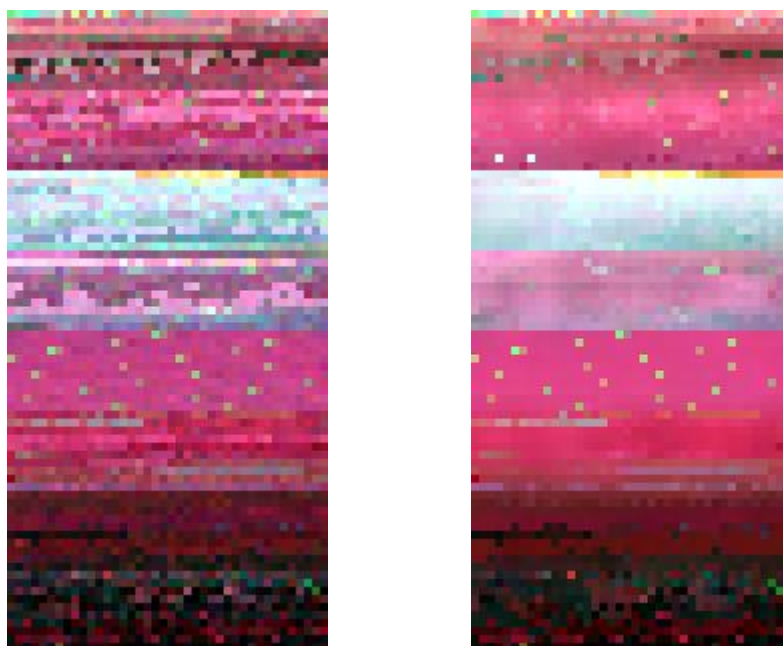


图 3.4.2 空间维优化的标准假彩色影像(RGB: 832.8nm,650.8nm,553.7nm)

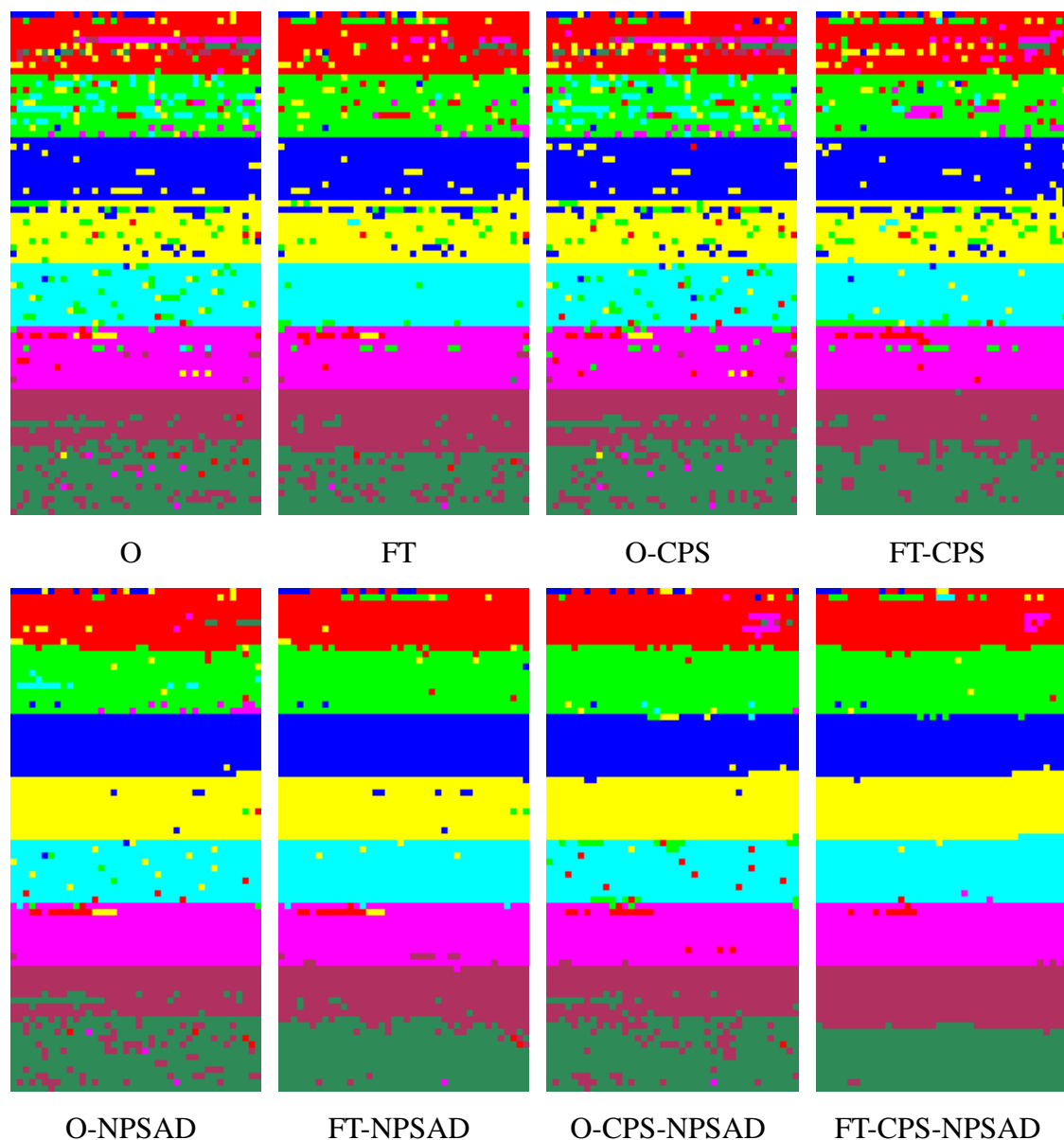


图 3.4.3 各方法对应分类结果图

图 3.4.3 为各分类数据及方法对应的分类结果图，从图中可以清晰的看出，通过构建植被特征库，各类别间的区分度明显提高，尤其是 C7 与 C8 类别。单纯利用原始光谱波段的分类结果中，两类间的混淆非常严重，而利用植被特征库，则得到了较好的划分。而进行光谱维优化后，无论是针对原始光谱集还是构建的植被特征库，波段数量都得到了大幅度减少，而分类结果基本不变。对于空间维优化，从图中可以看出，无论是原始光谱集还是构建的特征库，光谱维优化前还是光谱维优化后，经过空间维优化后，分类结果均得到了大幅度改善，第一行分类结果中的椒盐噪声得到了有效抑制。从图中可以看出，FT-CPS-NPSAD 得到的

分类结果准确度最高，椒盐噪声最少。

表 3.4.2 各方法对应分类结果精度表

注：PA，UA，OAA，KC 分别为生产者精度，用户精度，总体精度和 Kappa 系数。

	O		FT		O-CPS		FT-CPS	
	PA	UA	PA	UA	PA	UA	PA	UA
C1	72.75%	88.18%	81.00%	89.50%	71.75%	87.50%	76.50%	90.27%
C2	70.75%	77.11%	86.50%	83.98%	70.75%	77.11%	88.00%	79.64%
C3	94.75%	87.94%	95.00%	87.96%	94.75%	87.13%	95.50%	88.22%
C4	82.25%	78.90%	84.00%	84.63%	82.50%	80.68%	83.25%	83.88%
C5	87.50%	84.34%	97.75%	99.24%	89.00%	84.76%	92.00%	97.35%
C6	90.50%	87.65%	92.50%	93.91%	89.00%	87.25%	94.25%	93.55%
C7	80.75%	80.35%	94.75%	84.41%	77.50%	80.73%	90.00%	91.84%
C8	82.75%	77.88%	82.75%	91.94%	85.00%	75.72%	92.25%	88.70%
OAA(KC)	82.75% (0.80)		89.28% (0.88)		82.53% (0.80)		88.97% (0.87)	
波段数	80		140		6		18	
	O-NPSAD		FT-NPSAD		O-CPS-NPSAD		FT-CPS-NPSAD	
	PA	UA	PA	UA	PA	UA	PA	UA
C1	87.25%	93.57%	89.50%	95.47%	85.75%	92.70%	88.25%	96.71%
C2	90.50%	93.78%	97.75%	93.54%	97.75%	88.46%	98.00%	92.89%
C3	98.75%	95.18%	100%	95.01%	95.75%	95.99%	96.50%	97.72%
C4	98.50%	91.20%	97.75%	96.31%	99.00%	94.74%	97.75%	95.37%
C5	94.75%	96.19%	99.00%	99.25%	91.75%	98.92%	98.75%	97.29%
C6	95.25%	95.25%	93.25%	99.47%	95.00%	95.96%	97.25%	97.49%
C7	76.50%	84.53%	88.00%	93.62%	81.00%	87.33%	98.00%	98.49%
C8	84.50%	77.17%	95.75%	89.07%	88.75%	81.99%	99.25%	98.02%
OAA(KC)	90.75% (0.89)		95.13% (0.94)		91.84% (0.91)		96.72% (0.96)	
波段数	80		140		6		18	

从表 3.4.2 可以看出, 利用所构建的植被特征库得到的分类精度比单纯利用原始波段所得分类精度有大幅度提升; 光谱维优化后, 分类器运算所用的特征波段明显减少, 分类精度基本不变; 空间维优化后, 分类精度显著提高, 最高可提高 9% 以上。

基于植被特征库构建与优化的分类方法 (FT-CPS-NPSAD) 在仅用 18 个波段的情况下, 得到了最高的总体分类精度 96.72% 和 Kappa 系数 0.96。与单纯利用原始波段集 80 个波段 (O) 相比, 分类精度提高约 14%, Kappa 系数提高 0.16; 与直接利用植被特征库或光谱维优化后植被特征库 (FT 或 FT-CPS) 相比, 在波段数减少或一致的情况下, 总体精度提高大于 7%, Kappa 系数提高 0.08; 与原始波段直接进行空间维优化分类 (O-NPSAD) 相比, 总体精度提高约 6%, Kappa 系数提高 0.07。

同时, FT-CPS-NPSAD 分类结果中各类别的 PA 与 UA 也基本达到了 95% 以上 (除了 C1 类别的 PA=85.50% 与 C2 类别的 UA=92.89%), 尤其是对于最难分的 C7 与 C8 类别, 二者 PA 与 UA 均大于 98%, 这是其他分类方法均无法达到的精度。

综上, 实验结果表明, 基于植被特征库构建的方法通过增加植被精细分类中的特征波段, 可以有效提高各植被类别的识别能力, 增强各类别间可分性; 基于 CPS 的植被特征库光谱维优化算法则可以自动搜索出对分类最有帮助的特征波段, 大幅度降低数据冗余, 提高分类器运算效率; 基于 NPSAD 的植被特征库空间维优化算法则可以有效抑制分类结果中的椒盐噪声, 显著提高分类精度。

3.5 本章小结

本章从植被在高光谱影像中的光谱特征、空间纹理特征出发, 重点结合对植被生化参量敏感的光谱指数特征, 构建了针对植被精细识别分类的植被特征库; 同时在分析已有高光谱数据波段质量评价优化与特征选择算法的基础上, 提出了基于 CPS 的特征库光谱维优化算法与基于 NPSAD 的特征库空间维优化算法, 并以包含 8 个植被类别的实验数据对所构建的特征库, 以及光谱维、空间维优化的效果进行了测试。结果表明, 基于植被特征库构建与优化的分类方法可以大幅度提高分类精度和效率。

针对高光谱植被精细分类的特征库构建主要基于植被的光谱、空间纹理和生化参量三方面属性，综合了原始光谱、空间纹理和对植被生化参量敏感的光谱指数三类特征。其中空间纹理特征主要包括基于灰度共生矩阵和局域指示空间分析两类纹理特征，共选取了 10 个纹理特征。而对植被生化参量敏感的光谱指数特征中，主要考虑到不同植被类型间叶绿素、胡萝卜素、花青素和氮素含量的不同，重点选择了 50 个对这些生化组分含量变化敏感的植被指数，以提高不同植被的识别和分类精度。

本章在对植被特征库进行光谱维优化中，重点考虑不同植被类别在不同特征中的可分性，即“类对可分性”，在特征库光谱维优化中，保留对各类对具有最高识别能力的特征波段，通过迭代使各类别间均达到较高的区分度，并利用最优索引因子法进一步降低数据冗余。通过该算法，可以有效的降低数据冗余，减少分类器运算时间，提高分类效率。

在植被特征库的空间维优化中，本章主要基于地物分布通常具有一定的连续性这一理论，提出了基于 NPSAD 计算的空间维优化算法。在进行监督或半监督分类时，该算法可以根据训练样本自动计算 SAD 阈值；在进行非监督分类或其他应用时，则需要用户根据先验知识来设置阈值。利用该算法，可以有效的去除分类结果中的椒盐噪声，提高分类精度和分类图像平滑度，并保留一定的细节信息。

基于植被特征库构建与优化的分类流程为：首先选取一定数量训练样本，基于各类别样本点及所在窗口计算植被特征库中所有特征；然后针对训练样本数据，利用基于 CPS 的光谱维优化算法得到光谱维优化后的植被特征库；之后对于整幅影像构建光谱维优化后的植被特征库；再利用基于 NPSAD 的空间维优化算法对上一步得到的特征库进行空间维优化；最后选用合适的分类器对整幅影像进行分类，计算分类精度。

第4章 基于 SVM 的高光谱植被半监督分类算法优化

在传统高光谱植被监督分类中,分类器对于标记训练样本有较高的需求,这些样本的获取需要耗费大量的人力物力。而相对于标记样本,遥感图像中存在大量易获取的未标记样本。因此,有必要研究同时利用少量已标记样本和大量未标记样本的半监督分类算法。从第二章的研究结果可以看出,与其他分类器相比,支持向量机分类算法无论是在小样本情况下,还是在样本充足情况下,无论是对高维空间还是降维后的特征空间,均能得到较好的分类效果。因此,本章主要研究基于支持向量机的高光谱植被半监督分类算法。

本章将首先介绍传统支持向量机(SVM)原理,然后分别介绍本文提出的基于 SAD-ED 双重判别的渐进直推 SVM 分类算法(SAD/ED-PTSVM),以及结合混合像元分解进行主动学习的 SVM 分类算法(SUAL-SVM),并以图 2.2.1 中包含 8 个植被类别 80 个波段的示例 PHI 数据进行实验,分析所提半监督分类算法的性能。

4.1 支持向量机基本原理

SVM(Support Vector Machines, SVM)作为新一代的统计学习分类方法,是 20 世纪 90 年代开始发展起来的,并在文本分类、图像分类、手写字符识别等领域得到了广泛应用。Gualtieri 和 Crompton(1999)最早将 SVM 引入到遥感图像分类中。在遥感领域, SVM 在高光谱遥感数据分类中应用广泛。与一般监督分类(如最大似然分类方法)不同的是, SVM 在分类时不需要数据一定符合高斯分布,同时 SVM 对 Hughes 现象(Hughes, 1968)敏感程度也较低。另外,遥感影像分类中, SVM 通常能够用少量的训练样本获得比常规方法更高的分类精度(Melgani 和 Bruzzone, 2004; Mountrakis 等, 2011; Foody 和 Mathur, 2004)。

然而,研究表明虽然 SVM 分类性能卓越,精度较高,但其参数设置复杂,分类结果对于参数的设置非常敏感。因此,本节将在介绍 SVM 基本原理与核函数的基础上,分析可能对分类精度带来较大影响的参数,并对目前常用的参数优化策略进行简单的介绍。

4.1.1 支持向量机基本原理

SVM 起初是为两类分类设计的，在给定训练样本集的情况下，SVM 分类器的目标就是找到最优线性超平面，来使两种类别的训练样本尽可能分开。

考虑到机器学习过程不但要求经验风险最小，同时还应该使 VC 维（Vapnik-Chervonenkis Dimension）尽可能小，以此获得对未来样本较好的预测能力。基于这一点，Vapnik（1998）提出了结构风险最小化原则（Structural Risk Minimization, SRM）。支持向量机就是采用结构风险最小化原则，在使样本误差最小化的同时缩小模型泛化误差的上界，进而提高模型的泛化能力。

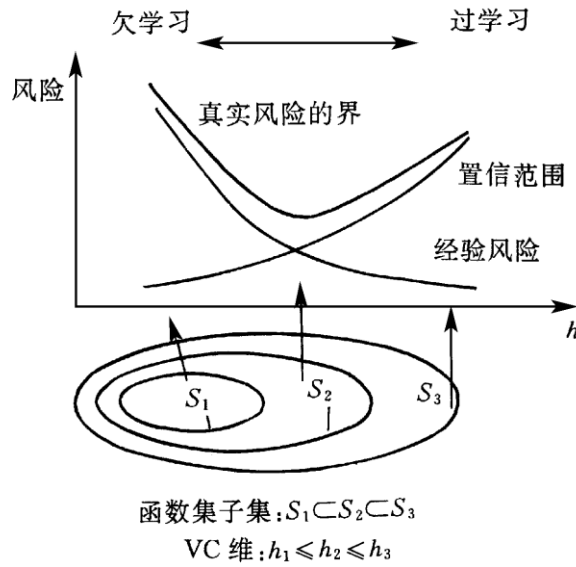


图 4.1.1 有序风险最小化示意图（张学工，2000）

SVM 的基本原理如下：

对于一组给定的标记训练样本集 $\{(x_1, y_1), \dots, (x_n, y_n)\}$ ，其中 $x_i \in R^N$ 且 $y_i \in \{-1, +1\}$ ，以及一个非线性高维（Hilbert）空间映射 $\phi(): R^N \rightarrow H$ 。SVM 为求解以下最优化问题：

$$\min_{w, \xi_i, b} \left\{ \frac{1}{2} \|w\|^2 + C \sum_i \xi_i \right\} \quad (4-1-1)$$

约束条件为：

$$y_i (\langle \phi(x_i), w \rangle + b) \geq 1 - \xi_i, \quad \forall i = 1, \dots, n \quad (4-1-2)$$

$$\xi_i \geq 0, \quad \forall i = 1, \dots, n \quad (4-1-3)$$

这里, w 和 b 用来定义特征空间中的线性分界面, 如图 4.1.2。非线性映射 ϕ 需符合 Cover 定理 (Cover, 1965), 以确保映射到高维空间后样本线性可分。 C 是惩罚系数, 用来控制分类器的泛化能力, 而 ξ_i 则是松弛变量, 用来控制允许的误差, 如图 4.1.2 (b)。

最优化问题 (4-1-1) 可以通过求解其如下拉格朗日对偶问题来解决:

$$\max_{\alpha_i} \left\{ \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \langle \phi(x_i), \phi(x_j) \rangle \right\} \quad (4-1-4)$$

约束条件为:

$$0 \leq \alpha_i \leq C \quad (4-1-5)$$

$$\sum_i \alpha_i y_i = 0, \quad \forall i = 1, \dots, n \quad (4-1-6)$$

其中, α_i 为拉格朗日乘子。这里映射 ϕ 在 SVM 学习中以内积的形式表现, 可通过核函数实现。核函数具体可定义为:

$$K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle \quad (4-1-7)$$

非线性 SVM 可以不用精确的计算具体映射而只通过核函数即可实现。通过核函数的引入, 可以得到对偶问题 (4-1-4) 的解:

$$w = \sum_{i=1}^n y_i \alpha_i \phi(x_i) \quad (4-1-8)$$

以及决策函数:

$$f(x) = \text{sgn} \left(\sum_{i=1}^n y_i \alpha_i K(x_i, x) + b \right) \quad (4-1-9)$$

其中 b 可以通过不为 0 或 C 的 α_i 进行简单计算而得到。利用该决策函数, 我们可以对任何测试向量 x 进行分类。

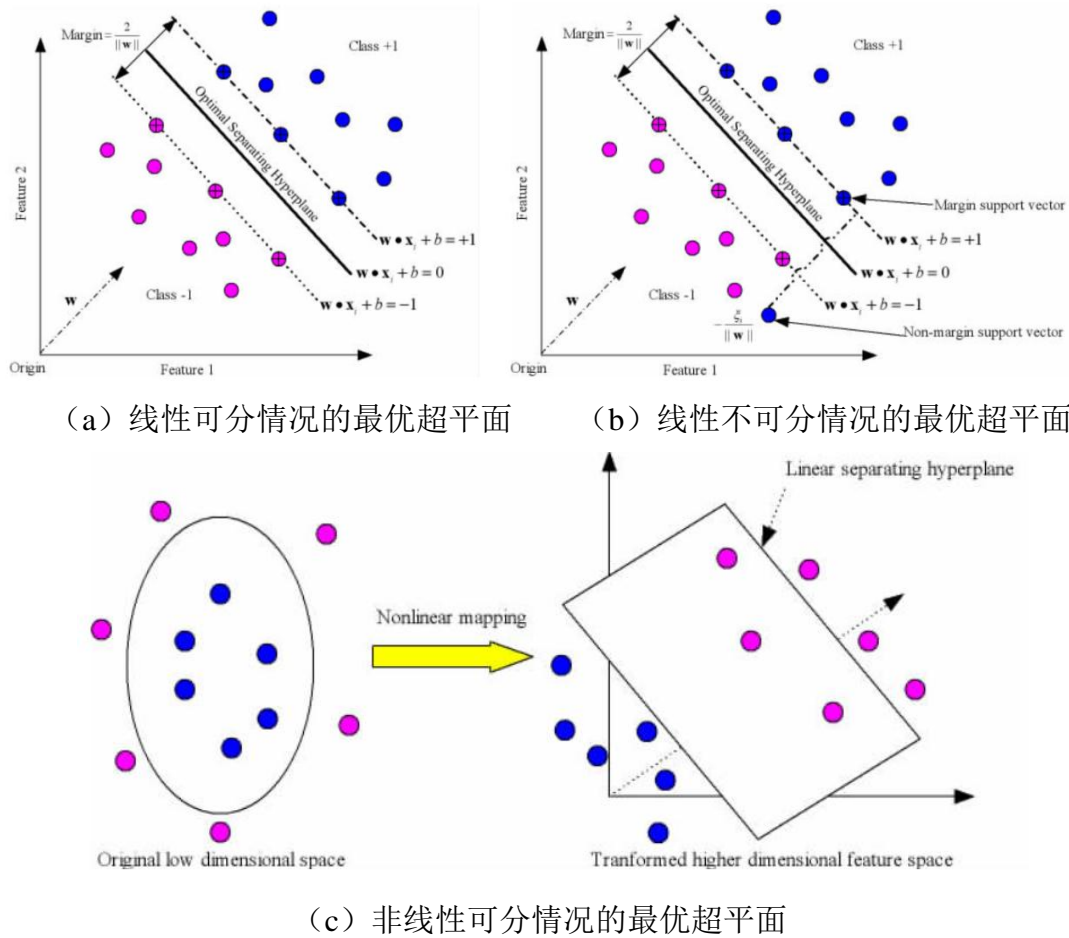


图 4.1.2 SVM 最优超平面示意图 (边肇祺和张学工, 2000)

另外, 考虑到 SVM 起初是为两类分类设计的, 在利用 SVM 进行多类分类时, 需要进行一定扩展。常用的 SVM 多类分类器主要有两种构造方法: 一是对于每两个类别, 构造一个用于识别这两个类别的两类 SVM 分类器, 然后各两类分类器判别结果通过一定方式组合进行多类分类, 即“1-V-1 (one-versus-one)”方法; 二是综合各分类面的参数求解得到一个最优化问题, 然后再求解这个最优化问题一次地实现多类分类, 即“1-V-R (one-versus-rest)”方法。与第一种方法相比, 第二种方法在最优化求解问题时中变量过多, 训练速度慢, 分类精度也较差, 因此一般研究多采用第一种方法 (杜培军等, 2006)。

目前, 有很多机构都开发了自己的 SVM 核心算法程序, 这些核心程序只提供针对特定格式的数据进行训练和分类的基本程序。针对不同应用领域, 如文本分类、遥感图像分类等, 必须开发专门的应用程序, 将数据转换成 SVM 程序可以处理的格式, 然后将这些核心程序嵌套在具体的程序环境中去应用 (程涛,

2006)。对于 SVM 分类器的具体实现，目前比较流行的软件包有 SVM^{light} (Joachims, 1999) 和 LIBSVM (Chang 和 Lin, 2011)。

4.1.2 支持向量机核函数及参数介绍

从上一节的介绍可知，核函数的存在使得原始空间中不可分的向量在高维空间中变得可以区分，因此核函数的选择在 SVM 分类结果有重要影响。核函数的构造需要满足 Mercer 准则(Mercer's Condition)。目前常用的核函数主要有四种，具体如下：

(1) 线性核函数 (Linear Kernel Function)

$$K(x_i, x_j) = x_i^T x_j \quad K(x_i, x_j) = x_i^T x_j \quad (4-1-10)$$

(2) 多项式核函数 (Polynomial Function)

$$K(x_i, x_j) = (\gamma x_i^T x_j)^d, \quad \gamma > 0 \quad (4-1-11)$$

(3) 径向基核函数 (Radial Basis Function)

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \quad \gamma > 0 \quad (4-1-12)$$

(4) S 型核函数 (Sigmoid Function)

$$K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r) \quad (4-1-13)$$

上面公式中， d ， γ 和 r 都是核参数。

这四种核函数，总体可以归结为两大类：局部性核函数和全局性核函数。局如径向基核函数就是一个典型的局部性核函数，而多项式核函数则是典型的全局性核函数。一些研究中会将两类核函数进行结合，即合成核函数 (Camps-Valls 等, 2006; Marconcini 等, 2009; Li 等, 2013)。合成核函数在进行多源数据综合分类方面具有重要作用 (尚坤等, 2011)。

Hsu 等(2003)对于常用的核函数进行了一定的比较研究，提出通常情况下，径向基核函数应该是首选核函数：

首先，与线性核函数不同的是，径向基核函数可以处理类别之间的标签与属性关系为非线性的情况。线性核函数也可以看作是径向基核函数的一种特例 (Keerthi 和 Lin, 2003)，只是系数 (C , γ) 中的惩罚系数 C 不变。而 S 型核函

数在某些参数设置情况下，与径向基核函数相似（Lin 和 Lin，2003）。

其次，高维参数的数量影响分类模型选择的复杂度，而径向基核函数的参数数量远小于多项式核函数。

最后，径向基核函数在数值计算上的困难更少，其核函数的值通常在 0 到 1 之间；而当维度很大时，多项式核函数的值则在 0 到正无穷；而 S 型核函数在某些参数设置情况下则是无效的。

4.1.3 支持向量机参数优化策略

由于 SVM 算法参数设置复杂，且分类精度对参数变化非常敏感，在实际应用中，通常会选择一些优化策略来调整参数。常用的参数优化算法主要有粒子群优化算法（Particle Swarm Optimization, PSO）、遗传算法（Genetic Algorithms, GA）和格网搜索算法（Grid Search, GS），具体算法参考文献（Kennedy 和 Eberhart, 1995; Goldberg, 1989; Hsu 等, 2003）。其中，以格网搜索算法应用最为广泛。格网搜索算法主要是利用交叉验证方法来自动确定参数，为了提高搜索效率，通常先进行粗搜索得到目标区所在格网，再针对格网逐步细化搜索，最终得到最佳参数。

由于本文采用的核函数为径向基核函数，所以主要利用格网搜索算法确定惩罚系数 C 与参数 γ 。

4.2 基于 SAD-ED 双重判定的渐进直推支持向量机分类

直推学习是机器学习领域处理未标记样本的一种有效手段，它通过对未标记样本的挖掘，将其与已标记样本综合，来构造一个更优的分类器。Joachims（1999）提出直推支持向量机（Transductive Support Vector Machine, TSVM）这一概念并将其成功应用在文本分类中，有效的改善了分类性能。与传统 SVM 相比，TSVM 只需要利用较少的标记样本，便可以获取相对较高的分类精度。考虑到遥感图像具有覆盖区域广、标记样本获取困难，同时未标记样本丰富、相同植被类别具有一定相关性等特点，直推学习无论是理论上还是实践中，对于高光谱遥感植被精细分类都具有重要价值。

本节将首先介绍传统 TSVM 的原理及目前的改进算法，之后针对现有算法

的不足同时结合高光谱遥感数据特点，提出改进的基于 SAD-ED 双重判定的渐进 TSVM 高光谱植被精细分类算法，并通过图 2.2.1 中包含 8 个植被类别的 PHI 示例数据测试分类器性能，最后对改进算法的意义、适用性以及不足进行讨论。

4.2.1 直推支持向量机原理

Joachims (1999) 所介绍的 TSVM 的基本原理如下：首先给定一组独立同分布的标记训练样本点：

$$(x_1, y_1), \dots, (x_n, y_n), \text{ 其中 } x_i \in R^m, y_i \in \{-1, +1\} \quad (4-2-1)$$

和另一组来自同一分布的未标记样本点：

$$SAD(T, R) = \cos^{-1} \left(\frac{\sum_{i=1}^{nb} T_i R_i}{\left(\sum_{i=1}^{nb} T_i^2 \right)^{\frac{1}{2}} \left(\sum_{i=1}^{nb} R_i^2 \right)^{\frac{1}{2}}} \right) \quad (4-2-2)$$

在一般线性不可分的条件下，Joachims 的 TSVM 的训练过程可以描述为如下的优化问题 Minimize over $(y_1^*, \dots, y_n^*, \vec{w}, b, \xi_1, \dots, \xi_n, \xi_1^*, \dots, \xi_k^*)$ ：

$$\frac{1}{2} \|\vec{w}\|^2 + C \sum_{i=0}^n \xi_i + C^* \sum_{j=0}^k \xi_j^* \quad (4-2-3)$$

约束条件为：

$$\forall_{i=1}^n : y_i [\vec{w} \cdot \vec{x}_i + b] \geq 1 - \xi_i \quad (4-2-4)$$

$$\forall_{j=1}^k : y_j^* [\vec{w} \cdot \vec{x}_j^* + b] \geq 1 - \xi_j^* \quad (4-2-5)$$

$$\forall_{i=1}^n : \xi_i > 0 \quad (4-2-6)$$

$$\forall_{j=1}^k : \xi_j^* > 0 \quad (4-2-7)$$

其中 C 和 C^* 为用户设置的惩罚系数。

TSVM 的一般步骤如下 (Joachims, 1999)：

第一步：设置已标记样本和未标记样本的影响因子，使用标准 SVM 对标记样本进行一次初始学习，得到一个初始的分类器，然后按照一定规则设置未标记样本中的正类样本个数 N ；

第二步：利用初始分类器对未标记样本进行初始分类，根据未标记样本判别函数的输出情况，对输出值最大的 N 个未标记样本暂时赋予正类标记，其余未标记样本赋予负类标记，同时设置一个临时影响因子；

第三步：重新训练所有的样本，得到一个新分类器，然后按照一定的规则交换一对标记值不同的测试样本的标注，使 SVM 求解的最优化问题目标函数值获得最大幅度的下降。反复执行这一步骤，直至没有符合交换条件的样本为止；

第四步：均匀的增加临时影响因子的值，并返回到第三步，当临时影响因子大于未标记样本的影响因子时，运行结束，输出分类结果。

与传统 SVM 相比，TSVM 在训练样本误差最小化方面更符合实际应用需要，分类效果更好，但是复杂度也高了很多，尤其是在未标记样本数量较多的情况下。另外，TSVM 在算法执行之前需人为设定未标记样本中的正类样本数目，而通常情况下这个数值是很难估计的，一旦设置的正类样本数目与实际偏差较大，会导致分类器性能迅速下降。

针对这些问题，Chen 等 (2003) 提出了一种渐进直推支持向量机 (Progressive Transductive Support Vector Machine, PTSVM)。该算法的基本思想是在训练开始之前不对未标记样本的分布做任何预先估计，而是在训练过程中利用成对标注法将新选择的样本加入到已标记样本中，进行新一轮训练。PTSVM 的渐进赋值规则是指，通常新添加的样本会一定程度影响下一轮的训练，促使当前的分类超平面发生少量偏移。而标签重置法可以使迭代过程早期出现的对未标记样本的误标，在后期得以重置为未标记状态，继续执行迭代，并有可能在之后的训练中得到更准确的标记。标签重置法使 PTSVM 具备自动修复差错的能力。

PTSVM 的主要步骤如下 (Chen 等, 2003)：

第一步：分别设置已标记样本与未标记样本的惩罚系数 C 和 C^* ，然后使用归纳式学习对已标记样本进行初始学习，获得初始的分类器；

第二步：用第一步得到的分类器对未标记样本进行学习，计算每个未标记样本对应的决策函数值，用成对标注法则在当前边界区域内选择具有最大、最小决策值的未标记样本，分别为这两个样本添加正、负类别标注；

第三步：对所有样本重新进行训练，计算每一个未标记样本的判别函数输出，

若发现某一个早期未标记样本的标注值和当前判别函数的输出不符,则按照标签重置法则删除对该样本的标注;

第四步:用成对标注法寻找当前边界区域内符合新添加标注条件而未标注的未标记样本:如存在,则添加标注并返回到上一步;如果不存在,则用当前的决策函数对全部样本进行分类。算法结束,输出结果。

PTSVM 可以解决传统 TSVM 需要预先估计样本中正负样本比例这一问题,同时采用成对标记法使初始分划面能够逐步调整到最优,并利用标签重置法控制误标记样本对分划面造成的影响。

4.2.2 基于 SAD-ED 双重判定的渐进直推支持向量机

虽然 PTSVM 可以在一定程度上解决传统 TSVM 分类中的问题,并具有一定推广能力,但是也不可避免存在一定的局限性,如根据未标记样本决策函数进行标记的方法性能不稳定、训练速度慢等。PTSVM 每次只是选择决策函数值最高的两个未标记样本分别记为正、负类,添加到训练样本集中优化当前分划面。这在样本空间维度较高时,很容易导致未标记样本的误标注。虽然后续通过标签重置,有可能使之前误标记的样本得以纠正,但是在初始标记样本代表性较差、初始分划面很不准确的情况下,分划面很容易向错误的方向发展而无法得到较好的分类结果。

为了解决 PTSVM 在样本标记方面的问题,同时考虑到高光谱遥感影像中同类地物具有相似的光谱曲线、不同地物的光谱曲线通常存在一定差异的特点,本文提出了一种基于光谱角距离 (Spectral Angle Distance, SAD) 与欧氏距离 (Euclidian Distance, ED) 双重判定的 PTSVM 分类算法 (SAD/ED-PTSVM)。该算法在 PTSVM 样本标注中,利用 SAD 与 ED 分别对未标记样本标签进行判定,同时结合其到分划边界的距离对未标记样本进行标注,增加未标记样本标注的准确度,进而加速分划面的优化进程。

SAD/ED-PTSVM 原理如下:与传统 TSVM 相似,首先给定一组标记训练样本点:

$$(x_1, y_1), \dots, (x_n, y_n), \text{ 其中 } x_i \in R^m, y_i \in \{-1, +1\} \quad (4-2-8)$$

和一组未标记样本点：

$$x_1^*, x_2^*, x_3^*, \dots, x_k^* \quad (4-2-9)$$

在传统 TSVM 中, C^* 是逐渐增加的, 而在 PTSVM 中, 由于对于未标记样本的标注是渐进完成的, 希望先前的被标注样本有足够的力量向好的方向调整当前的分界面, 所以一开始就给 C^* 赋予了一个较大的值。

考虑到本文所选的未标记样本已经过初始筛选和 SAD 与 ED 的双重判定, 样本类别标注的确定度较高, 另一方面, SVM 参数设置复杂, 减少参数可以有效减少参数优化过程所耗时间。因此, 本文在最优化问题 (4-2-3) 中, 将 C^* 与 C 进行合并, 用同样的惩罚系数处理初始标记样本和新添加标注的未标记样本。

在未标记样本选择中, 只有当未标记样本满足以下条件时, 才会添加标注并加入到训练样本集中:

- (1) 在当前分界面的边界区域内: 对于未标记样本, 应满足:

$$\text{Max} \|f(x_i^*)\|, \text{ s.t. } \|f(x_i^*)\| < 1 \quad (4-2-10)$$

- (2) 到正负决策边界距离最近: 本算法保留了 PTSVM “成对标注” 渐进优化分界面这一特点, 每次选择最靠近决策边界的一对未标记样本, 即:

$$\text{Max}(f(x_i^*)), \text{ s.t. } 0 < f(x_i^*) < 1 \quad (4-2-11)$$

$$\text{Min}(f(x_j^*)), \text{ s.t. } -1 < f(x_j^*) < 0 \quad (4-2-12)$$

- (3) SAD 与 ED 类别判定结果相同: 与先前算法不同的是, 本算法在对未标记样本标注过程中, 考虑到同类地物光谱相似、不同地物光谱差异较大这一特点, 引入 SAD 进行光谱相似性度量, 同时为解决光谱角距离只考虑波形不考虑数值差异这一问题, 加入了 ED 进行双重判定。对于前两步筛选出的正/负未标记样本, 分别找到与其 SAD 与 ED 最近的已标记样本, 如果这两个已标记样本为相同类别, 则标注该未标记样本为此类别; 否则, 放弃此点搜索下一个满足条件的点。需要注意的是, 为了从不同角度对样本进行判定, 计算 SAD 与 ED 均是在原始空间进行, 而不是变换后的核空间。

- (4) 正负类未标记样本同时添加：为了保证正、负训练样本比例，防止分划面朝某一方向过分偏移，需对上一步得到的两个未标记样本进行类别检验：如果二者标注为不同类别，则添加到训练样本集中进行后续训练；如果二者标注为相同类别，则保留决策函数值绝对值较大的样本，继续搜索第二步中下一个满足条件的未标记样本。

SAD/ED-PTSVM 分类算法流程如图 4.2.2，具体实现步骤如下：

第一步：构建初始标记样本集 **LS** 和未标记样本集 **ULS**，以及未标记训练样本集 **ULTS**（初始时空），设置参数。

第二步：利用 **LS** 建立初始决策函数、构建分划面和分划边界。

第三步：计算 **ULS** 与 **ULTS** 中各样本点的决策值，将 **ULS** 中位于当前分划面边界区域内的样本点添加到待标注样本集 **TempUL** 中，同时对 **ULTS** 中判别结果与之前类别标注不同的点，从 **ULTS** 中删除并加回到 **ULS** 中。

第四步：选择 **TempUL** 中到正负决策边界距离最近的两个点，记为 x_i^* 和 x_j^* ；

第五步：对于 x_i^*/x_j^* ，分别计算 **LS** 与其 SAD 与 ED 最近的点，如果两个点（或同一个点）类别相同，则为 x_i^*/x_j^* 添加该标注；否则，放弃此点，返回第四步，继续搜索下一点，直到找到满足条件的两个点 x_p^* 和 x_n^* 。

第六步：判断 x_p^* 和 x_n^* 的类别，如果不同，则将这两个点添加到 **ULTS** 中，并从 **ULS** 中删除；如果相同，保留二者更靠近分划边界（决策值的绝对值较大）的点，删除另一点，返回第四步，搜索下一点，直到找到满足条件的点为止，将其添加到 **ULTS** 中并从 **ULS** 中删除。

第七步：将 **ULTS** 与 **LTS** 结合得到新的决策函数，返回第三步；直到 **ULS** 中找不到满足条件的点为止，用当前的决策函数为影像中所有像元分类。

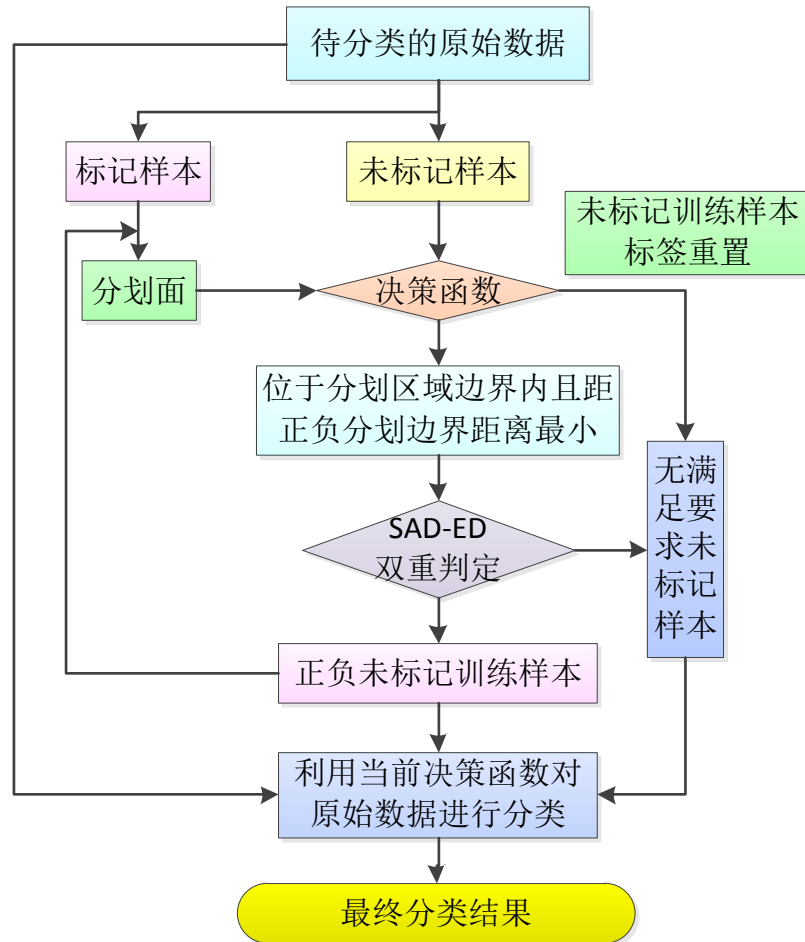


图 4.2.2 SAD/ED-PTSVM 分类算法流程图

4.2.3 实验与结果

为了验证 SAD/ED-PTSVM 分类算法性能，本文利用图 2.2.1 包含 8 个植被类别 80 个波段的示例 PHI 数据，主要从以下两方面对该算法进行了分析和测试：

- (1) 分析不同初始标记样本数量 (n_{Lab})、未标记样本数量 (n_{ULab}) 对分类精度的影响；
- (2) 比较加入 SAD-ED 双重判定与单纯依靠决策函数值进行类别标注的 PTSVM 分类结果。

在参数优化方面，采用网格搜索参数优化策略，搜索的最优 C 区间为 0.1 到 100，最优 γ 区间为 0.01 到 1000，分类精度见图 4.2.3-图 4.2.6。

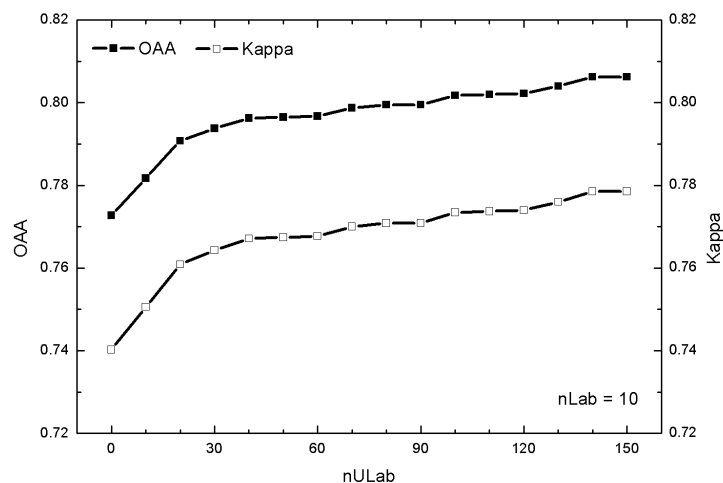


图 4.2.3 分类精度随未标记样本变化（各类别初始标记样本数为 10）

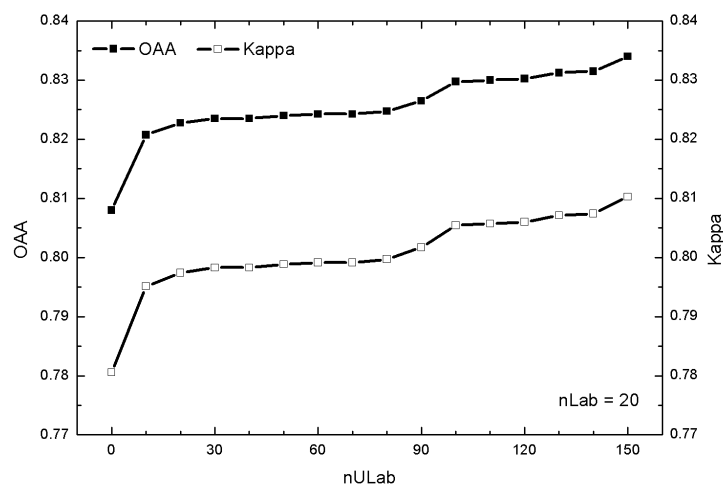


图 4.2.4 分类精度随未标记样本变化（各类别初始标记样本数为 20）

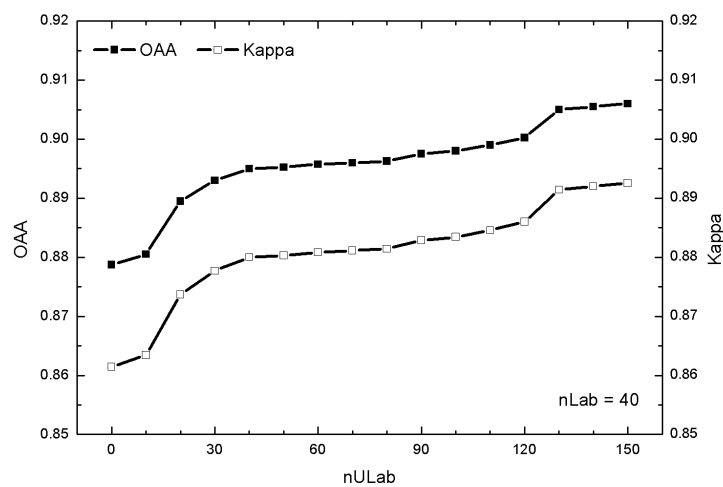


图 4.2.5 分类精度随未标记样本数变化（各类别初始标记样本数为 40）

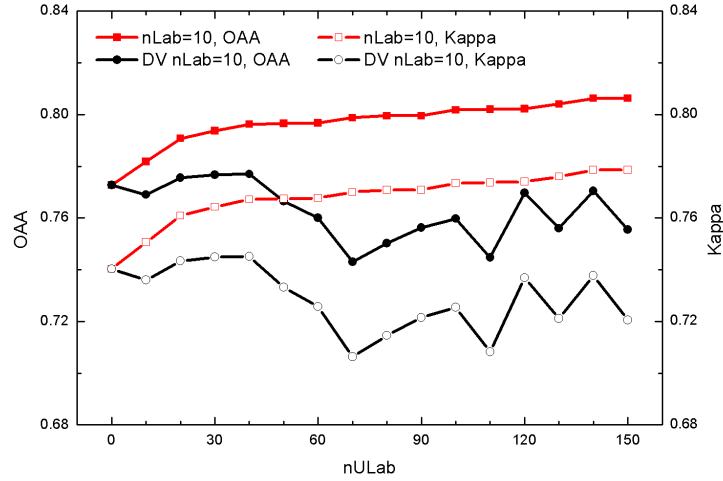


图 4.2.6 利用 SAD-ED 双重判定与利用决策值 (DV) 的 PTSVM 分类精度

从图 4.2.3-图 4.2.5 可以看出，SAD/ED-PTSVM 分类算法可以在不需要额外标记样本的情况下，有效提高分类精度，同时针对不同数量的初始标记样本，分类精度均随未标记样本数量增加而提高。

从图 4.2.6 中可以看出，SAD/ED-PTSVM 分类算法，与单纯依靠决策值的 PTSVM 分类算法，在初始标记样本与未标记样本均相同的情况下，前者分类精度随未标记样本数量稳步提高，而后者在未标记样本较少时与不添加未标记样本相比略有提高，随着未标记样本的增加，分类精度有所下降且波动较大，非常不稳定。

4.2.4 本节小结与讨论

本节主要提出了基于 SAD-ED 双重判定的渐进直推支持向量机分类算法 (SAD/ED-PTSVM)，并以示例数据进行实验，测试了其对于不同初始标记样本数量与未标记样本数量的分类精度及精度变化趋势，同时与单纯利用决策函数值的 PTSVM 分类结果进行了比较，验证了本文所提算法的有效性。

SAD/ED-PTSVM 分类算法保留了传统 PTSVM 中“成对标记”与“标签重置”的特点，同时有效的利用了高光谱数据中“图谱合一”的特点，以 SAD 与 ED 进行双重判定，为未标记样本添加标注，有效的减少未标记样本的误标记概率，从而降低后期标签重置带来的时间成本。

另一方面，由于本文为未标记样本添加的标注已经过双重判定，标记准确度

较高,所以对后添加样本与初始标记样本采用相同的惩罚系数,可以有效的简化传统 TSVM 的参数设置,减少参数优化带来的时间成本,提高分类效率。

4.3 结合混合像元分解进行主动学习的支持向量机分类

近年来,主动学习逐渐成为机器学习领域的研究热点。主动学习的基本思想是:允许学习算法在大量未标记样本中通过询问机制来选取信息量最丰富的样本,通过人机交互进行类别标注,以此来实现用最少的标记样本获得最高分类精度的目的。与传统被动学习模式相比,主动学习策略可以有效的降低样本复杂度,提高运算效率,使训练样本的类别标注可以发挥更大的作用。

在高光谱遥感植被精细分类方面,考虑到遥感图像覆盖区域广、标记样本获取困难等特点,基于主动学习的分类算法可以有效提高野外调查的质量和效率,降低人力物力消耗,通过利用有限的标记样本尽可能提高植被精细分类的准确度,在高光谱遥感植被精细分类方面具有重要的应用意义。

本节将首先介绍混合像元分解与主动学习算法的基本原理,在此基础上结合高光谱遥感数据特点,提出结合混合像元分解进行主动学习的支持向量机分类算法,并通过图 2.2.1 中包含 8 个植被类别的 PHI 实例数据测试分类器性能,最后对改进算法的意义、适用性以及不足进行讨论。

4.3.1 混合像元分解与主动学习基本原理

与被动学习随机选取样本的方式相比,主动学习(Active Learning, AL)算法经过一定的判别选择出信息量最大、对改善分类性能最有帮助的未标记样本给用户,主动与用户进行交互,而不是被动的接受数据。AL 的一般过程如下(张江红, 2011):

第一步:设置已标记样本集为 L , 待标记的未标记样本集为 U , 每次从 U 中采样的个数为 l , 分类器为 f ;

第二步:从候选样本集 U 中选择 N 个样本并标记其类别,构造初始训练样本集 L , 使 L 中每类至少包含一个样本 $L = L \cup \{(x_i, l(x_i))\}_{i=1}^N$, $U = U / \{(x_i, l(x_i))\}_{i=1}^N$;

第三步:根据已标记样本集 L 构造初始分类器 f ;

第四步：用 f 对 U 中所有样本进行划分，获得各样本点标注；

第五步：根据一定的算法从待标记样本集 U 中获得信息量最丰富的样本 x^* ，并获得其相应标签 $l(x^*)$ ；

第六步：更新已标记样本集 L 和待标记样本集 U ，令 $L = L \cup \{(x^*, l(x^*))\}_{i=1}^N$ ， $U = U / \{(x^*, l(x^*))\}_{i=1}^N$ ；

第七步：若分类精度满足需求，则算法终止；否则，计算 f ，返回第三步。

从算法流程中可以看出，如何从待标记样本集中获得最有信息量的样本点，是 AL 研究的核心问题。目前常用的策略包括：基于委员会的策略、基于最大间隔的策略和基于后验概率的策略（Tuia 等，2011）。在高光谱遥感图像分类中，应用较多的是基于后验概率的 AL 策略（Tuia 等，2011；Li 等，2010；Li 等，2011）。

本文从高光谱遥感图像中地物的分布特点出发，考虑到实际分类中最易误分、对决策面影响最大的主要是混合像元这一问题，利用混合像元分解（Spectrum Unmixing, SU）技术，同时结合在高光谱数据分类中性能卓越的 SVM，来进行 AL，选取出对分类器最有帮助的样本来提高分类精度。

SU 技术主要用于分析混合像元中各类别成分所占的比例，是高光谱数据分析中最基本的内容之一。相比于非线性光谱混合模型，线性混合模型具有模型简单、物理含义明确等特点。最简单的线性混合模型基本原理如下（Bioucas-Dias 等，2012）：对于任一像元 y ，其在第 i 波段（ $i \in \{1, \dots, B\}$ ， B 为波段数）的光谱 y_i 可以表示为：

$$y_i = \sum_{j=1}^p \rho_{ij} \alpha_j + w_i \quad (4-3-1)$$

其中， $\rho_{ij} \geq 0$ 为端元 $j \in \{1, \dots, p\}$ 在第 i 波段的特征值， $\alpha_j \geq 0$ 为端元 j 的丰度值， w_i 为误差项， p 为端元数目。线性混合模型一般可以分为无约束线性混合模型、半约束线性混合模型和全约束线性混合模型（Bioucas-Dias 等，2012）。公式(4-3-1)即为无约束线性混合模型。考虑到端元比例的实际物理含义，通常会加上一些约

束条件，即非负性约束（Abundance Nonnegativity Constraint, ANC）与和为 1 约束（Abundance Sum Constraint, ASC），满足如下约束条件的混合模型即为全约束混合模型：

$$\text{ANC: } \alpha_{ij} \geq 0, \quad j=1, \dots, p \quad (4-3-2)$$

$$\text{ASC: } \sum_{j=1}^p \alpha_j = 1 \quad (4-3-3)$$

由于在高光谱图像分类中，可以假定图像中任一像元一定属于且仅属于各类别中的一类，所以在利用混合像元分解技术辅助高光谱遥感图像分类研究中，我们选择全约束混合模型，并利用各类别标记样本平均光谱作为端元光谱。在解混算法选择方面，本文采用 Heinz 和 Chang（2001）提出的全约束最小二乘（Fully Constrained Least Squares, FCLS）线性光谱混合分析方法提取丰度信息。

全约束最小二乘求解方法首先在原方程组里添加 ASC 约束条件（4-3-3），并给予其很大的权重，然后利用 ANC 约束条件对新的方程组进行求解，使求得的解能够近似满足两个约束条件。该算法可以同时满足两个约束条件且运算效率较高，所以得到广泛的认可。

4.3.2 结合混合像元分解进行主动学习的支持向量机分类

在高光谱遥感数据处理中，混合像元分解（SU）与图像分类一直是重要的研究内容，SU 技术可以给出各个像元的各组分含量，这对于改善图像分类效果具有重要意义。而目前的研究中，结合 SU 进行高光谱图像分类还不多见。近年来，机器学习领域发展起来的 AL，通过一定的策略选择出信息量最大、对分类精度提高最有帮助的训练样本交由用户进行标注，通过人机交互改善分类结果，这对于降低标注工作量、提高分类效率具有重要作用。

考虑到高光谱遥感图像分类中，最难区分的通常是多种组分丰度比较平均、或主要由两种组分构成且二者丰度差别较小的混合像元，因此本文提出了一种结合 SU 进行 AL 的 SVM 分类算法（SUAL-SVM）。该算法首先根据初始标记样本对整幅图像进行解混，得到各组分丰度信息，然后利用丰度信息，根据预先设置的权重、提取最难区分的两种类型混合像元，同时利用 SVM 分划函数对提取的

像元进行筛选，得到对分类最有帮助的样本，进行人机交互标注，以此提高分类精度和分类效率。

SUAL-SVM 分类原理如下：

首先给定一组初始标记训练样本点：

$$(x_1, y_1), \dots, (x_n, y_n), \text{ 其中 } x_i \in R^m, y_i \in \{c1, c2, \dots, cN\} \quad (4-3-4)$$

和待标记样本集：

$$x_1^*, x_2^*, x_3^*, \dots, x_k^*, \text{ 其中 } x_i^* \in R^m \quad (4-3-5)$$

以及待标记样本数量 nAL 和权重系数 λ 。根据初始的标记训练样本，分别计算各类平均光谱曲线 $\bar{x}_{c1}, \bar{x}_{c2}, \dots, \bar{x}_{cN}$ ，并利用这些平均光谱曲线作为端元光谱，对所有待标记样本点进行全约束最小二乘线性解混，得到各点丰度信息：

$$x_{i,y_i}^* = abundance(x_i^*, y_i), \text{ 其中 } x_i^* \in R^m, y_i \in \{c1, c2, \dots, cN\} \quad (4-3-6)$$

根据丰度信息进行两方面计算：

计算各点最大最小丰度之差： $\max(abundance(x_i^*)) - \min(abundance(x_i^*))$

按照二者之差将各待标记样本由小到大排序，取前 $nAL \times \lambda$ 个样本（丰度最平均的样本）加入到主动学习样本集中；

另一方面，取各点丰度值最高的两个类别，按照这两个类别进行分组，如 $c1_c2, c1_c3, \dots, cN-1_cN$ ，类别数为 N 时，可分为 $CP = \frac{N \times (N-1)}{2}$ 组，对每组分别计算丰度最大的两个类别之间的丰度值之差，按差值从小到大进行排序，各组分别选择前 $nAL \times (1-\lambda)$ 个样本（最易误分样本），进行以下筛选：

根据当前已标记样本 $(x_1, y_1), \dots, (x_n, y_n)$ ，利用 1-V-1 SVM 计算各类对决策函数，对于“最易误分样本”，保留其位于分划边界区域内的样本，舍弃其他样本。依据各组保留下来的样本数量 NCP_i ，按比例选择前 NCP_i^* 个加入到主动学习样本集中：

$$NCP_i^* = nAL \times (1-\lambda) \times NCP_i / \sum_{i=1}^{CP} NCP_i \quad (4-3-7)$$

将 AL 样本集加入到已标记样本集中，更新已标记样本集。对于待标记的 nAL

个样本，可以一次性加入，也可以分批加入。在分批加入的情况下，设置迭代次数，然后每次都利用更新后的标记样本集重新进行丰度反演，得到“丰度最平均样本”和“最易误分样本”，并进行后续的计算，最终加入所有需要的待标记样本，得到满意的精度。

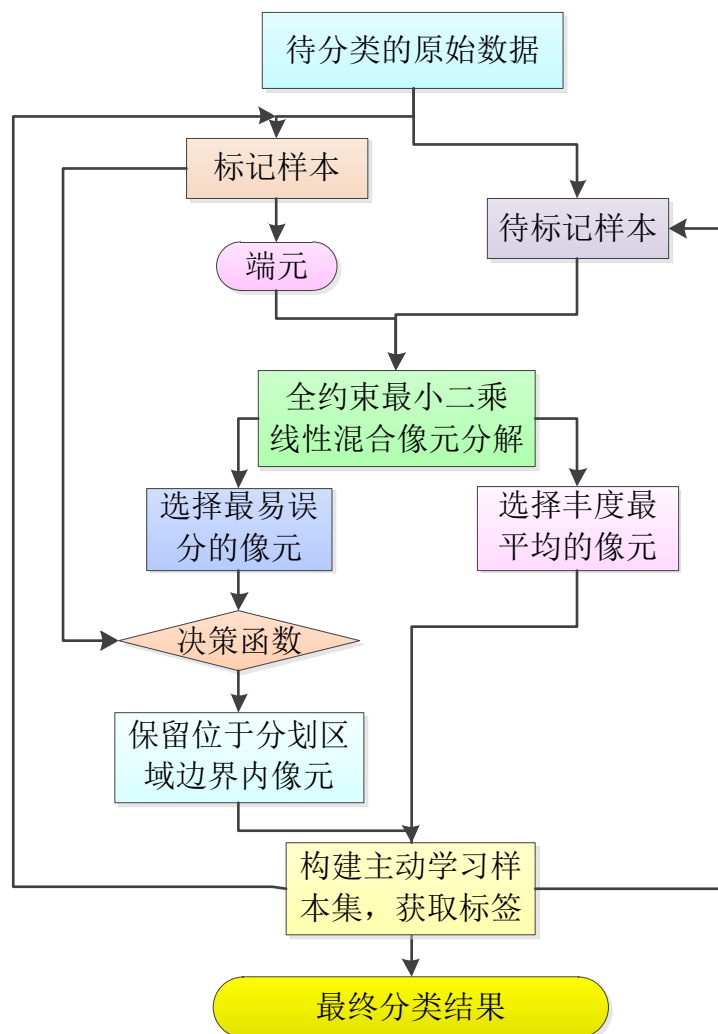


图 4.3.1 SUAL-SVM 分类算法流程图

结合混合像元分解进行主动学习的支持向量机分类算法流程如图 4.3.1，具体实现步骤如下：

第一步：构建初始标记样本集 LTS ，设置主动学习样本数量 $NALS$ ，迭代次数 $Iter$ 和权重参数 λ ；

第二步：以标记样本集 LTS 中各类别的平均光谱曲线作为端元，进行全约束最小二乘混合像元分解；

第三步：对各像元计算其最大最小丰度值之差，并按由小到大进行排序，取前 $NALS*\lambda/Iter$ 个点，添加到主动学习样本集 $ALTS$ 中；

第四步：根据第二步结果，统计各像元最主要的两个丰度对应类别，按照对应的类别分组，计算丰度最高的两个类别丰度之差，并按从小到大进行排序，保留各组前 $NALS*(1-\lambda)/Iter$ 个像元；

第五步：针对各类对利用已标记样本进行 1-V-1 支持向量机分类，对上一步各组保留的像元进行筛选，保留各组在分划边界区域内的像元，删除其它像元；

第六步：将第五步保留的像元按照各类别数量比例，选择各组排序靠前的像元，共添加 $NALS*(1-\lambda)/Iter$ 个像元到主动学习样本集 $ALTS$ 中；

第七步：更新标记样本集 $LTS = LTS \cup ALTS$ ，返回第二步并重复第三步至第七步，直到分类精度满足需求，或迭代次数达到 $Iter$ 。

4.3.3 实验与结果

为了验证 $SUAL-SVM$ 分类算法性能，本节利用图 2.2.1 包含 8 个植被类别 80 个波段的示例 PHI 数据，主要从以下几方面对该算法进行了分析和测试：

- (1) 分析不同初始标记样本数量 (nLS)、 AL 样本数量 ($nALS$) 对分类精度的影响；
- (2) 比较结合 SU 进行 AL 得到的训练样本与随机选择 (RS) 相同数量的训练样本得到的分类精度；
- (3) 分析总样本数 ($nLS+nALS$) 一定情况下，初始标记样本与 AL 样本数对分类精度影响；
- (4) 分析初始标记样本数量与 AL 样本数量一定的情况下，迭代次数 $Iter$ 对于分类精度的影响；
- (5) 分析初始标记样本数量与 AL 样本数量一定的情况下，权重系数 λ 对于分类精度的影响。

在图 4.3.2-图 4.3.5 中，权重系数均为 0.5，迭代次数随 AL 样本数量增加而增加，即 AL 样本数为 0-100 时，迭代次数为 1； AL 样本数在 101-200 时，迭代次数为 2； AL 样本数在 201-300 时，迭代次数为 3； AL 样本数在 301-400 时，

迭代次数为4。

图 4.3.6 与图 4.3.7 中，初始标记样本数目与 AL 样本数目分别为 40 和 400，图 4.3.6 中权重系数为 0.5，图 4.3.7 中迭代次数为 4。

在参数优化方面，各参数优化策略搜索的最优 C 区间为 0.1 到 100，最优 γ 区间为 0.01 到 1000，分类精度见图 4.3.2-图 4.3.7。

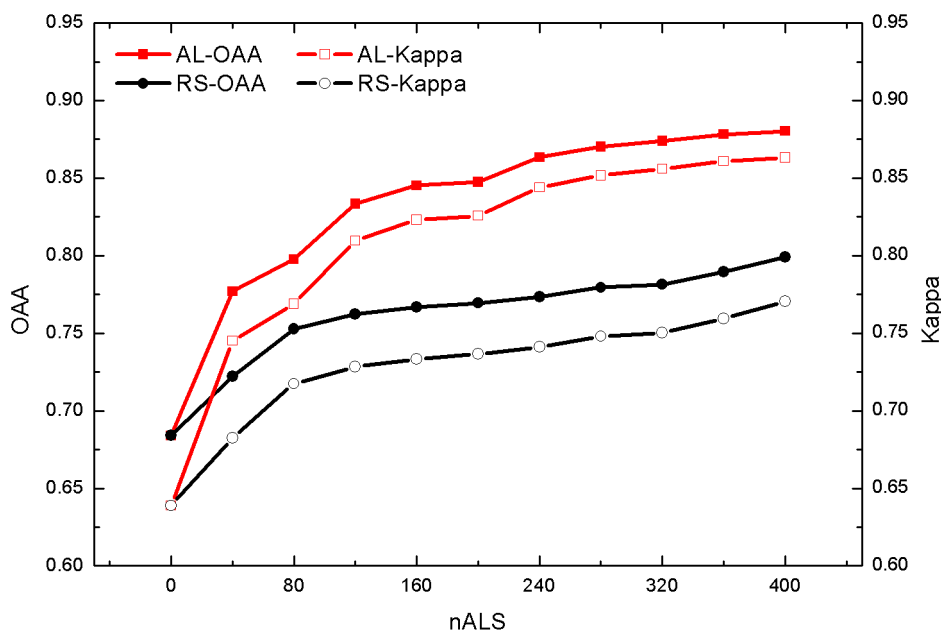


图 4.3.2 分类精度随 AL 样本数量变化（所有类别初始标记样本总数为 40）

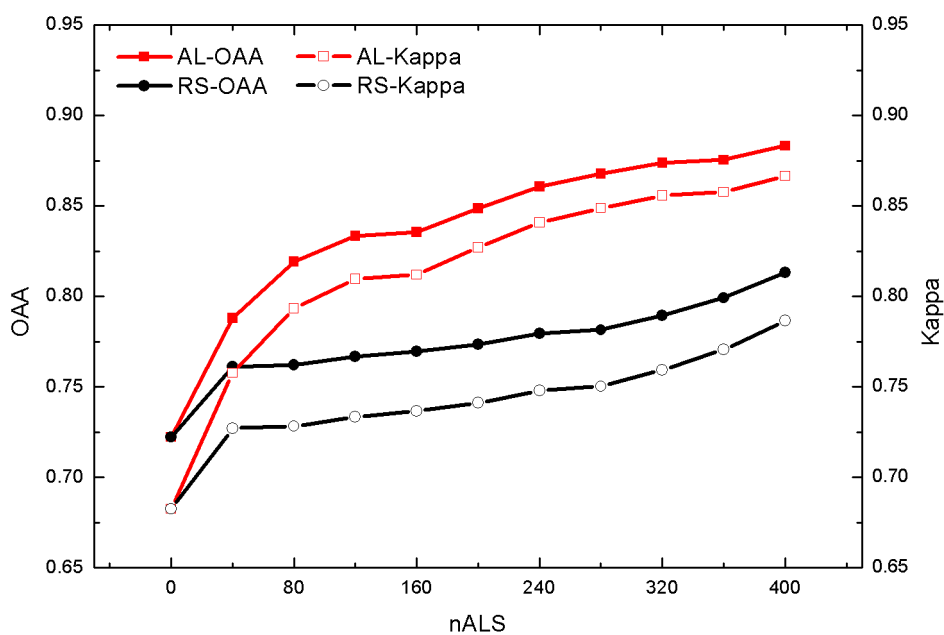


图 4.3.3 分类精度随 AL 样本数量变化（所有类别初始标记样本总数为 80）

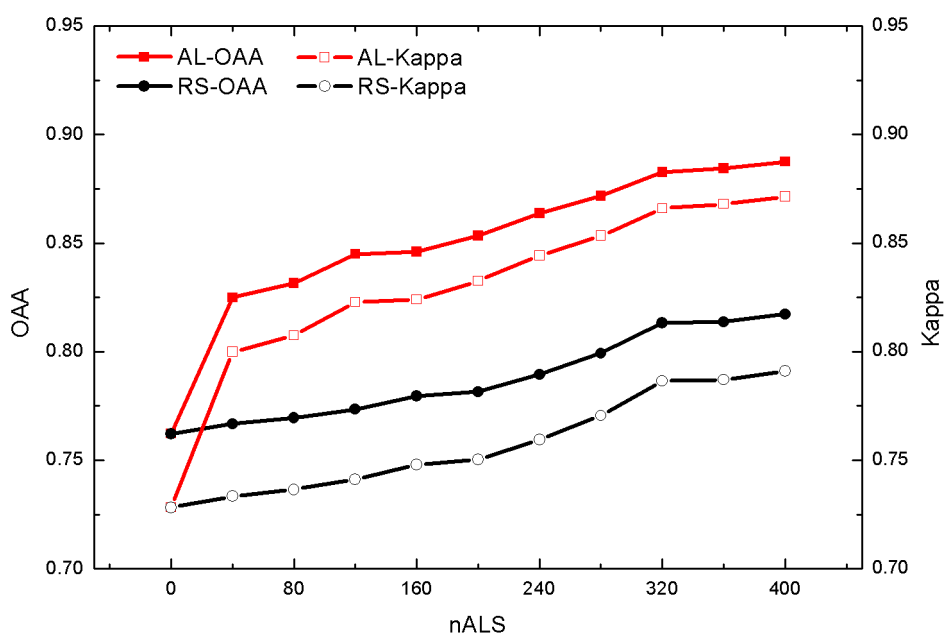


图 4.3.4 分类精度随 AL 样本数量变化（所有类别初始标记样本总数为 160）

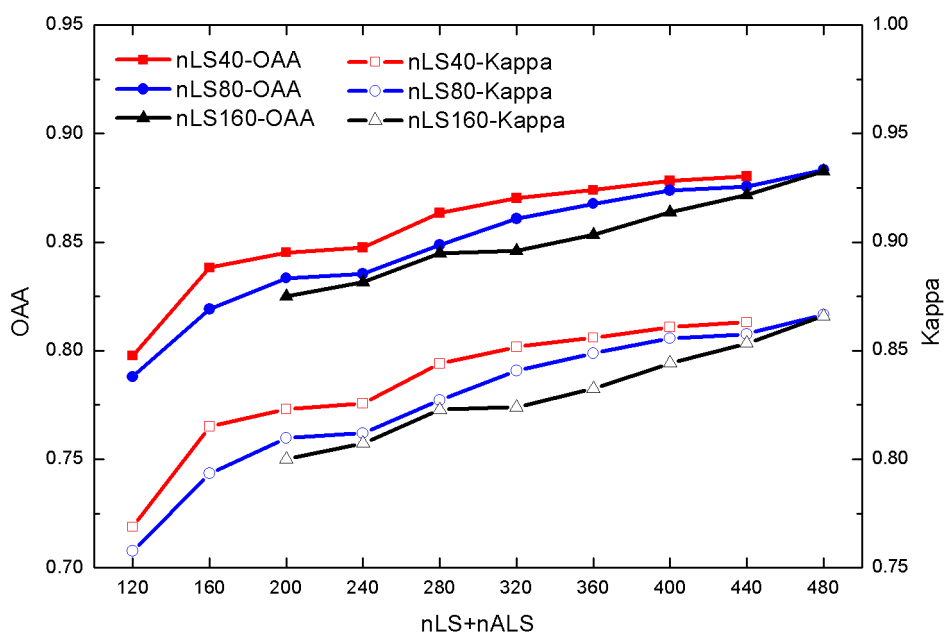


图 4.3.5 总样本数一定情况下，初始标记样本与 AL 样本数对分类精度影响
（图中横坐标为总样本数目，红、蓝、黑分别对应初始标记样本数目 40、80、160，总数目减初始标记样本数目即为各点对应 AL 样本数目）

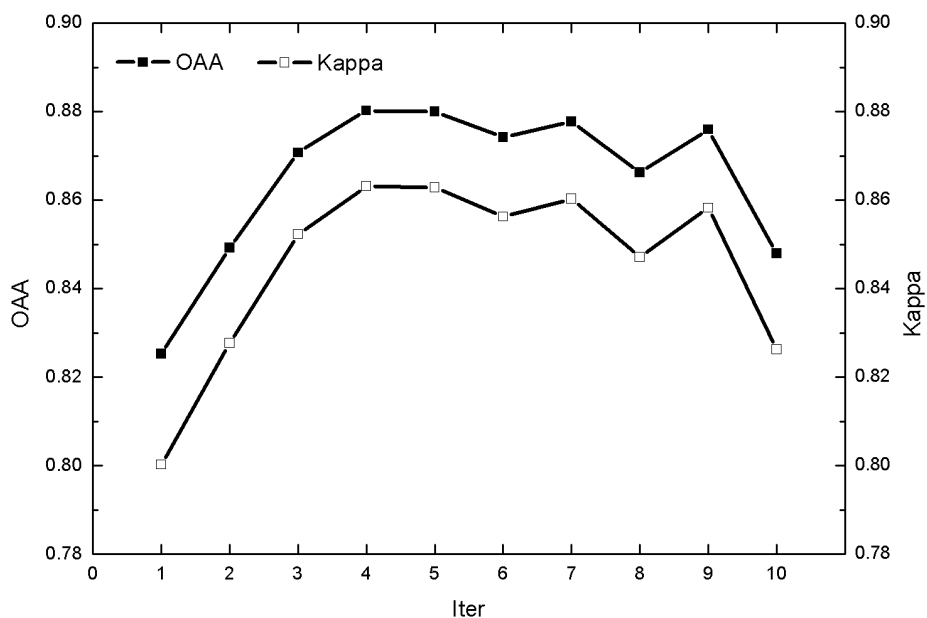


图 4.3.6 迭代次数对分类精度影响
（初始标记样本与 AL 样本数目分别为 40 和 400）

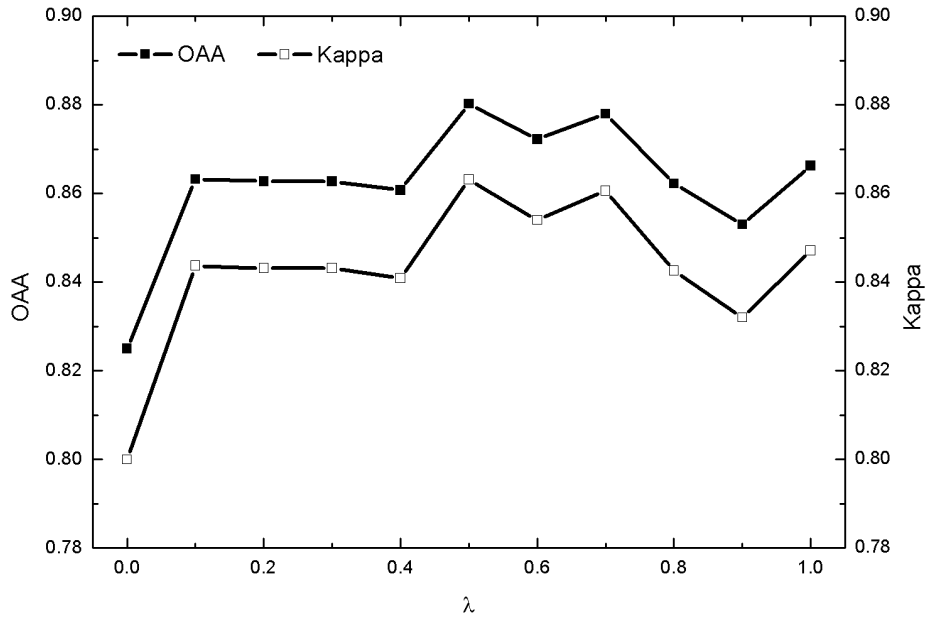


图 4.3.7 权重系数对分类精度影响

(初始标记样本与 AL 样本数目分别为 40 和 400)

从图 4.3.2-图 4.3.4 中可以看出, SUAL-SVM 分类算法与随机选择训练样本的分类结果相比, 总体精度与 Kappa 系数具有大幅度提高; 同时, 在初始标记样本一定的情况下, 分类精度随 AL 样本数量增加先是显著提高, 后趋于平稳。

从图 4.3.5 中可以看出, 在总样本数目一定的情况下, AL 样本所占比例越高, 分类精度越高, 该结果从侧面反映了本文所提算法的有效性。

图 4.3.6 反映了迭代次数对分类精度的影响。可以看出, 并非迭代次数越多分类精度越好, 当迭代次数在 4 至 5 次 (单次主动学习样本数为 80 至 100) 时, 取得了最优分类精度, 同时可以看出当迭代次数在 1 至 10 区间、以 1 为间隔变化时, 分类精度的变化区间为 82.5% 到 88%, 可见该算法的分类精度对于迭代次数变化是非常敏感的。

图 4.3.7 反映了权重系数对于分类精度的影响。可以看出, 权重系数为 0.5 时分类精度最高, 权重系数为 0 时 (只加入最易误分样本而不考虑丰度最平均样本) 分类精度最低, 这从一定程度上反映了 SUAL-SVM 分类算法的有效性。当权重系数在 0.0-1.0 区间以 0.1 间隔变化时, 分类精度变化区间约为 82.5%-88%, 可见该算法的分类精度对于权重系数的变化也是较为敏感的。

4.3.4 本节小结与讨论

本节提出了结合混合像元分解进行主动学习的支持向量机 (SUAL-SVM) 分类算法, 并以示例数据进行实验, 测试其对于不同初始标记样本数量与主动学习样本数量的分类精度及精度变化趋势, 并与随机选择相同数目训练样本进行支持向量机分类的结果进行对比; 同时测试在总样本数目一定的情况下, 不同初始标记样本与主动学习样本所占比例对分类精度的影响。由于本文所提算法在初始时需要进行迭代次数与权重系数的设置, 所以对于实验中初始标记样本与主动学习样本数目一定的情况, 也分析了迭代次数与权重系数对分类精度的影响。

实验结果表明, 与随机选择训练样本分类结果相比, SUAL-SVM 算法可以利用更少的标记样本实现更高的分类精度, 有效的减轻样本标记的工作量, 减少分类器样本训练所需时间; 在参数设置方面, SUAL-SVM 算法的分类精度对于权重系数与迭代次数变化较为敏感, 推荐权重系数为 0.5 左右, 最佳迭代次数取决于单次添加的主动学习样本数目, 单次添加的主动学习样本数目以 80 到 100 为宜。

SUAL-SVM 分类算法有效地将高光谱研究中的混合像元分解技术与机器学习领域中的主动学习策略进行有机结合, 充分利用了各像元的丰度信息来辅助分类决策; 同时, 算法中设置了权重系数, 用来综合考虑两类类别归属最困难的像元, 即丰度最均匀像元与最易误分像元, 使得筛选出来的主动学习样本能够有效优化分类决策函数; 另外, 由于不同类别之间的区分难易度不同, SUAL-SVM 算法在添加最易误分像元时, 利用分划边界函数进行优选, 根据各类对分划边界区域内剩余的主动学习样本数量, 按比例添加, 使各类对的主动学习样本数目按类对区分难易度由多到少不等, 该方法使得添加的主动学习样本更加具有针对性。

4.4 本章小结

在传统高光谱植被精细分类研究中, 通常利用监督分类算法, 该类算法对于标记训练样本具有较高的要求, 而标记样本的获取耗时耗力。因此, 本文在分析传统支持向量机原理基础上, 结合半监督学习算法, 提出了两种基于支持向量机的半监督植被精细分类算法, 即: 基于 SAD-ED 双重判定的渐进直推支持向量

机 (SAD/ED-PTSVM) 分类算法与结合混合像元分解进行主动学习的直推支持向量机 (SUAL-SVM) 分类算法。

SAD/ED-PTSVM 分类算法, 是在传统 PTSVM 样本标注中, 利用 SAD 与 ED 分别对未标记样本标签进行判定, 同时结合其到分划边界的距离对未标记样本进行“自动”标注。该算法保留了传统 PTSVM 中“成对标记”与“标签重置”的特点, 有效利用了高光谱数据中“图谱合一”的特点, 减少未标记样本的误标记概率, 从而降低后期标签重置带来的时间成本, 同时有效的简化传统 TSVM 的参数设置, 减少参数优化所需的时间, 提高分类效率。

SUAL-SVM 分类算法, 首先根据初始标记样本对整幅图像进行解混, 得到各组分丰度信息, 然后利用丰度信息, 根据预先设置的权重、提取“丰度最均匀像元”与“最易误分像元”, 同时利用 SVM 分划函数对提取的像元进行筛选, 得到对分类最有帮助的样本, 进行“人机交互”标注, 以此提高分类精度和分类效率。在参数设置方面, SUAL-SVM 分类精度对于权重系数与迭代次数变化较为敏感。与随机选择训练样本相比, SUAL-SVM 可利用更少的标记样本实现更高的分类精度, 有效减轻样本标记工作量, 减少分类器样本训练所需时间。

第5章 基于地面成像光谱数据的作物/杂草精细分类

地面成像光谱数据在具备高光谱分辨率的同时，具有极高的空间分辨率，可以支持地面尺度的植被精细识别。目前，针对地面尺度的高光谱植被精细识别研究正处于起步阶段。本章主要验证本文所提的基于特征库构建及优化分类算法、基于支持向量机的半监督分类优化算法，在地面尺度高光谱遥感植被精细分类中的有效性，以及在杂草识别、作物产量精细估算等方面的应用潜力。

本章具体可以分为以下几个部分：首先，介绍实验所用的地面成像光谱系统；其次，介绍地面高光谱数据的预处理及研究区概况；然后，分别将本文所提特征库光谱维与空间维优化算法与其他特征选择及滤波算法进行比较，再利用所提出的两种基于支持向量机的半监督分类优化算法进行植被精细分类，并与其他分类结果进行比较分析；最后，提出一套适用于地面尺度高光谱遥感数据植被精细分类的方法流程，为地面高光谱遥感数据在精准农业中更广泛的应用提供参考。

5.1 地面成像光谱数据预处理

本章所用的地面尺度高光谱遥感图像由地面成像光谱辐射测量系统（Field Imaging Spectrometer System, FISS）获取。FISS 系统是我国第一套自行研制开发的地面成像光谱系统，由中国科学院遥感应用研究所与上海技术物理所共同研制开发。

FISS 的成像方式如下图 5.1.1 所示（童庆禧等，2010；方俊永等，2009）。

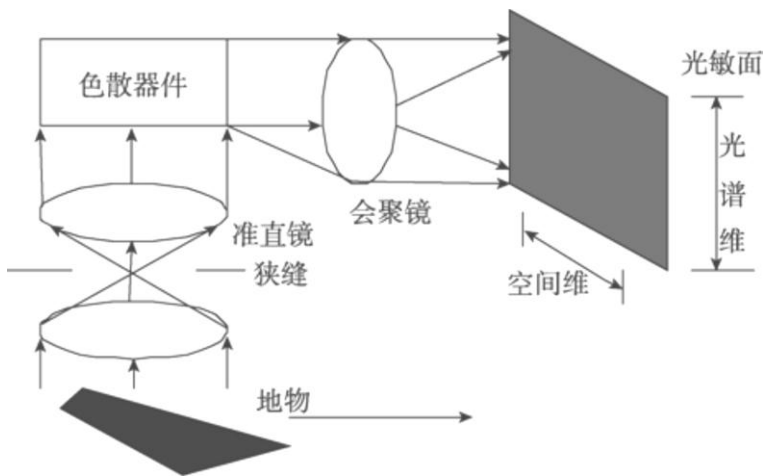


图 5.1.1 FISS 原理图（童庆禧等，2010）

FISS 主要由五部分构成：多用途平台、电子学系统、光机系统、计算机系统和辅助设备等（刘波，2010）。FISS 的光机系统实物图见图 5.1.2，野外测量示意图和实际测量工作图见图 5.1.3。



图 5.1.2 FISS 光机系统实物图（刘波，2010）

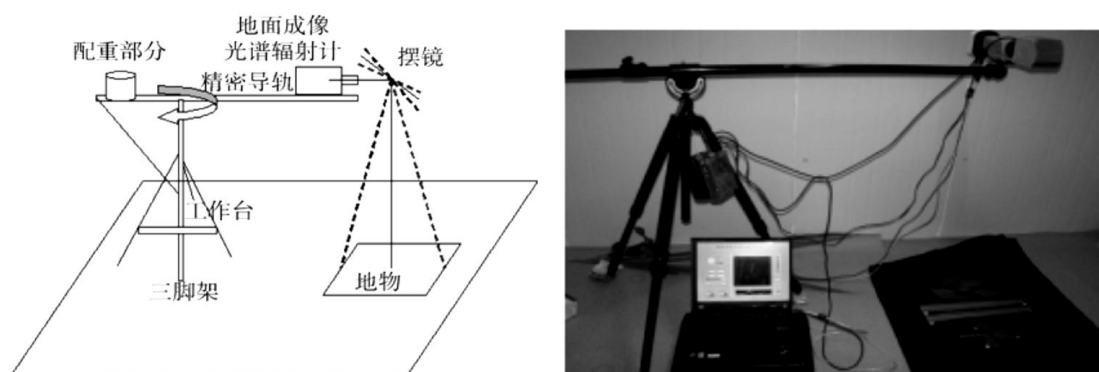


图 5.1.3 FISS 野外测量示意图和实际测量工作图（童庆禧等，2010）

FISS 的主要指标参数包括光谱分辨率、光谱范围、空间分辨率和帧速率等，具体的指标参数和性能见表 5.1.1。

表 5.1.1 FISS 主要指标参数和性能（童庆禧等，2010）

指标	参数	指标	参数
工作方式	面阵 CCD 推扫式	信噪比	60%波段>500
扫描视场	-20° 至 20°	瞬时视场角	1.0mrad
波段数	344	光谱范围	379nm-870nm
光谱分辨率	4-7nm	空间维	464
光谱采样	约 1.4nm	量化位数	12 位
成像速率	最高 20 帧/秒	实验室辐射定标精度	优于 5%

本章所用的地面成像光谱数据为 2009 年 7 月 22 日在中国农业大学西校区实验田获取的数据，成像时间为中午 12 点 34 分，采取垂直观测方式，仪器距离地

面高度约为 1.2m，对应的图像空间分辨率约 1.2mm。

实验中对原始数据进行了光谱定标、辐射校正和几何校正等基本预处理，具体方式见文献（童庆禧等，2010）。由于是地面尺度成像，所以不需要进行大气校正。在利用 FISS 进行影像获取的同时，对地面植被分布进行了拍照记录，如图 5.1.4。



图 5.1.4 实验场景及实际植被分布情况

研究区共包含 4 种植被类型，分别是西瓜苗、牛筋草、野苋菜和马齿苋。研究区的 FISS 影像以及实际植被分布如下图。

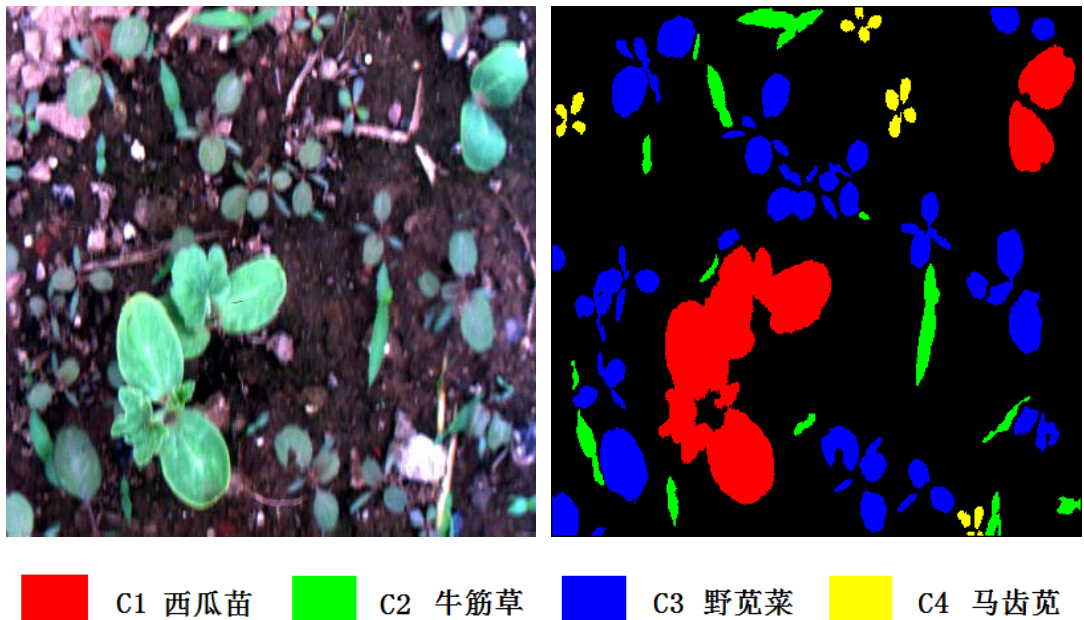


图 5.1.5 研究区的 FISS 影像（真彩色合成）及实际植被分布图

在分类精度评价方面，主要根据实际植被分布图对整幅影像进行计算，比较总体精度（OAA）、Kappa 系数（KC），以及各类别的生产者精度（PA）和用户精度（UA）等。

5.2 植被特征库构建与优化

在本节实验中，所用的训练样本为基于实际植被分布图（图 5.1.5）各类别随机生成的 25 个训练样本，且植被特征库光谱维优化中所用的训练样本，与本节后续分类实验中所用训练样本相同。

5.2.1 植被特征库构建

在植被特征库构建中，原始波段为反射率数据；空间纹理特征的获取则是首先对原始光谱波段进行主成分分析，选取前 3 个主成分（累积贡献率超过 98%）分别计算 10 个纹理特征，得到 30 个纹理波段；光谱指数特征直接利用原始光谱波段间计算获得，共计 50 个光谱指数特征。构建的植被特征库具体组成如表 5.2.1。

表 5.2.1 FISS 影像植被特征库对应波段

波段号	对应波段
1-344	原始反射率波段
345-351	第 1 主成分对应 7 个纹理波段，依次为： 均值、均质性、对比度、异质性、熵、二阶矩和相关性
352-358	第 2 主成分对应 7 个纹理波段（次序同上）
359-365	第 3 主成分对应 7 个纹理波段（次序同上）
366-368	第 1-3 主成分对应的 Moran 指数波段
369-371	第 1-3 主成分对应的 Geary 系数波段
372-374	第 1-3 主成分对应的 G 统计量波段
375-424	表 3.1.1 中所列的 50 个光谱指数波段

各植被类别在特征库中各波段（已经过归一化处理）的均值图 5.2.1 所示。

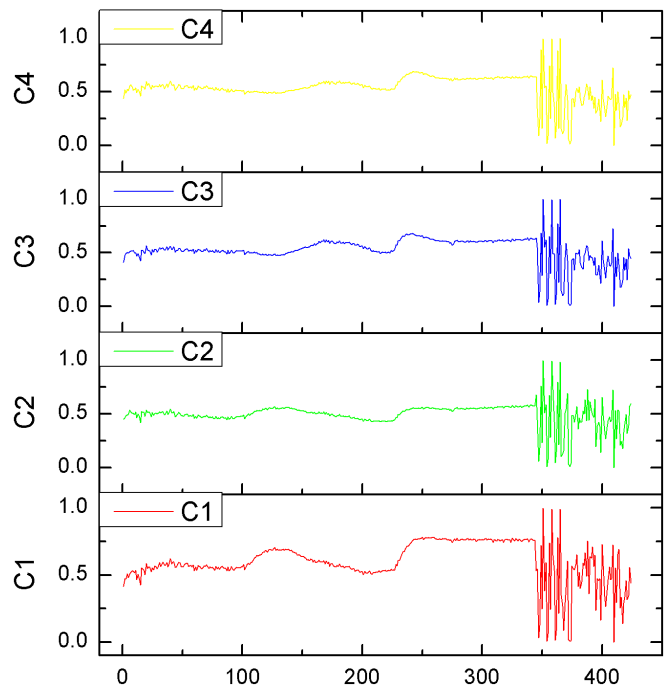


图 5.2.1 各植被类别在特征库中各波段上的均值分布

5.2.2植被特征库优化

在对所构建的植被特征库进行光谱维优化时，根据前面随机选择的各类别 25 个训练样本，利用第三章介绍的基于类对可分性的植被特征库光谱维优化算法，得到光谱维优化后的植被特征库。光谱维优化后最终保留了 28 个植被特征，其中原始波段 10 个，纹理特征 8 个，光谱指数特征 10 个。光谱维优化后保留的植被特征波段具体见表 5.2.2。

表 5.2.2 光谱维优化后保留的植被特征波段

具体波段信息	
原始特征	对应波长(nm): 418.8, 562.8, 581.5, 628.0, 683.9, 725.3, 747.6, 769.8, 772.8, 795.0
纹理特征	第一主成分: 相关性、Moran 指数、G 统计量 第二主成分: 二阶矩、相关性、Moran 指数 第三主成分: 相关性、Moran 指数
光谱指数特征	表 3.1.1 中序号分别为: 13, 14, 18, 29, 32, 35, 36, 42, 43, 48

在进行植被特征库空间维优化时，首先计算各类别训练样本的平均光谱角，具体计算结果如图 5.2.2，取其中最小值确定参数，即 $MinSAD$ 为 0.1047，同时设置窗口大小为 9×9 ，迭代次数为 2 次，之后按照第 3 章第 3 节中介绍的步骤进行基于邻域 SAD 的植被特征库空间维优化操作。

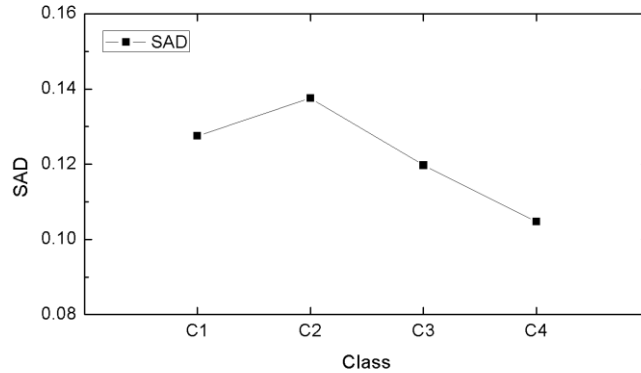


图 5.2.2 各类别训练样本的平均光谱角距离（SAD）

5.2.3 基于植被特征库构建与优化的地面高光谱数据分类

为了客观的验证基于植被特征库构建与优化的分类算法在地面高光谱数据植被精细分类中的有效性，本节比较了不同的特征选择方法与滤波算法，在分类器方面统一采用 ENVI 软件自带的传统支持向量机，选择 RBF 核函数，参数方面均采用 ENVI 软件默认参数，未进行参数优化。

具体的分类方法见表 5.2.3，相应的分类结果与精度如图 5.2.3 与表 5.2.4。

表 5.2.3 实验中比较的分类方法

名称	光谱维优化算法	特征波段数
CPS	基于类对可分性的光谱维优化算法	28
OIF	最优索引因子法（Chavez 等，1982）	28
名称	空间维优化算法	窗口大小
NPSAD	基于邻域 SAD 的空间维优化算法	9×9
MID	中值滤波	9×9
LP	低通滤波	9×9
ELee	加强型 Lee 滤波	9×9
EFrost	加强型 Frost 滤波	9×9

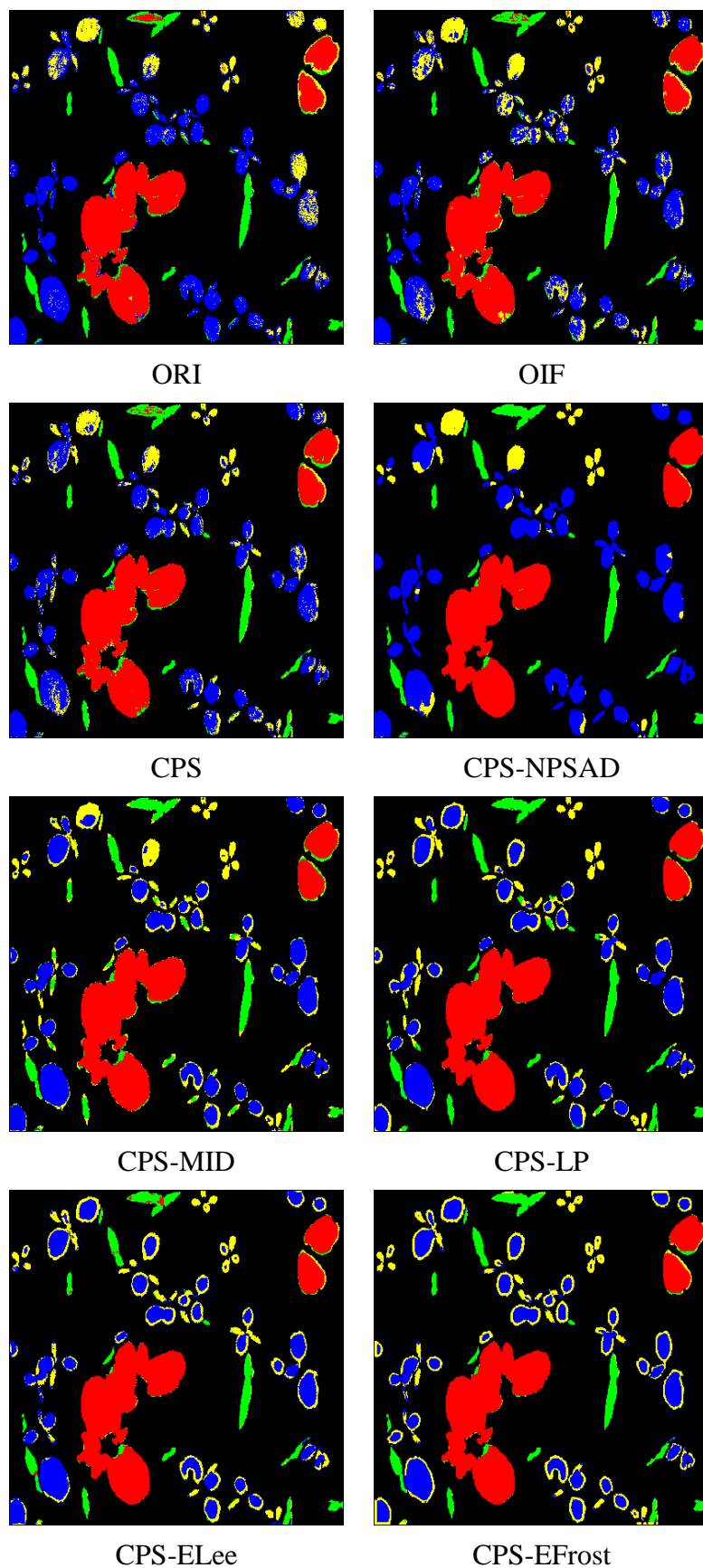


图 5.2.3 各方法对应分类结果图

注：ORI 为直接利用原始数据 344 个波段进行 SVM 分类结果

表 5.2.4 各方法对应分类结果精度表

注：ORI 为直接利用原始数据 344 个波段进行 SVM 分类结果。

	ORI		CPS		OIF		CPS-NPSAD	
	PA	UA	PA	UA	PA	UA	PA	UA
C1	90.47%	98.13%	92.55%	99.16%	91.31%	98.93%	98.18%	99.99%
C2	87.71%	78.42%	90.76%	80.45%	91.22%	79.05%	97.87%	94.15%
C3	77.30%	93.39%	74.64%	96.11%	68.08%	94.61%	85.29%	97.62%
C4	59.75%	16.45%	70.25%	17.13%	60.94%	12.59%	78.36%	29.78%
OAA(KC)	83.18%(0.7462)		83.55%(0.7546)		79.83%(0.7043)		91.67%(0.8710)	
	CPS-MID		CPS-LP		CPS-ELee		CPS-EFrost	
	PA	UA	PA	UA	PA	UA	PA	UA
C1	93.71%	100.00%	97.03%	99.96%	96.63%	99.06%	96.34%	99.94%
C2	82.12%	73.77%	98.46%	85.35%	91.52%	91.03%	94.03%	91.13%
C3	67.17%	98.95%	66.36%	99.60%	63.24%	97.04%	58.17%	97.93%
C4	91.81%	17.33%	98.25%	19.83%	82.74%	15.00%	90.21%	14.38%
OAA(KC)	80.56%(0.7173)		83.55%(0.7597)		80.68%(0.7190)		78.84%(0.6976)	

从图 5.2.3 与表 5.2.4 中可以看出，通过特征库构建及光谱维优化，仅用 28 个波段便可得到比原始 344 个波段更高的分类精度，对于作物类别(C1 西瓜苗)，生产者精度与用户精度均明显提高，这是最优索引因子法特征选择分类结果所无法比拟的，可见通过特征库构建与基于 CPS 的光谱维优化可以更好的进行作物/杂草的识别，促进精准农业的发展。同时，从分类结果图中可以看出，单纯利用特征库构建及光谱维优化的分类结果中椒盐噪声较多，植被叶片信息不完整，这在一定程度上影响地块作物/杂草的面积估算，影响精准施肥的进行。

为解决光谱维优化后分类结果椒盐噪声多的问题，本节比较了不同的空间维优化算法。从分类结果图中可以看出，除本文所提的基于 NPSAD 的空间维优化算法，其他空间维优化后的分类结果均有明显的边缘效应，这严重影响杂草识别的结果。同时，从分类精度表可以看出，基于 NPSAD 的空间维优化分类结果，对于作物类别(C1 西瓜苗)生产者精度达到 98% 以上，用户精度更是达到 99.99%，

这远远高于其他算法分类结果,可以认为,该算法能够非常准确的提取作物类别。

5.2.4 本节小结

本节利用基于植被特征库构建与优化的分类方法进行了地面尺度的高光谱植被精细分类实验,得到如下结论:

首先,通过植被特征库的构建,可以有效的增加用于作物/杂草识别的特征波段,提高作物识别的精度;

其次,与其他空间维优化算法相比,基于 NPSAD 的空间维优化算法可以利用光谱维信息进行空间维平滑,使同类地物内部更加平滑、不同地物间边缘信息得以准确的保留,有效的解决其他算法具有的边缘效应,在地面尺度植被精细分类中,可以更完整的提取叶片信息,提高作物/杂草识别精度;

最后,从本节实验可以看出,基于植被特征库构建与优化的分类方法,可以准确的提取出作物信息,提高不同类型作物/杂草的识别精度,有效提高地面高光谱遥感数据植被精细分类的精度和效率。

5.3 基于支持向量机的半监督分类

针对本文所提的基于 SAD-ED 双重判定的渐进直推支持向量机分类算法、结合混合像元分解进行主动学习的支持向量机分类算法,本节主要利用地面高光谱遥感数据进行植被精细分类实验,将这两种半监督分类算法分别与单纯基于决策值进行类别标记的直推支持向量机分类、通过随机选择确定训练样本的支持向量机分类结果进行对比,验证本文所提算法的优越性。

本节分类实验中所用的数据为上一节经过植被特征库构建与优化后,最终得到的 28 个特征波段。本节实验主要基于编写的 Matlab 程序实现,采用 RBF 核函数,惩罚系数与核参数通过网格搜索进行优化,搜索的最优 C 区间为 0.1 到 1000,最优 γ 区间为 0.01 到 1000。

5.3.1 基于 SAD-ED 双重判定的渐进直推支持向量机分类

在对地面高光谱遥感数据进行基于 SAD-ED 双重判定的渐进直推支持向量机分类 (SAD/ED-PTSVM) 实验时,初始标记样本为基于实际植被分布图 (图

5.1.5) 各类别随机生成的 25 个训练样本，未标记样本则在各类对分类时，从剩余像元中随机生成的 50 个与初始标记样本不重叠的像元。为验证该算法优势，实验中将分类结果与单纯基于决策值实现类别标记的直推支持向量机（DV-PTSVM）和传统支持向量机（SVM）的分类结果进行了对比分析。DV-PTSVM 的样本获取方法与 SAD/ED-PTSVM 相同，而 SVM 则是直接利用前面随机生成的 25 个训练样本进行分类。

具体分类结果与精度如图 5.3.1 与表 5.3.1 所示。

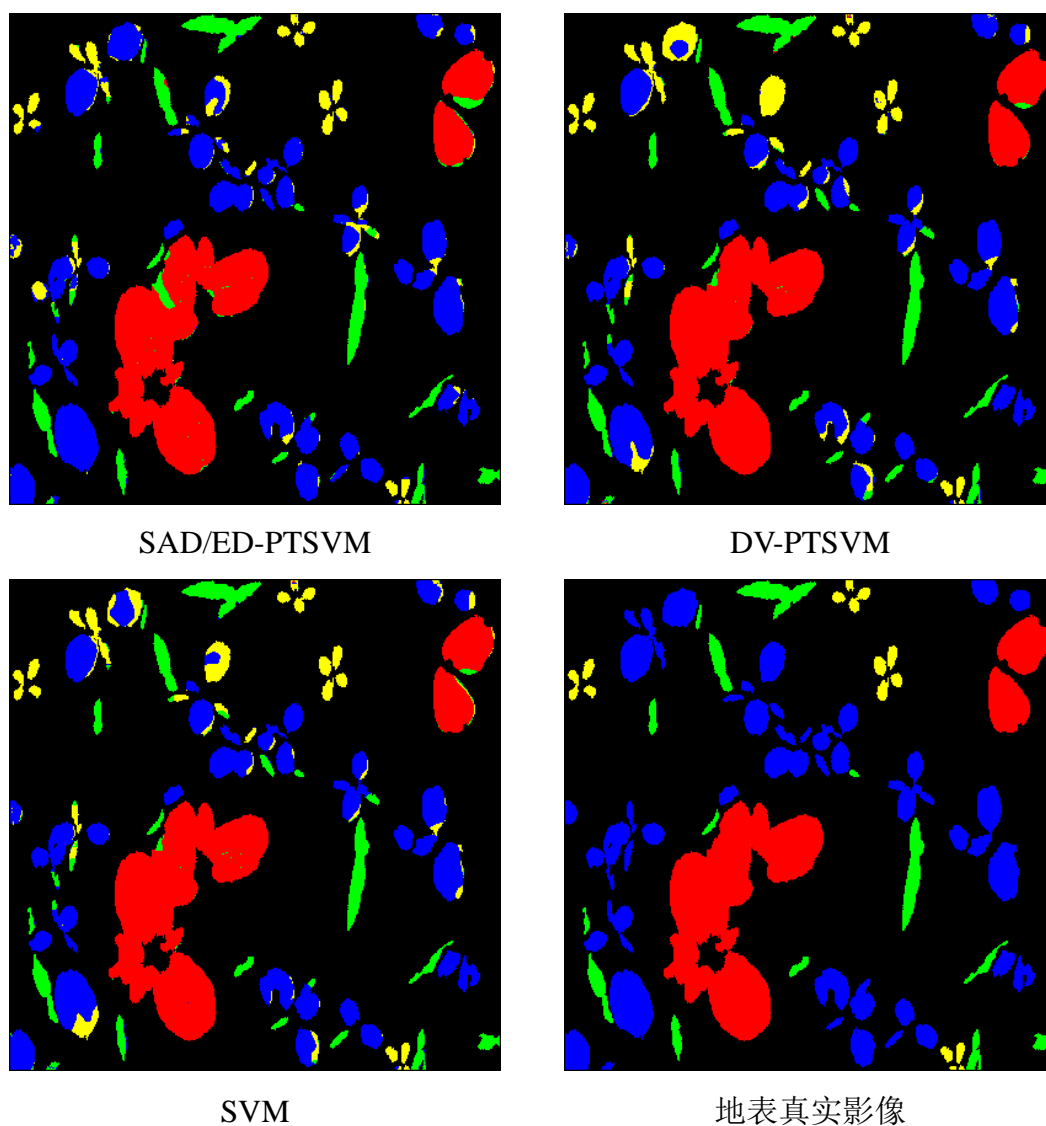


图 5.3.1 各方法对应分类结果图

表 5.3.1 各方法对应分类结果精度表

	SAD/ED-PTSVM		DV-PTSVM		SVM	
	PA	UA	PA	UA	PA	UA
C1	97.21%	99.98%	99.07%	99.84%	98.68%	99.97%
C2	98.68%	81.89%	98.22%	82.57%	98.76%	82.38%
C3	87.84%	99.51%	77.49%	99.25%	81.03%	99.50%
C4	98.01%	45.80%	96.18%	29.98%	98.33%	34.77%
OAA(KC)	93.22%(0.8947)		89.25%(0.8370)		90.80%(0.8595)	

从图 5.3.1 与表 5.3.1 可以看出,与传统 SVM 分类结果相比,SAD/ED-PTSVM 分类算法对于各类对仅添加 50 个未标记样本,便可有效的提高总体精度和 Kappa 系数;而 DV-PTSVM 分类结果与传统 SVM 结果相比,各类别精度变化不大,总体精度和 Kappa 系数稍有下降。可见,SAD/ED-PTSVM 分类算法可以有效提高未标记样本的标注准确度,进而提高分类结果的总体精度以及杂草类别(C2, C3 与 C4)之间的识别精度。

值得一提的是,由于 PTSVM 初始标记样本较少,且后续未标记样本的标注添加是在利用初始标记获得分划函数基础上进行的,所以最终分类结果受初始标记样本分布影响较大。本节实验中未标记样本主要通过随机算法生成,通过人工优选初始标记样本,可以在此基础上进一步提高分类精度。

5.3.2结合混合像元分解进行主动学习的支持向量机分类

在结合混合像元分解进行主动学习的支持向量机(SUAL-SVM)分类实验中,初始标记样本为基于实际植被分布图(图 5.1.5)各类别随机生成的 5 个训练样本,之后再利用 AL 通过人机交互补充 20 个训练样本,然后采用传统 SVM 进行分类;用来进行对比的随机选择样本(RS)分类算法中,则是直接利用实际植被分布图(图 5.1.5)随机生成 25×4=100 个训练样本,然后采用传统 SVM 进行分类。

具体的分类结果与精度如图 5.3.2 与表 5.3.2 所示。

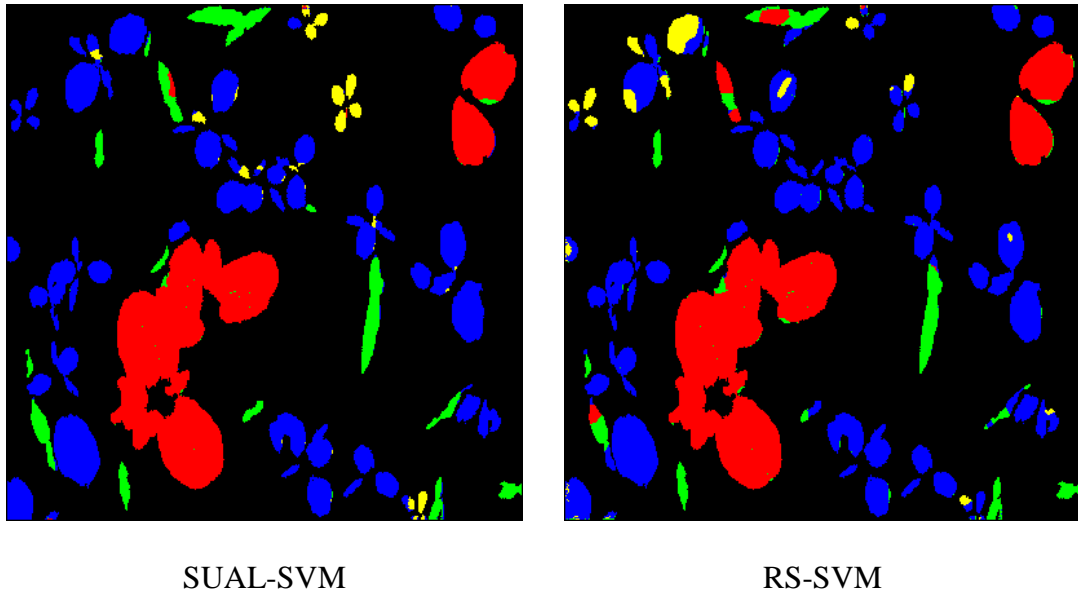


图 5.3.2 各方法对应分类结果图

表 5.3.2 各方法对应分类结果精度表

	SUAL-SVM		RS-SVM	
	PA	UA	PA	UA
C1	99.16%	99.40%	98.51%	95.59%
C2	95.06%	97.00%	68.67%	88.42%
C3	98.05%	96.20%	92.66%	91.61%
C4	62.29%	73.18%	36.44%	30.97%
OAA(KC)	96.88%(0.9494)		90.42%(0.8443)	

从图 5.3.2 与表 5.3.2 可以看出，与 RS-SVM 分类结果相比，SUAL-SVM 分类结果在总体精度与 Kappa 系数方面都有大幅度的提高，总体精度达到 96% 以上，Kappa 系数接近 95%，同时，各类别的生产者精度与用户精度也更加均匀，尤其是对于实际分布较少的 C4（马齿苋），生产者精度与用户精度均有显著提高；而利用 RS-SVM 分类，容易导致某些类别由于训练样本太少而识别度较低，如 C2（牛筋草）和 C4（马齿苋）。

从分类结果可以看出，SUAL-SVM 算法选择出的训练样本更具有针对性，不仅可以更为准确的提取出作物类别（C1 西瓜苗），对于不同的杂草类型，也可以更加完整、准确的进行识别，尤其是对于实际分布较少的类别，可以显著改善

识别精度。

5.4 本章小结

本章针对所提的特征库构建与优化算法、基于 SAD-ED 双重判定的渐进直推支持向量机 (SAD/ED-PTSVM) 分类算法与结合混合像元分解进行主动学习的支持向量机 (SUAL-SVM) 分类算法, 以真实地面成像光谱数据进行实验, 验证了这些算法在地面尺度作物/杂草种类识别、叶片提取等方面的有效性。

本章的实验结果表明, 基于植被特征库构建与优化的分类算法可以有效提高地面尺度高光谱遥感影像的植被分类精度, 与其他方法相比, 该算法可以显著提高作物类别的识别精度, 更好的区分不同杂草类型, 同时大幅度减少分类结果中的椒盐噪声。同时, 本文所提的 SAD/ED-PTSVM 分类算法与 SUAL-SVM 分类算法, 与其他分类算法相比, 可以在标记样本数量相同的情况下, 不同程度的提高分类精度, 尤其是对于分布较少的植被类别, 可以有效提高其识别精度。利用 SAD/ED-PTSVM 分类算法和 SUAL-SVM 分类算法, 可以在提高分类精度和分类效率的同时, 降低样本标注所需的人力物力, 这对于高光谱数据植被精细分类具有重要意义。

根据本章的实验结果, 可以提出针对地面尺度的高光谱遥感数据植被精细分类流程, 具体如下:

第一步: 结合分类目的确定植被分类体系, 并依此选择地面高光谱遥感数据;

第二步: 对获取到的地面高光谱遥感数据进行预处理, 具体包括: 光谱定标、辐射校正和几何校正;

第三步: 根据研究区实际地表覆盖情况, 对各类别选择一定的初始训练样本, 对于 SAD/ED-PTSVM 分类, 应尽量选择容易与其他类别误分的像元, 使分划函数更加准确; 而对于 SUAL-SVM 分类, 初始标记样本只需选择各类别中比较确定的、“纯度”较高的像元, 以进行混合像元分解;

第四步: 根据所选择的训练样本计算光谱指数特征, 对训练样本所在窗口像元进行主成分分析, 保留前 n 个使累积贡献率达到一定数值 (90%, 95% 或 98%) 的特征波段, 计算其纹理特征, 构建植被特征库;

第五步: 根据训练样本, 计算各波段对于不同类对的可分性, 采用基于类对

可分性（CPS）的特征库光谱维优化算法对植被特征库进行光谱维优化；

第六步：通过计算训练样本中各类别的平均光谱角距离设置 $MinSAD$ 参数，并调整窗口大小，进行基于 $NPSAD$ 的特征库空间维优化；

第七步：选择半监督分类算法，进行 $SAD/ED-PTSVM$ 分类或 $SUAL-SVM$ 分类；

第八步：得到最终分类结果，按照应用需求对分类结果进行分类后处理，并对分类结果进行精度评价。

第6章 基于航空高光谱遥感数据的作物精细分类

航空尺度的高光谱遥感数据是农业遥感监测的重要数据源，同时具有较高的光谱分辨率与空间分辨率，在农作物品种识别、种植面积估算、产量估算方面具有很大的作用。本章主要利用航空尺度的高光谱数据，验证所提基于特征库构建及优化的分类算法和基于支持向量机的半监督分类算法，在航空尺度高光谱植被精细分类中的优势，以及在作物精细识别与监测中的有效性。

本章具体可以分为以下几个部分：首先，介绍所用的航空成像光谱系统；其次，介绍航空高光谱数据的预处理及研究区概况；然后，分别将本文所提特征库光谱维与空间维优化算法与其他特征选择及滤波算法进行比较，再利用所提出的两种基于支持向量机的半监督分类算法分别得到分类结果，并与其他分类结果进行比较分析；最后提出一套适用于航空高光谱遥感数据植被精细分类的方法流程，为航空高光谱遥感数据在作物精细识别与监测中更广泛的应用提供参考。

6.1 航空高光谱遥感数据预处理

本文所用的航空高光谱数据为 PHI (Pushbroom Hyperspectral Imager) 图像。PHI 由上海技术物理所研制生产，为推扫式 CCD 面阵成像光谱仪，总视场角约为 0.36 弧度，瞬时视场角约为 1 毫弧度，光谱分辨率小于 5nm，光谱采样间隔约为 1.86nm，有效像元数目为 367 个，仪器具体参数见表 6.1.1。本文实验所用数据共有 80 个光谱波段，对应的具体波长见表 6.1.2。

表 6.1.1 PHI 成像光谱系统主要参数

指标	参数	指标	参数
工作方式	面阵 CCD 推扫式	信噪比	300
总视场角	0.36rad (21°)	瞬时视场角	1.0mrad
光谱采样	1.86nm	光谱范围	400nm-850nm
光谱分辨率	<5nm	像元数	367pixels/line
帧频	60Fr/Sec	数据速率	7.2Mb/Sec

本文实验中所用的数据为 2000 年 8 月在日本长野南牧村 (Minamimaki) 进

行航拍实验获取的数据，实验区地物以植被为主，包括不同生长期和覆盖条件下的白菜、卷心菜、生菜等多种作物。实验区属平原地形，设计相对航高约 1700 米，对应的地面分辨率约 1.7m。

表 6.1.2 实验所用 PHI 数据各波段对应波长

波段号	波长 (nm)	波段号	波长 (nm)	波段号	波长 (nm)	波段号	波长 (nm)
1	411.90	21	569.30	41	647.00	61	723.30
2	419.70	22	573.20	42	650.80	62	727.00
3	427.50	23	577.10	43	654.60	63	730.70
4	435.40	24	581.00	44	658.50	64	734.40
5	443.20	25	585.00	45	662.30	65	738.20
6	451.10	26	588.90	46	666.10	66	741.90
7	458.90	27	592.80	47	669.90	67	745.60
8	466.70	28	596.70	48	673.80	68	749.20
9	474.50	29	600.60	49	677.60	69	752.90
10	482.30	30	604.60	50	681.50	70	760.20
11	490.10	31	608.50	51	685.30	71	767.40
12	498.00	32	612.40	52	689.20	72	774.70
13	506.10	33	616.30	53	693.00	73	782.00
14	514.10	34	620.20	54	696.90	74	789.20
15	522.10	35	624.20	55	700.70	75	796.40
16	530.20	36	628.10	56	704.60	76	803.60
17	538.10	37	632.00	57	708.40	77	810.90
18	545.90	38	635.80	58	712.10	78	818.20
19	553.70	39	639.50	59	715.80	79	825.50
20	561.50	40	643.30	60	719.60	80	832.80

实验所用高光谱图像都进行了光谱定标和辐射校正，具体的参数及方法由仪器制造者上海技术物理所提供。在进行航空高光谱遥感实验的同时，进行了地面同步或准同步定标，采用经验线性法进行大气校正。另外，在进行航空实验的前

后短时间内，还对研究区内的地物分布进行了详细调查。研究区共包含 8 种植被类型，分别是日本白菜、中国白菜、卷心菜、生菜、萝卜、牧草、架豆和林地。研究区的 PHI 影像以及实际植被分布如下图。

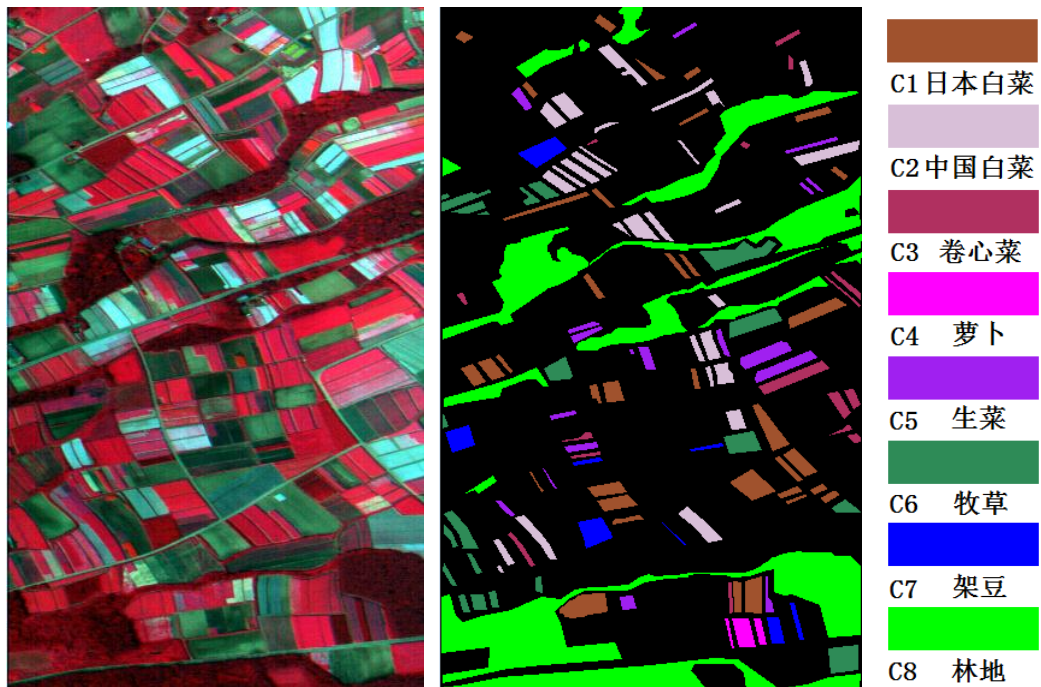


图 6.1.1 研究区的 PHI 影像（标准假彩色合成）及实际植被分布图

在分类精度评价方面，主要根据实际植被分布图对整幅影像进行计算，本文比较了总体精度（OAA）、Kappa 系数（KC），以及各类别的生产者精度（PA）和用户精度（UA）。

6.2 植被特征库构建与优化

在本节实验中，所用的训练样本为基于实际植被分布图（图 6.1.1）各类别随机生成的 25 个训练样本，而植被特征库光谱维优化中所用的训练样本，与本节后续分类实验中所用的训练样本相同。

6.2.1 植被特征库构建

在植被特征库构建中，光谱指数特征直接利用原始光谱波段间的计算获得，共计 50 个光谱指数特征；空间纹理特征的获取则是首先对原始光谱波段进行主成分分析，选取前 4 个主成分（累积贡献率超过 98%）分别计算 10 个纹理特征，得到 40 个纹理波段。构建的植被特征库具体组成如表 6.2.1。

6.2.1 PHI 影像植被特征库对应波段

波段号	对应波段
1-80	原始反射率波段
81-87	第 1 主成分对应 7 个纹理波段，依次为： 均值、均质性、对比度、异质性、熵、二阶矩和相关性
88-94	第 2 主成分对应 7 个纹理波段（次序同上）
95-101	第 3 主成分对应 7 个纹理波段（次序同上）
102-108	第 4 主成分对应 7 个纹理波段（次序同上）
109-112	第 1-4 主成分对应的 Moran 指数波段
113-116	第 1-4 主成分对应的 Geary 系数波段
117-120	第 1-4 主成分对应的 G 统计量波段
121-170	表 3.1.1 中所列的 50 个光谱指数波段

各植被类别在特征库中各波段(已经过归一化处理)的均值如图 6.2.1 所示。

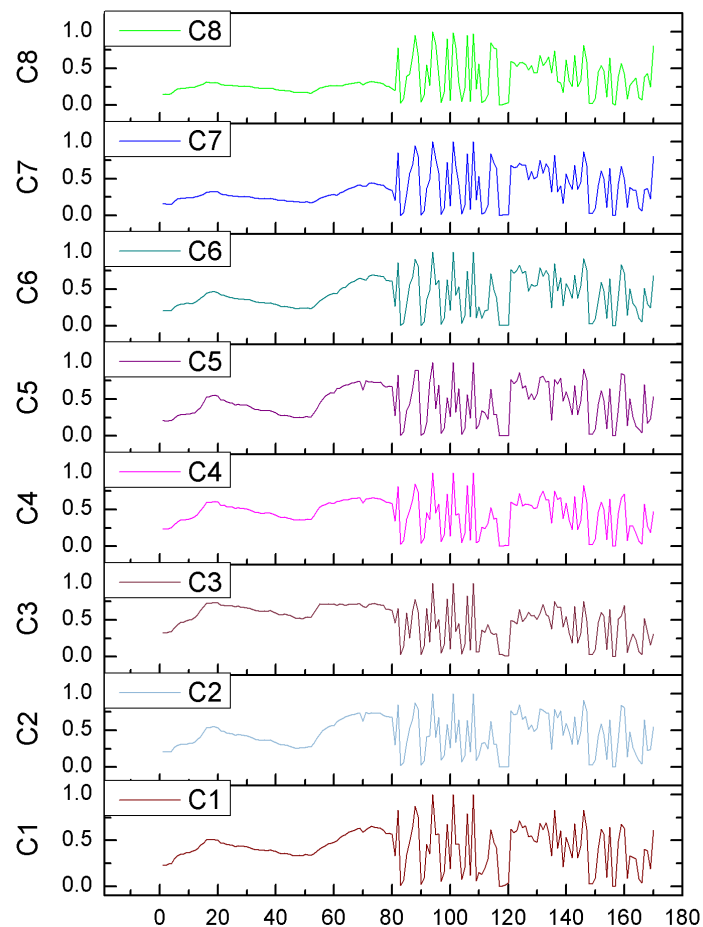


图 6.2.1 各植被类别在特征库中各波段上的均值分布

6.2.2植被特征库优化

在对所构建的植被特征库进行光谱维优化时，根据前面随机选择的各类别 25 个训练样本，利用第三章介绍的基于类对可分性的植被特征库光谱维优化算法，得到光谱维优化后的植被特征库。光谱维优化后最终保留了 19 个植被特征，其中原始波段 5 个，纹理特征 8 个，光谱指数特征 6 个。光谱维优化后保留的植被特征波段具体见表 6.2.2。

表 6.2.2 光谱维优化后保留的植被特征波段

具体波段信息	
原始特征	对应波长(nm): 561.5, 658.5, 700.7, 734.4, 796.4
纹理特征	第一主成分: 对比度、G 统计量
	第二主成分: Moran 指数、Geary 系数
	第三主成分: 均值、Moran 指数
	第四主成分: Geary 系数、G 统计量
光谱指数特征	表 3.1.1 中序号分别为: 3, 5, 18, 19, 21, 29

在进行植被特征库空间维优化时，首先计算各类别训练样本的平均光谱角，具体计算结果如图 6.2.2，取其中最小值确定参数，即 *MinSAD* 为 0.0380，同时设置窗口大小为 5×5，迭代次数为 1 次，之后按照第 3 章第 3 节中介绍的步骤进行基于 NPSAD 的植被特征库空间维优化操作。

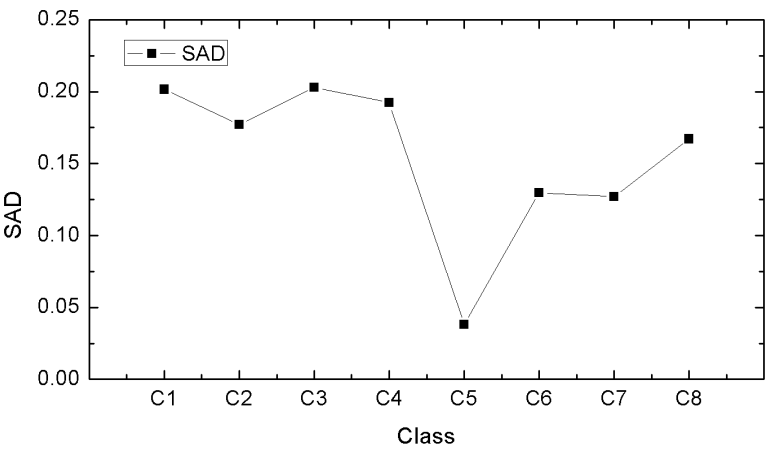


图 6.2.2 各类别训练样本的平均光谱角距离（SAD）

6.2.3 基于植被特征库构建与优化的航空高光谱数据分类

为了客观的验证基于植被特征库构建与优化的分类算法在航空高光谱数据植被精细分类中的有效性,本节比较了不同的特征选择方法与滤波算法,在分类器方面统一采用 ENVI 软件自带的传统支持向量机,选择 RBF 核函数,参数方面均采用 ENVI 软件默认参数,未进行参数优化。

具体的分类方法见表 6.2.3, 相应的分类结果与精度如图 6.2.3 与表 6.2.4。

表 6.2.3 实验中比较的分类方法

名称	光谱维优化算法	特征波段数
CPS	基于类对可分性的光谱维优化算法	19
OIF	最优索引因子法 (Chavez 等, 1982)	19
名称	空间维优化算法	窗口大小
NPSAD	基于邻域 SAD 的空间维优化算法	5×5
MID	中值滤波	5×5
LP	低通滤波	5×5
ELee	加强型 Lee 滤波	5×5
EFrost	加强型 Frost 滤波	5×5

从图 6.2.3 与表 6.2.4 中可以看出,通过特征库构建及光谱维优化,仅用 19 个特征波段便可得到比原始 80 个波段更高的分类精度。同时,基于 CPS 的特征库光谱维优化后分类结果要明显优于最优索引因子法特征选择的分类结果。从细节图可看出,通过构建特征库,增强了相似作物间的识别能力,但是单纯利用特征库构建及光谱维优化的分类结果中椒盐噪声较多,作物地块较为破碎,不利于完整的提取作物地块信息。

光谱维优化后出现的椒盐噪声多,地块破碎度高等问题,可以通过在植被特征库的空间维优化进行解决。从分类结果中可以看出,通过空间维优化,分类结果的椒盐噪声明显减少,作物地块更加完整,总体精度与 Kappa 系数都有显著提高(分别提高 5.7% 和 0.07),同时,大部分作物类别的生产者精度与用户精度也都有不同幅度的提高。

从不同空间维优化算法的分类结果比较中可看出,基于 NPSAD 的空间维优

化算法可以有效避免加强型 Lee 滤波与加强型 Frost 滤波算法造成的边缘效应，同时也可以更好的对地块内部进行平滑，更完整的提取出作物地块，进而提高作物种植面积以及产量估算的准确度。

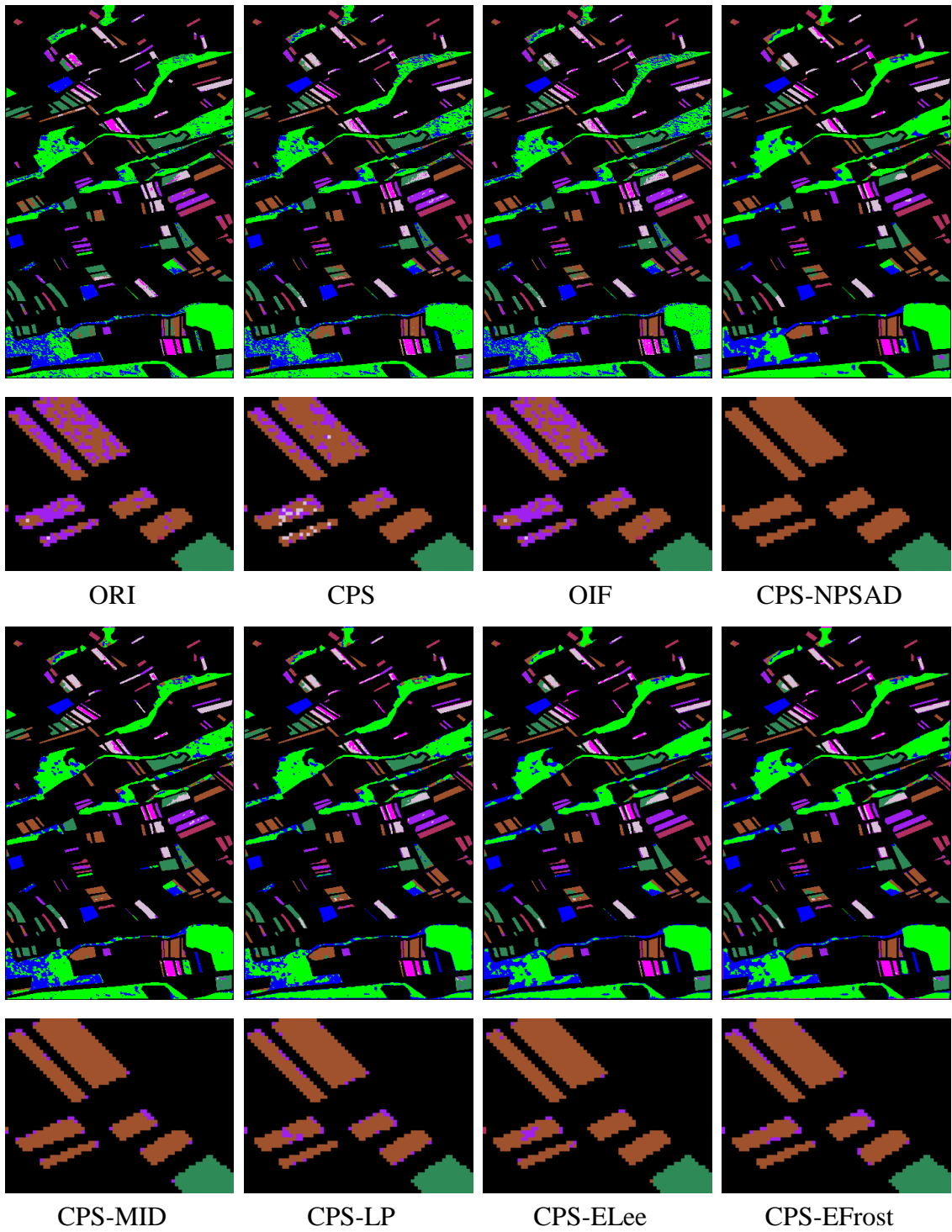


图 6.2.3 各方法对应分类结果图

注：ORI 为直接利用原始数据 80 个波段进行 SVM 分类结果

表 6.2.4 各方法对应分类结果精度表

注：ORI 为直接利用原始数据 80 个波段进行 SVM 分类结果。

	ORI		CPS		OIF		CPS-NPSAD	
	PA	UA	PA	UA	PA	UA	PA	UA
C1	51.46%	84.16%	66.93%	74.08%	66.36%	74.00%	77.45%	76.93%
C2	53.88%	82.28%	59.72%	83.06%	55.86%	80.47%	62.56%	92.73%
C3	78.55%	76.72%	93.63%	72.73%	92.16%	74.79%	95.28%	77.71%
C4	82.67%	64.53%	85.78%	69.07%	87.01%	62.82%	89.27%	83.97%
C5	87.90%	40.19%	71.64%	36.76%	67.11%	33.91%	87.33%	36.55%
C6	88.97%	53.43%	82.73%	68.15%	80.03%	63.33%	86.92%	76.18%
C7	74.60%	18.48%	81.23%	17.52%	81.87%	15.78%	83.74%	21.41%
C8	75.49%	96.31%	71.00%	97.34%	66.04%	97.18%	76.81%	97.67%
OAA(KC)	72.07%(0.61)		72.28%(0.62)		68.82%(0.58)		77.97%(0.69)	
	CPS-MID		CPS-LP		CPS-ELee		CPS-EFrost	
	PA	UA	PA	UA	PA	UA	PA	UA
C1	75.72%	73.71%	75.05%	72.72%	74.14%	75.87%	72.77%	79.54%
C2	61.14%	87.52%	50.13%	86.03%	52.65%	87.53%	51.45%	88.33%
C3	92.95%	72.51%	90.13%	74.86%	93.57%	75.48%	87.87%	74.84%
C4	85.78%	67.43%	90.63%	71.94%	91.95%	79.08%	92.80%	71.16%
C5	78.64%	39.06%	86.77%	31.96%	86.39%	34.70%	87.52%	24.39%
C6	82.84%	76.65%	81.64%	72.23%	82.27%	67.17%	82.91%	67.77%
C7	80.75%	20.10%	86.74%	21.13%	90.27%	20.09%	89.36%	21.64%
C8	74.72%	97.57%	74.72%	97.86%	72.60%	97.96%	73.86%	97.92%
OAA(KC)	75.63%(0.66)		74.81%(0.65)		74.13%(0.65)		74.42%(0.65)	

6.2.4 本节小结

本节利用基于植被特征库构建与优化的分类方法进行了航空尺度的高光谱植被精细分类实验，得到如下结论：

首先，通过构建植被特征库，可以增加用于植被精细分类的特征波段，增强不同作物之间的可分性，进而提高不同作物类型的识别能力；

其次，与最优索引因子法相比，基于 CPS 的光谱维优化算法可以有针对性

的提取出对各作物类别间区分最有帮助的特征波段,针对不同研究区的不同作物类型,提取的特征波段是不同的;

再次,与其他空间维优化算法相比,基于 NPSAD 的空间维滤波算法可以有效的利用光谱维信息,更为完整的提取作物地块信息,同时避免边缘效应;

最后,从本节实验可以看出,基于植被特征库构建与优化的分类方法,适用于航空尺度的高光谱遥感植被精细分类,可以更为准确的识别不同作物类型/品种,得到较为完整的作物地块信息,有效提高分类的精度和效率。

6.3 基于支持向量机的半监督分类

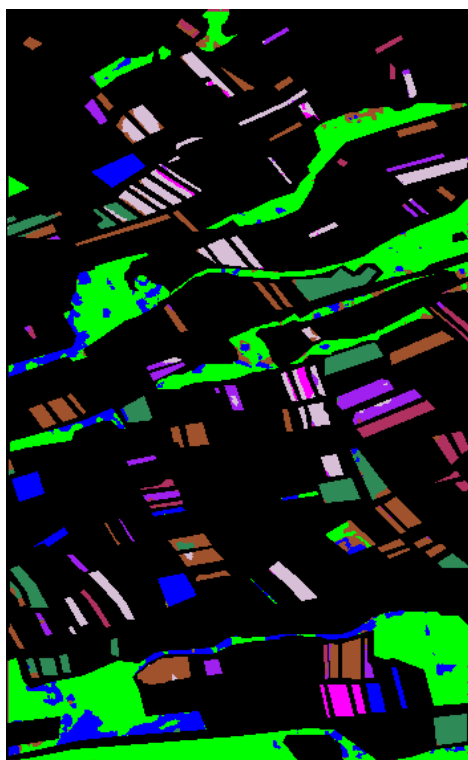
针对本文所提的基于 SAD-ED 双重判定的渐进直推支持向量机分类算法、结合混合像元分解进行主动学习的支持向量机分类算法,本节主要利用航空高光谱遥感数据进行植被半监督分类实验,将这两种半监督分类算法分别与单纯基于决策值进行类别标记的直推支持向量机分类、通过随机选择确定训练样本的支持向量机分类结果进行对比,验证本文所提算法的优越性。

本节分类实验中所用数据为上一节经过植被特征库构建与优化后,最终得到的 35 个特征波段。本节实验主要基于编写的 Matlab 程序实现,采用 RBF 核函数,惩罚系数与核参数通过网格搜索进行优化,搜索的最优 C 区间为 0.1 到 100,最优 γ 区间为 0.01 到 1000。

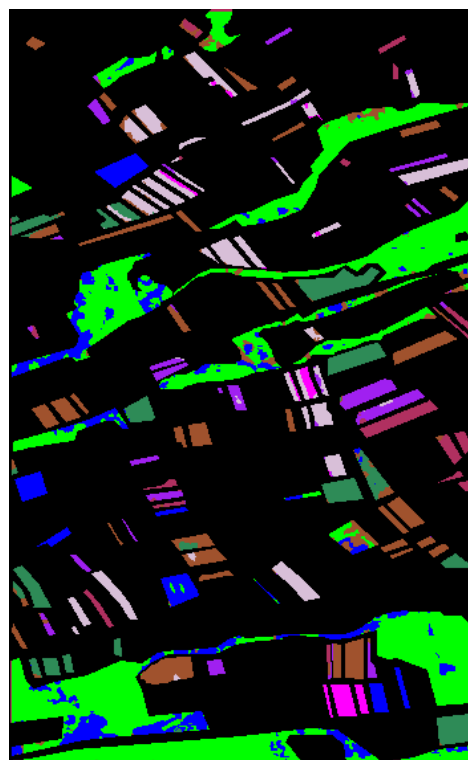
6.3.1 基于 SAD-ED 双重判定的渐进直推支持向量机分类

在对航空高光谱遥感数据进行基于 SAD-ED 双重判定的渐进直推支持向量机分类 (SAD/ED-PTSVM) 实验时,初始标记样本为基于实际植被分布图 (图 6.1.1) 各类别随机生成的 25 个训练样本,未标记样本则在各类对分类时,从剩余像元中随机生成的 20 个与初始标记样本不重叠的像元。为验证该算法优势,实验中将分类结果与单纯基于决策值实现类别标记的直推支持向量机 (DV-PTSVM) 和传统支持向量机 (SVM) 的分类结果进行了对比分析。DV-PTSVM 的样本获取方法与 SAD/ED-PTSVM 相同,而 SVM 则是直接利用前面随机生成的 25 个训练样本进行分类。

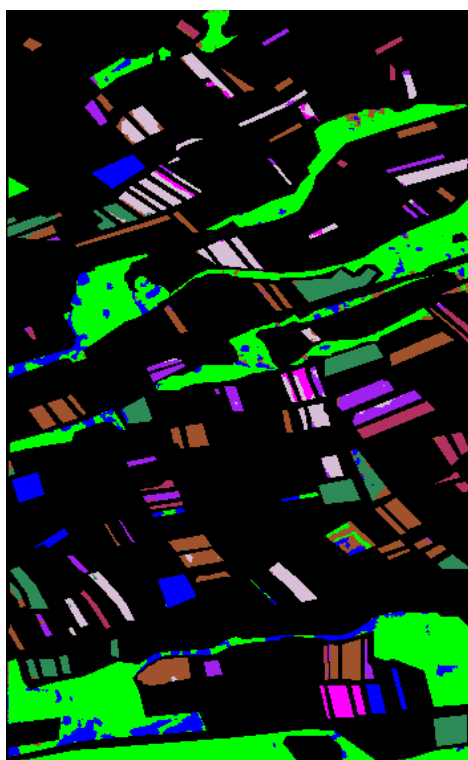
具体分类结果与精度如图 6.3.1 与表 6.3.1 所示。



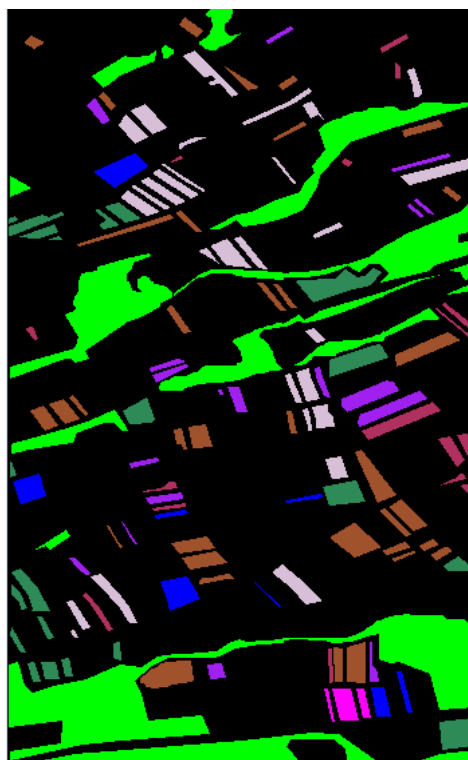
SVM



DV-PTSVM



SAD/ED-PTSVM



地表真实影像

图 6.3.1 各方法对应分类结果图

表 6.3.1 各方法对应分类结果精度表

	SAD/ED-PTSVM		DV-PTSVM		SVM	
	PA	UA	PA	UA	PA	UA
C1	87.89%	83.36%	86.76%	81.12%	86.04%	80.27%
C2	85.84%	96.26%	87.82%	95.13%	88.11%	93.38%
C3	91.42%	89.77%	92.95%	71.66%	93.14%	74.58%
C4	94.55%	88.98%	92.97%	89.03%	89.82%	90.09%
C5	100.00%	57.50%	99.62%	63.57%	99.62%	62.22%
C6	90.30%	92.02%	89.31%	90.78%	88.90%	90.00%
C7	95.35%	35.19%	95.45%	29.23%	96.58%	30.13%
C8	86.28%	99.07%	80.17%	98.85%	80.80%	98.93%
OAA(KC)	87.87%(0.82)		84.42%(0.78)		84.57%(0.78)	

从图 6.3.1 与表 6.3.1 可以看出,与传统 SVM 分类结果相比,SAD/ED-PTSVM 分类算法对于各作物类别间仅添加 20 个未标记样本,便可有效提高总体精度和 Kappa 系数;而 DV-PTSVM 分类结果与传统 SVM 相比,各作物类别精度变化不大,总体精度和 Kappa 系数稍有下降。可见, SAD/ED-PTSVM 分类算法可以有效提高未标记样本的标注准确度,进而提高作物识别分类的精度。

另外,结合本章第二节分类结果可看出,通过 Matlab 平台编程调用 LibSVM 工具箱,同时采用网络搜索参数优化策略,分类结果明显优于直接利用 ENVI 软件并采用默认参数的分类结果。

6.3.2结合混合像元分解进行主动学习的支持向量机分类

在结合混合像元分解进行主动学习的支持向量机（SUAL-SVM）分类实验中,初始标记样本为基于实际植被分布图（图 6.1.1）各类别随机生成的 5 个训练样本,后再利用 AL 通过人机交互补充 20 个训练样本,然后采用传统 SVM 进行分类;用来进行对比的随机选择样本分类算法中,则是直接利用实际植被分布图（图 6.1.1）随机生成 25×8=200 个训练样本,然后采用传统 SVM 进行分类。

具体的分类结果与精度如图 6.3.2 与表 6.3.2 所示。

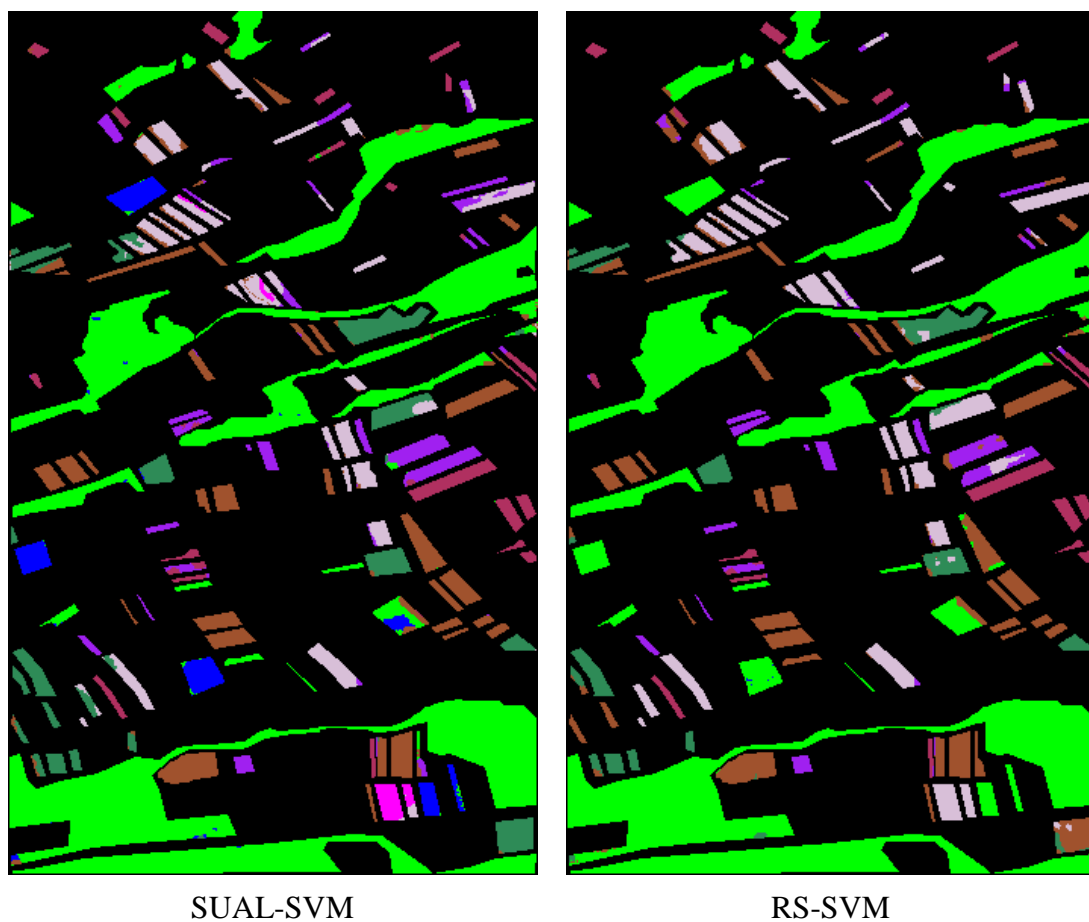


图 6.3.2 各方法对应分类结果图

表 6.3.2 各方法对应分类结果精度表

	SUAL-SVM		RS-SVM	
	PA	UA	PA	UA
C1	84.91%	86.89%	86.79%	75.67%
C2	79.15%	92.40%	90.03%	73.85%
C3	95.96%	75.11%	92.22%	80.78%
C4	89.61%	79.61%	76.66%	88.15%
C5	78.26%	83.13%	0.00%	0.00%
C6	90.21%	95.77%	61.35%	98.09%
C7	87.49%	88.96%	0.48%	100.00%
C8	99.03%	97.58%	99.26%	92.01%
OAA(KC)	93.27%(0.90)		87.20%(0.80)	

从图 6.3.2 与表 6.3.2 可以看出, 与 RS-SVM 分类结果相比, SUAL-SVM 分类结果, 在总体精度与 Kappa 系数方面都有大幅度的提高, 同时, 各作物类别的生产者精度与用户精度也更加均匀, 而利用 RS-SVM 分类, 容易导致某些作物类别由于训练样本太少而无法识别, 如 C5 (生菜) 与 C7 (架豆), 影响作物的识别和分类效果。

从分类结果可以看出, SUAL-SVM 算法选择出的训练样本更具有针对性, 可用于指导地面调查实验, 同时有效的提高各作物间的识别精度, 在样本数目相同的情况下, 可以得到更高的分类精度; 而满足某一精度, 则需要较少的训练样本, 可以有效减少样本标注的工作量, 这对于大尺度的作物识别与监测具有重要意义。

6.4 本章小结

本章利用航空高光谱遥感影像实验了本文所提的特征库构建与优化算法、基于 SAD-ED 双重判定的渐进直推支持向量机 (SAD/ED-PTSVM) 分类算法与结合混合像元分解进行主动学习的支持向量机 (SUAL-SVM) 分类算法, 并验证了这些算法在航空尺度作物监测中的有效性。

本章的实验结果表明, 基于植被特征库构建与优化的分类算法适用于航空高光谱遥感影像的植被精细分类, 与其他方法相比, 该算法可以显著提高不同作物类别/品种的分类精度, 提高所提取作物地块的完整度; 同时本文所提的 SAD/ED-PTSVM 分类算法与 SUAL-SVM 分类算法, 与其他分类算法相比, 可以显著提高作物分类结果的精度和分类效率, 降低样本标注所需的人力物力, 使得这两种算法在航空尺度高光谱数据植被精细分类中具有广阔的应用前景。

根据本章实验结果, 提出针对航空尺度高光谱遥感数据植被精细分类的方法流程如下:

第一步: 结合具体分类目的, 确定植被分类体系, 并依此选择合适的航空高光谱遥感数据;

第二步: 对航空高光谱遥感数据进行预处理, 具体包括: 辐射校正、大气校正和几何校正;

第三步: 根据研究区实际地表覆盖情况, 对各类别选择一定的初始训练样本,

对于 SAD/ED-PTSVM 分类，应尽量选择易与其他类别发生误分的像元，使分划函数更加准确；而 SUAL-SVM 分类，初始标记样本只需选择各类别中比较确定的、“纯度”较高的像元，以进行混合像元分解；

第四步：根据所选择的训练样本计算光谱指数特征，对训练样本所在窗口像元进行主成分分析，保留前 n 个使累积贡献率达到一定数值（90%，95%或 98%）的特征波段，计算其纹理特征，构建植被特征库；

第五步：根据训练样本，计算各波段对于不同作物类别的可分性，采用基于类对可分性的特征库光谱维优化算法对植被特征库进行光谱维优化；

第六步：通过计算训练样本中各类别的平均光谱角距离设置 $MinSAD$ 参数，并设置窗口大小，进行基于 NPSAD 的特征库空间维优化；

第七步：选择半监督分类算法，进行 SAD/ED-PTSVM 分类或 SUAL-SVM 分类；

第八步：得到最终分类结果，根据应用需求对分类结果进行分类后处理，并对分类结果进行精度评价。

第7章 结论和展望

高光谱遥感数据一直是植被精细分类研究中最重要数据源,其较高的光谱分辨率和较宽的波段覆盖范围使得很多在多光谱数据中无法识别的特征变得可以识别、不可区分的植被类别变得可以区分。但是,高光谱遥感数据在植被精细分类中主要存在以下问题:第一,随着类别精细度的不断提高,单纯利用光谱信息得到的分类结果已经不再能满足应用需求;第二,其成百上千的光谱波段带来了巨大的运算量,对监督分类中训练样本的需求也大幅增加;第三,随着传感器空间分辨率的提高,分类结果中椒盐噪声严重影响分类结果的应用。

近几年发展起来的半监督分类算法,对高光谱遥感数据的小样本分类有着重要意义。半监督分类算法同时利用了少量已标记样本和大量未标记样本,综合对整个数据进行分类。而目前的高光谱数据半监督分类算法大多只是单纯利用地物光谱的统计特征,从机器学习的角度实现半监督学习过程。为此,本文重点开展植被特征库的构建与优化算法和针对小样本的支持向量机半监督分类改进算法研究,提出改进的直推支持向量机与主动学习支持向量机分类算法研究。本文的研究成果对于高光谱遥感数据在作物品种识别、入侵物种监测、精准农业等领域的应用具有重要意义。

7.1 本文主要成果与结论

本文在充分调研国内外研究现状的基础上,系统分析比较了目前主流的监督与非监督高光谱植被精细分类算法,并在此基础上开展基于植被特征库构建与优化的分类算法研究;同时针对高光谱遥感分类中的标记样本难以获取这一问题,重点研究基于直推学习与主动学习的支持向量机半监督分类算法;最后分别采用具有较高空间分辨率的地面成像光谱数据与航空高光谱数据,进行植被精细分类实验,验证所提算法在地面尺度与航空尺度植被精细分类中的有效性。论文的主要研究成果和结论如下:

1. 系统分析比较了目前主流的高光谱植被监督分类算法,从已标记训练样本数量、先验知识掌握情况、数据分类的特征空间维度等几方面,对不同分类器用于高光谱遥感植被精细分类的适应性进行了分析:光谱角度

制图在高光谱植被精细分类时最好与基于统计特征的分类器结合；最大似然分类器对于样本数目与波段数比要求较高；决策树算法更适合样本充足的情况下针对原始光谱空间直接分类；支持向量机分类器性能稳定，在不同训练样本数量和特征空间维度的情况下，均能得到较好的分类结果，更适用于高光谱数据的植被精细分类；

2. 提出了基于植被特征库构建与优化的植被分类策略：基于植被的光谱、空间纹理和生化参量三方面属性，综合了原始光谱、空间纹理和对植被生化参量敏感的光谱指数三类特征构建植被特征库。空间纹理特征主要包括基于灰度共生矩阵和局域指示空间分析两类纹理特征；光谱指数特征主要考虑到不同植被类型间叶绿素、胡萝卜素、花青素和氮素含量的不同，重点选择了 50 个对这些生化组分含量变化敏感的植被指数。实验结果表明，通过构建植被特征库，有效的提高了不同植被类别间的可分性；
3. 提出了基于类对可分性（CPS）的植被特征库光谱维优化算法：该算法重点考虑不同植被类别在不同特征波段中的可分性，即“类对可分性”，在特征库光谱维优化中，分别保留对各类对具有最大 Bhattacharyya 距离的原始波段、纹理特征和光谱指数特征，并计算各类对的 Jeffries-Matusita 距离，通过迭代使各类别间均达到较高的区分度，并利用最优索引因子法进一步降低波段相关性。该算法可有效降低数据冗余，提高分类效率；
4. 提出了基于邻域光谱角距离（NPSAD）的植被特征库空间维优化算法：考虑到地物分布通常具有一定的连续性，即相邻像元属于同一地物类别概率较高。在有训练样本的情况下，该算法可以根据训练样本自动计算光谱角阈值；否则，需由用户根据先验知识设置阈值。对于窗口大小的设置，通常图像空间分辨率越高，窗口越大。与其他空间维优化算法相比，该算法可以在去除分类结果中椒盐噪声的同时，避免边缘效应，保留细节信息，提高分类精度和分类图像平滑度；
5. 给出了基于植被特征库构建与优化的分类流程：首先选取训练样本，基于各类别样本点及所在窗口计算植被特征库中所有特征，然后利用训练

样本进行基于 CPS 的光谱维优化,之后对于整幅影像构建光谱维优化后的植被特征库,再基于 NPSAD 进行空间维优化,最后选用合适的分类器对整幅影像进行分类,计算分类精度;

6. 提出了基于光谱角距离-欧式距离双重判定的渐进直推支持向量机 (SAD/ED-PTSVM) 分类算法: 该算法在传统渐进直推支持向量机基础上,保留了“成对标记”与“标签重置”特点,通过计算光谱角距离与欧式距离分别对未标记样本标签进行判定,并结合其到分划边界距离,实现对未标记样本的“自动”标注。该算法有效利用高光谱数据的光谱维信息,降低未标记样本的误标记概率,进而降低标签重置带来的时间成本,并有效简化了传统直推支持向量机的参数设置,减少参数优化所需时间,提高分类效率;
7. 提出了结合混合像元分解进行主动学习的支持向量机 (SUAL-SVM) 分类算法: 该算法将高光谱研究中的混合像元分解技术与机器学习领域中的主动学习策略进行有机结合,充分利用了各像元的丰度信息来辅助分类决策;同时,算法中设置了权重系数,用来调节“丰度最均匀”与“最易误分”像元比重。在主动学习样本添加时,该算法按各类对区分难易度由多到少添加样本,使样本添加更具有针对性。该算法可有效提高分布较少的类别的识别精度,利用更少的标记样本实现更高的总体分类精度,有效减轻样本标记工作量,减少分类器样本训练所需时间;
8. 地面与航空尺度的植被精细分类实验结果表明: 基于植被特征库构建与优化的分类算法可以有效的提高地面/航空尺度的不同种类作物/杂草的分类精度,提取出更完整的叶片/作物地块;而基于光谱角距离-欧氏距离双重判定的渐进直推支持向量机与结合混合像元分解进行主动学习的支持向量机分类算法,在各类别仅有 25 个标记样本的情况下,便可得到较高的分类精度,有效的提高了分类精度和分类效率,降低样本标注所需的人力物力。实验结果表明本文所提算法在地面与航空尺度的植被精细分类研究中具有广阔的应用前景。

7.2 本文创新点

本文的创新之处主要有以下三点：

1. 针对高光谱植被精细分类的特征库构建与优化策略：首先，在植被特征库构建中，在传统的光谱与空间纹理特征基础上，重点添加了对植被生化参量敏感的光谱指数特征；其次，从植被精细分类的目的出发，提出基于类对可分性的植被特征库光谱维优化算法，以保证各植被类别的分类精度；最后，在传统窗口滤波算法基础上，重点考虑邻域像元光谱角度距离，提出了基于邻域光谱角距离的植被特征库空间维优化算法流程。
2. 基于光谱角距离-欧式距离双重判定的渐进直推支持向量机分类算法：针对传统渐进直推支持向量机根据未标记样本决策函数进行标记的方法性能不稳定、在样本空间维度较高时容易导致未标记样本的误标注、以及参数设置复杂等问题，提出了针对高光谱遥感植被精细分类的改进算法，即基于光谱角距离-欧式距离双重判定的渐进直推支持向量机分类算法，该算法在未标记样本标注时，利用光谱角距离与欧式距离分别对未标记样本标签进行判定、同时结合未标记样本到分划边界的距离对其进行标注，增加未标记样本标注的准确度，加速分划面的优化进程，减少参数优化所需时间，提高分类的精度和效率。
3. 结合混合像元分解进行主动学习的支持向量机分类算法：考虑到高光谱植被精细分类中，最难区分的通常是混合像元分解结果中多种组分丰度比较平均、或主要由两种组分构成且二者丰度差别较小的像元，本文引入机器学习领域的主动学习算法，提出了结合混合像元分解进行主动学习的支持向量机分类算法，该算法首先根据初始标记样本对整幅图像进行解混，然后根据预先设置的权重、提取出最难区分的混合像元，利用支持向量机分划函数对提取的像元进行筛选，并按照类对区分的难易度按比例添加，使标记样本的添加更具有针对性，有效减轻样本标记工作量和分类器训练所需时间。

7.3 不足与展望

本文在高光谱植被精细分类算法特征库与优化和基于直推学习与主动学习的支持向量机分类算法方面进行了深入研究,取得了一些进展,但仍存在许多不足或需要进一步深入研究的问题,具体如下:

1. 初始标记样本选择问题:基于类对可分性的光谱维优化算法主要是在初始训练样本的基础上进行的,所以样本的代表性直接影响植被特征库的光谱维优化结果;基于 SAD-ED 双重判定的渐进直推支持向量机与结合混合像元分解进行主动学习的支持向量机分类算法,首次迭代的计算过程主要也是基于初始标记样本进行的。所以,初始标记样本的选择直接影响分类结果的精度,如何选择合适的初始标记样本是高光谱数据小样本分类中的重要问题,后续研究中将重点考虑如何结合非监督聚类结果、针对不同算法的需求来改善初始标记样本的分布;
2. 基于 SAD-ED 双重判定的渐进直推支持向量机分类算法中,本文实验所用的未标记样本是通过随机算法生成的,在进行支持向量机分类时,由于采用的“1V1”多类分类策略,可能导致生成的某些未标记样本对一些类对的区分完全没有帮助,后续将考虑结合其他数据分析技术来优选未标记样本,在对不同类对进行区分时,添加对应的未标记样本,进而改善各类对决策面的优化过程;
3. 结合混合像元分解进行主动学习的支持向量机分类算法,分类结果对于迭代次数、权重系数的变化非常敏感,同时支持向量机自身对于惩罚系数和核参数的设置也非常敏感,虽然本文已给出推荐的迭代次数与权重系数设置,但是对于不同的数据类型,并不能保证其适应性,如何自动、快速的找到最优的参数组合,是下一步重点研究内容之一;
4. 在今后的研究中,还将继续挖掘对植被生化参量敏感的光谱指数特征和更多的空间形状、结构特征等,不断更新植被特征库,提高相似植被类型的可分性。

参考文献

- [1] 白继伟, 赵永超, 张兵等. 基于包络线消除的高光谱图像分类方法研究[J]. 计算机工程与应用, 2003, 39(13): 88-90.
- [2] 边肇祺, 张学工. 模式识别[M]. 北京: 清华大学出版社, 2000.
- [3] 陈荣, 曹永锋, 孙洪. 基于主动学习和半监督学习的多类图像分类[J]. 自动化学报, 2011, 37(8): 954-962.
- [4] 陈述彭, 童庆禧, 郭华东. 遥感信息机理研究[M]. 北京: 科学出版社, 1998.
- [5] 承继成, 郭华东, 史文中等. 遥感数据的不确定性问题[M]. 北京: 科学出版社, 2004.
- [6] 程涛. 基于多元地统计学的多波段纹理及其图像分类应用[D]. 北京大学, 2006.
- [7] 邓乃扬, 田英杰. 支持向量机-理论、算法与拓展[M]. 北京: 科学出版社, 2009.
- [8] 董连英. Hyperion 影像森林植被分类方法与应用研究[D]. 吉林大学, 2013.
- [9] 杜培军, 谭琨, 夏俊士. 高光谱遥感影像分类与支持向量机应用研究[M]. 北京: 科学出版社, 2012.
- [10] 方俊永, 亓洪兴, 刘学, 等. 地面成像光谱辐射计的设和初步结果[A]. 第七届成像光谱技术与应用研讨会文集, 2009.
- [11] 高恒振, 万建伟, 徐湛, 等. 基于光谱加权直推式支持向量机的高光谱图像半监督分类[J]. 信号处理, 2011, 27(001): 122-127.
- [12] 高恒振, 万建伟, 许可, 等. 基于聚类核函数的最小二乘支持向量机高光谱图像半监督分类[J]. 信号处理, 2011, 27(002): 276-280.
- [13] 高连如. 高光谱遥感目标探测中的信息增强与特征提取研究[D]. 中国科学院遥感应用研究所, 2007.
- [14] 耿修瑞. 高光谱遥感图像目标探测与分类技术研究[D]. 中国科学院遥感

- 应用研究所, 2005.
- [15]耿修瑞等. 一种基于空间连续性的高光谱图像分类方法[J]. 红外与毫米波学报, 2004, 23(4): 299-302.
- [16]雷彤. 基于高光谱的苹果花期果期光谱特征分析及其果量估测[D]. 山东农业大学, 2010.
- [17]李颖. 基于成像光谱数据的田间植物识别研究[D]. 北京大学, 2012.
- [18]厉小润, 朱洁尔, 王晶, 等. 组合核支持向量机高光谱图像分类[J]. 浙江大学学报 (工学版), 2013, 8: 012.
- [19]刘波. 基于地面成像光谱系统的植被信息提取研究[D]. 中国科学院遥感应用研究所, 2010.
- [20]刘峰, 龚健雅. 一种基于多特征的高光谱遥感图像分类方法[J]. 地理与地理信息科学, 2009, 25(3): 19-22.
- [21]钱育蓉, 贾振红, 于炯, 等. BP-ANN 在荒漠草地高光谱分类研究中的应用[J]. 计算机工程与应用, 2011, 47(12): 225-228.
- [22]钱育蓉, 于炯, 贾振红, 等. 新疆典型荒漠草地的高光谱特征提取和分析研究[J]. 草业学报, 2013, 1: 020.
- [23]尚坤, 李培军, 程涛. 基于合成核支持向量机的高光谱土地覆盖分类[J]. 北京大学学报: 自然科学版, 2011, 47(001): 109-114.
- [24]苏红军, 杜培军, 盛业华. 高光谱遥感数据光谱特征提取算法与分类研究 [J]. 计算机应用研究, 2008, 25(2).
- [25]童庆禧, 薛永祺, 王晋年, 等. 地面成像光谱辐射测量系统及其应用[J]. 遥感学报, 2010 (003): 409-422.
- [26]童庆禧, 张兵, 郑兰芬. 高光谱遥感—原理、技术与应用[M]. 北京: 高等教育出版社, 2006.
- [27]王楠楠. 高光谱图像异常目标检测[D]. 哈尔滨工程大学, 2009.
- [28]韦玮, 李增元. 基于高光谱影像融合的湿地植被类型信息提取技术研究 [J]. 林业科学研究, 2011, 24(3): 300-306.
- [29]温兴平, 胡光道, 杨晓峰. 从高光谱遥感影像提取植被信息[J]. 测绘科学,

- 2008, 33(3): 66-68.
- [30]杨国鹏, 余旭初, 冯伍法, 等. 高光谱遥感技术的发展与应用现状[J]. 测绘通报, 2008 (10): 1-4.
- [31]姚伏天, 钱运涛, 李吉明. 空间约束半监督高斯过程下的高光谱图像分类[J]. 浙江大学学报 (工学版), 2012, 7: 024.
- [32]张江红. 多分类主动学习方法在地表分类中的应用[D]. 南京理工大学, 2011.
- [33]张钧萍, 张晔. 基于多特征多分辨率融合的高光谱图像分类[J]. 红外与毫米波学报, 2004, 23(5): 345-348.
- [34]张连蓬. 基于投影寻踪和非线性主曲线的高光谱遥感图像特征提取及分类研究[D]. 青岛: 山东科技大学, 2003.
- [35]张良培, 张立福. 高光谱遥感[M]. 北京: 测绘出版社, 2011.
- [36]张文杰. 星载高光谱传感器模拟仿真系统研究[D]. 哈尔滨工业大学, 2008.
- [37]张西雅, 金慧然, 李培军. 多角度高光谱 CHRIS/PROBA 数据在土地覆盖分类中的应用[J]. 地理与地理信息科学, 2010, 26(5): 68-71.
- [38]张霞, 张兵. 中巴地球资源一号卫星多光谱扫描图象质量评价[J]. 中国图象图形学报: A 辑, 2002, 7(6): 581-586.
- [39]张学工. 关于统计学习理论与支持向量机[J]. 自动化学报, 2000, 26(1): 32-42.
- [40]朱子先. 基于高光谱遥感影像的扎龙湿地植被分类研究[D]. 哈尔滨师范大学, 2012.
- [41]赵英时. 遥感应用分析原理与方法[M]. 北京: 科学出版社, 2003.
- [42]Acito N., Diani M., Corsini G. Signal-dependent noise modeling and model parameter estimation in hyperspectral images[J]. Geoscience and Remote Sensing, IEEE Transactions on, 2011, 49(8): 2957-2971.
- [43]Adankon M. M., Cheriet M. Learning semi-supervised svm with genetic algorithm[C]//Neural Networks, 2007. IJCNN 2007. International Joint

- Conference on. IEEE, 2007: 1825-1830.
- [44] Al-Mallahi A., Kataoka T., Okamoto H.. Discrimination between potato tubers and clods by detecting the significant wavebands[J]. Biosystems Engineering, 2008, 100(3): 329-337.
- [45] Anselin L. Local indicators of spatial association—LISA[J]. Geographical analysis, 1995, 27(2): 93-115.
- [46] Barnes J. D., Balaguer L., Manrique E., et al. A reappraisal of the use of DMSO for the extraction and determination of chlorophylls a and b in lichens and higher plants[J]. Environmental and Experimental Botany, 1992, 32(2): 85-100.
- [47] Beaven S. G., Hazel G. G., Stocker A. D. Automated Gaussian spectral clustering of hyperspectral data[C]//AeroSense 2002. International Society for Optics and Photonics, 2002: 254-266.
- [48] Belluco E., Camuffo M., Ferrari S., et al. Mapping salt-marsh vegetation by multispectral and hyperspectral remote sensing[J]. Remote sensing of environment, 2006, 105(1): 54-67.
- [49] Benediktsson J. A., Palmason J. A., Sveinsson J. R. Classification of hyperspectral data from urban areas based on extended morphological profiles[J]. Geoscience and Remote Sensing, IEEE Transactions on, 2005, 43(3): 480-491.
- [50] Bioucas-Dias J. M., Plaza A., Camps-Valls G., et al. Hyperspectral Remote Sensing Data Analysis and Future Challenges[J]. Geoscience and Remote Sensing Magazine, IEEE, 2013, 1(2): 6-36.
- [51] Bioucas-Dias J. M., Plaza A., Dobigeon N., et al. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches[J]. Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of, 2012, 5(2): 354-379.
- [52] Blackburn G. A. Spectral indices for estimating photosynthetic pigment

- concentrations: a test using senescent tree leaves[J]. *International Journal of Remote Sensing*, 1998, 19(4): 657-675.
- [53] Borak J. S. Feature selection and land cover classification of a MODIS-like data set for a semiarid environment[J]. *International Journal of Remote Sensing*, 1999, 20(5): 919-938.
- [54] Borregaard T, Nielsen H., Norgaard L., Have H. Crop-weed discrimination by line imaging spectroscopy[J]. *J. agric. Engng. Res.*, 2000, 75: 389-400.
- [55] Breiman L., Friedman J., Stone C. J., et al. *Classification and regression trees*[M]. CRC press, 1984.
- [56] Broge N. H., Leblanc E. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density[J]. *Remote sensing of environment*, 2001, 76(2): 156-172.
- [57] Bruce L. M., Koger C. H., Li J. Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction[J]. *Geoscience and Remote Sensing, IEEE Transactions on*, 2002, 40(10): 2331-2338.
- [58] Bruzzone L., Chi M., Marconcini M. A novel transductive SVM for semisupervised classification of remote-sensing images[J]. *Geoscience and Remote Sensing, IEEE Transactions on*, 2006, 44(11): 3363-3373.
- [59] Bruzzone L., Persello C. A novel context-sensitive semisupervised SVM classifier robust to mislabeled training samples[J]. *Geoscience and Remote Sensing, IEEE Transactions on*, 2009, 47(7): 2142-2154.
- [60] Camps-Valls G., Bandos Marsheva T., Zhou D. Semi-supervised graph-based hyperspectral image classification[J]. *Geoscience and Remote Sensing, IEEE Transactions on*, 2007, 45(10): 3044-3054.
- [61] Camps-Valls G., Gomez-Chova L., Muñoz-Marí J., et al. Composite kernels for hyperspectral image classification[J]. *Geoscience and Remote Sensing Letters, IEEE*, 2006, 3(1): 93-97.

- [62]Camps-Valls G., Mooij J., Scholkopf B. Remote sensing feature selection by kernel dependence measures[J]. *Geoscience and Remote Sensing Letters*, IEEE, 2010, 7(3): 587-591.
- [63]Chang C. C., Lin C. J. LIBSVM: a library for support vector machines[J]. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2011, 2(3): 27. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [64]Chapelle O., Zien A. Semi-supervised classification by low density separation[J]. 2004.
- [65]Chavez P. S., Berlin G. L., Sowers L. B. Statistical method for selecting Landsat MSS ratios[J]. *Journal of applied photographic engineering*, 1982, 8(1): 23-30.
- [66]Chen Y., Nasrabadi N. M., Tran T. D. Hyperspectral image classification using dictionary-based sparse representation[J]. *Geoscience and Remote Sensing, IEEE Transactions on*, 2011, 49(10): 3973-3985.
- [67]Chen Y., Wang G., Dong S. Learning with progressive transductive support vector machine[J]. *Pattern Recognition Letters*, 2003, 24(12): 1845-1855.
- [68]Cheng X., Chen Y. R., Tao Y., et al. A novel integrated PCA and FLD method on hyperspectral image feature extraction for cucumber chilling damage inspection[J]. *Transactions-American Society of Agricultural Engineers*, 2004, 47(4): 1313-1320.
- [69]Cover T. M. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition[J]. *Electronic Computers, IEEE Transactions on*, 1965 (3): 326-334.
- [70]Dalla Mura M., Villa A., Benediktsson J. A., et al. Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis[J]. *Geoscience and Remote Sensing Letters*, IEEE, 2011, 8(3): 542-546.
- [71]Dasgupta S. Coarse sample complexity bounds for active learning[C]//NIPS.

- 2005, 3(3.2): 15.7.
- [72] Daughtry C. S. T., Walthall C. L., Kim M. S., et al. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance[J]. Remote sensing of Environment, 2000, 74(2): 229-239.
- [73] Demir B., Persello C., Bruzzone L. Batch-mode active-learning methods for the interactive classification of remote sensing images[J]. Geoscience and Remote Sensing, IEEE Transactions on, 2011, 49(3): 1014-1031.
- [74] Di W., Crawford M. M. Active learning via multi-view and local proximity co-regularization for hyperspectral image classification[J]. Selected Topics in Signal Processing, IEEE Journal of, 2011, 5(3): 618-628.
- [75] Feyaerts F., Van Gool L. Multi-spectral vision system for weed detection[J]. Pattern recognition letters, 2001, 22: 667-674.
- [76] Foody G. M. Status of land cover classification accuracy assessment[J]. Remote sensing of environment, 2002, 80(1): 185-201.
- [77] Foody G. M., Mathur A. A relative evaluation of multiclass image classification by support vector machines[J]. Geoscience and Remote Sensing, IEEE Transactions on, 2004, 42(6): 1335-1343.
- [78] Gamon J. A., Penuelas J., Field C. B. A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency[J]. Remote sensing of environment, 1992, 41(1): 35-44.
- [79] Gao B. C. An operational method for estimating signal to noise ratios from data acquired with imaging spectrometers[J]. Remote Sensing of Environment, 1993, 43(1): 23-33.
- [80] Gao L. R., Zhang B., Zhang X., et al. A new operational method for estimating noise in hyperspectral images[J]. Geoscience and Remote Sensing Letters, IEEE, 2008, 5(1): 83-87.
- [81] Getis A., Ord J. K. The analysis of spatial association by use of distance statistics[J]. Geographical analysis, 1992, 24(3): 189-206.

- [82]Gitelson A. A., Keydan G. P., Merzlyak M. N. Three - band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves[J]. Geophysical Research Letters, 2006, 33(11).
- [83]Gitelson A. A., Vina A., Ciganda V., et al. Remote estimation of canopy chlorophyll content in crops[J]. Geophysical Research Letters, 2005, 32(8).
- [84]Gitelson A., Merzlyak M. N. Spectral Reflectance Changes Associated with Autumn Senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. Leaves. Spectral Features and Relation to Chlorophyll Estimation[J]. Journal of Plant Physiology, 1994, 143(3): 286-292.
- [85]Gitelson A. A., Merzlyak M. N., Chivkunova O. B. Optical Properties and Nondestructive Estimation of Anthocyanin Content in Plant Leaves[J]. Photochemistry and photobiology, 2001, 74(1): 38-45.
- [86]Goetz A. F. H., Gao B. C., Wessman C. A., et al. Estimation of biochemical constituents from fresh, green leaves by spectrum matching techniques[C]//Geoscience and Remote Sensing Symposium, 1990. IGARSS'90.'Remote Sensing Science for the Nineties', 10th Annual International. IEEE, 1990: 971-974.
- [87]Goldberg D. E. Genetic algorithms in search, optimization, and machine learning[M]. Reading Menlo Park: Addison-wesley, 1989.
- [88]Gómez-Chova L., Camps-Valls G., Muñoz-Mari J., et al. Semisupervised image classification with Laplacian support vector machines[J]. Geoscience and Remote Sensing Letters, IEEE, 2008, 5(3): 336-340.
- [89]Gonzalez R. C., Woods R. E., Eddins S. L. Digital image processing using MATLAB[M]. Pearson Education India, 2004.
- [90]Green A. A., Berman M., Switzer P., et al. A transformation for ordering multispectral data in terms of image quality with implications for noise removal[J]. Geoscience and Remote Sensing, IEEE Transactions on, 1988, 26(1): 65-74.

- [91]Gualtieri J. A., Crompton R. F. Support vector machines for hyperspectral remote sensing classification[C]//The 27th AIPR Workshop: Advances in Computer-Assisted Recognition. International Society for Optics and Photonics, 1999: 221-232.
- [92]Haboudane D., Miller J. R., Pattey E., et al. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture[J]. Remote sensing of environment, 2004, 90(3): 337-352.
- [93]Haboudane D., Miller J. R., Tremblay N., et al. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture[J]. Remote sensing of environment, 2002, 81(2): 416-426.
- [94]Hansen M., Dubayah R., DeFries R. Classification trees: an alternative to traditional land cover classifiers[J]. International journal of remote sensing, 1996, 17(5): 1075-1081.
- [95]Haralick R. M., Shanmugam K., Dinstein I. H. Textural features for image classification[J]. Systems, Man and Cybernetics, IEEE Transactions on, 1973, (6): 610-621.
- [96]Heinz D. C., Chang C. I. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery[J]. Geoscience and Remote Sensing, IEEE Transactions on, 2001, 39(3): 529-545.
- [97]Hernández-Clemente R., Navarro-Cerrillo R. M., Suárez L., et al. Assessing structural effects on PRI for stress detection in conifer forests[J]. Remote Sensing of Environment, 2011, 115(9): 2360-2375.
- [98]Hirano A., Madden M., Welch R. Hyperspectral image data for mapping wetland vegetation[J]. Wetlands, 2003, 23(2): 436-448.
- [99]Hsu C. W., Chang C. C., Lin C. J. A practical guide to support vector

- classification[J]. 2003.
- [100] Huang C., Davis L. S., Townshend J. R. G. An assessment of support vector machines for land cover classification[J]. *International Journal of Remote Sensing*, 2002, 23(4): 725-749.
 - [101] Hughes G. On the mean accuracy of statistical pattern recognizers[J]. *Information Theory, IEEE Transactions on*, 1968, 14(1): 55-63.
 - [102] Jacquemoud S., Ustin S. L., Verdebout J., et al. Estimating leaf biochemistry using the PROSPECT leaf optical properties model[J]. *Remote sensing of environment*, 1996, 56(3): 194-202.
 - [103] Jia K., Wu B., Tian Y., et al. Vegetation classification method with biochemical composition estimated from remote sensing data[J]. *International Journal of Remote Sensing*, 2011, 32(24): 9307-9325.
 - [104] Jia X., Richards J. A. Binary coding of imaging spectrometer data for fast spectral matching and classification[J]. *Remote sensing of environment*, 1993, 43(1): 47-53.
 - [105] Jia X., Richards J. A. Efficient maximum likelihood classification for imaging spectrometer data sets[J]. *Geoscience and Remote Sensing, IEEE Transactions on*, 1994, 32(2): 274-281.
 - [106] Jia X., Richards J. A. Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification[J]. *Geoscience and Remote Sensing, IEEE Transactions on*, 1999, 37(1): 538-542.
 - [107] Joachims T. Svm-light: Support vector machine[J]. SVM-Light Support Vector Machine <http://svmlight.joachims.org/>, University of Dortmund, 1999, 19(4).
 - [108] Joachims T. Transductive inference for text classification using support vector machines[C]//ICML. 1999, 99: 200-209.
 - [109] Keerthi S. S., Lin C. J. Asymptotic behaviors of support vector machines

- with Gaussian kernel[J]. *Neural computation*, 2003, 15(7): 1667-1689.
- [110] Kekre H. B., Thepade S. D., Sarode T. K., et al. Image Retrieval using Texture Features extracted from GLCM, LBG and KPE[J]. *International Journal of Computer Theory and Engineering*, 2010, 2(5): 695-700.
- [111] Kennedy J., Eberhart R. Particle swarm optimization[C]//*Proceedings of IEEE international conference on neural networks*. 1995, 4(2): 1942-1948.
- [112] Kettig R L, Landgrebe D A. Classification of multispectral image data by extraction and classification of homogeneous objects[J]. *Geoscience Electronics, IEEE Transactions on*, 1976, 14(1): 19-26.
- [113] Kim D. H., Finkel L. H.. Hyperspectral image processing using locally linear embedding[C]//*Neural Engineering, 2003. Conference Proceedings. First International IEEE EMBS Conference on*. IEEE, 2003: 316-319.
- [114] Kruse F. A., Lefkoff A. B., Boardman J. W., et al. The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data[J]. *Remote sensing of environment*, 1993, 44(2): 145-163.
- [115] Kumar S., Ghosh J., Crawford M. M. Best-bases feature extraction algorithms for classification of hyperspectral data[J]. *Geoscience and Remote Sensing, IEEE Transactions on*, 2001, 39(7): 1368-1379.
- [116] Le Maire G., Francois C., Dufrene E. Towards universal broad leaf chlorophyll indices using PROSPECT simulated database and hyperspectral reflectance measurements[J]. *Remote sensing of environment*, 2004, 89(1): 1-28.
- [117] Li J, Marpu P R, Plaza A, et al. Generalized Composite Kernel Framework for Hyperspectral Image Classification[J]. 2013.
- [118] Li J., Bioucas-Dias J. M., Plaza A. Hyperspectral image segmentation using a new Bayesian approach with active learning[J]. *Geoscience and Remote Sensing, IEEE Transactions on*, 2011, 49(10): 3947-3960.
- [119] Li J., Bioucas-Dias J. M., Plaza A. Semisupervised hyperspectral image

- segmentation using multinomial logistic regression with active learning[J]. Geoscience and Remote Sensing, IEEE Transactions on, 2010, 48(11): 4085-4098.
- [120] Li J., Bioucas-Dias J. M., Plaza A. Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields[J]. Geoscience and Remote Sensing, IEEE Transactions on, 2012, 50(3): 809-823.
- [121] Li S., Zhang B., Li A., et al. Hyperspectral Imagery Clustering With Neighborhood Constraints[J]. Geoscience and Remote Sensing Letters, IEEE, 2013, 10(3): 588-592.
- [122] Lin H. T., Lin C. J. A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods[J]. submitted to Neural Computation, 2003: 1-32.
- [123] Mahlein A. K., Rumpf T., Welke P., et al. Development of spectral indices for detecting and identifying plant diseases[J]. Remote Sensing of Environment, 2013, 128: 21-30.
- [124] Mallinis G., Koutsias N., Tsakiri-Strati M., et al. Object-based classification using Quickbird imagery for delineating forest vegetation polygons in a Mediterranean test site[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2008, 63(2): 237-250.
- [125] Marconcini M., Camps-Valls G., Bruzzone L. A composite semisupervised SVM for classification of hyperspectral images[J]. Geoscience and Remote Sensing Letters, IEEE, 2009, 6(2): 234-238.
- [126] Martin-Herrero J. Comments on “A New Operational Method for Estimating Noise in Hyperspectral Images”[J]. Geoscience and Remote Sensing Letters, IEEE, 2008, 5(4): 705-709.
- [127] Mather P., Tso B. Classification methods for remotely sensed data[M]. CRC press, 2003.

-
- [128] Melgani F., Bruzzone L. Classification of hyperspectral remote sensing images with support vector machines[J]. *Geoscience and Remote Sensing, IEEE Transactions on*, 2004, 42(8): 1778-1790.
- [129] Meroni M., Rossini M., Guanter L., et al. Remote sensing of solar-induced chlorophyll fluorescence: Review of methods and applications[J]. *Remote Sensing of Environment*, 2009, 113(10): 2037-2051.
- [130] Mitri G. H., Gitas I. Z. Mapping post-fire forest regeneration and vegetation recovery using a combination of very high spatial resolution and hyperspectral satellite imagery[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2013, 20: 60-66.
- [131] Mountrakis G., Im J., Ogole C. Support vector machines in remote sensing: A review[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2011, 66(3): 247-259.
- [132] Nemani R., Running S. Land cover characterization using multitemporal red, near-IR, and thermal-IR data from NOAA/AVHRR[J]. *Ecological Applications*, 1997, 7(1): 79-90.
- [133] Niazmardi S., Naeini A. A., Homayouni S., et al. Particle swarm optimization of kernel-based fuzzy c-means for hyperspectral data clustering[J]. *Journal of Applied Remote Sensing*, 2012, 6(1): 063601-063601.
- [134] Noble S. D. Combining spectral and spatial information for automated plant identification[M]. 2007, 68(03).
- [135] Okamoto H., Lee W. S. Green citrus detection using hyperspectral imaging[J]. *Computers and Electronics in Agriculture*, 2009, 66(2): 201-208.
- [136] Paoli A., Melgani F., Pasolli E. Clustering of hyperspectral images based on multiobjective particle swarm optimization[J]. *Geoscience and Remote Sensing, IEEE Transactions on*, 2009, 47(12): 4175-4188.
- [137] Pasolli E., Melgani F., Bazi Y. Support vector machine active learning

- p>through significance space construction[J].
- Geoscience and Remote Sensing Letters, IEEE*
- , 2011, 8(3): 431-435.
- [138] Patra S., Bruzzone L. A batch-mode active learning technique based on multiple uncertainty for SVM classifier[J]. *Geoscience and Remote Sensing Letters, IEEE*, 2012, 9(3): 497-501.
- [139] Peñuelas J., Filella I. Visible and near-infrared reflectance techniques for diagnosing plant physiological status[J]. *Trends in plant science*, 1998, 3(4): 151-156.
- [140] Peñuelas J., Gamon J. A., Griffin K. L., et al. Assessing community type, plant biomass, pigment composition, and photosynthetic efficiency of aquatic vegetation from spectral reflectance[J]. *Remote Sensing of Environment*, 1993, 46(2): 110-118.
- [141] Plaza A., Benediktsson J. A., Boardman J. W., et al. Recent advances in techniques for hyperspectral image processing[J]. *Remote Sensing of Environment*, 2009, 113: S110-S122.
- [142] Plaza A., Martinez P., Perez R., et al. A new approach to mixed pixel classification of hyperspectral imagery based on extended morphological profiles[J]. *Pattern Recognition*, 2004, 37(6): 1097-1116.
- [143] Qi J., Chehbouni A., Huete A. R., et al. A modified soil adjusted vegetation index[J]. *Remote sensing of environment*, 1994, 48(2): 119-126.
- [144] Quinlan J. R. C4. 5: programs for machine learning[M]. Morgan kaufmann, 1993.
- [145] Rajan S., Ghosh J., Crawford M. M. An active learning approach to hyperspectral data classification[J]. *Geoscience and Remote Sensing, IEEE Transactions on*, 2008, 46(4): 1231-1242.
- [146] Ratle F., Camps-Valls G., Weston J. Semisupervised neural networks for efficient hyperspectral image classification[J]. *Geoscience and Remote Sensing, IEEE Transactions on*, 2010, 48(5): 2271-2282.

- [147] Richards J. A. Remote sensing digital image analysis[M]. Berlin et al.: Springer-Verlag, 2013.
- [148] Roberts D. A., Gardner M., Church R., et al. Mapping chaparral in the Santa Monica Mountains using multiple endmember spectral mixture models[J]. Remote Sensing of Environment, 1998, 65(3): 267-279.
- [149] Rondeaux G., Steven M., Baret F. Optimization of soil-adjusted vegetation indices[J]. Remote Sensing of Environment, 1996, 55(2): 95-107.
- [150] Roujean J. L., Breon F. M. Estimating PAR absorbed by vegetation from bidirectional reflectance measurements[J]. Remote Sensing of Environment, 1995, 51(3): 375-384.
- [151] Rouse Jr. J. W., Haas R. H., Schell J. A., et al. Monitoring vegetation systems in the Great Plains with ERTS[J]. NASA special publication, 1974, 351: 309.
- [152] Roweis S. T., Saul L. K. Nonlinear dimensionality reduction by locally linear embedding[J]. Science, 2000, 290(5500): 2323-2326.
- [153] Savitzky A., Golay M. J. E. Smoothing and differentiation of data by simplified least squares procedures[J]. Analytical chemistry, 1964, 36(8): 1627-1639.
- [154] Schleicher T. D., Bausch W. C., Delgado J. A., et al. Evaluation and refinement of the nitrogen reflectance index (NRI) for site-specific fertilizer management[C]//2001 ASAE Annual International Meeting, St-Joseph, MI, USA. ASAE Paper. 2001 (01-11151).
- [155] Sindhwani V., Keerthi S. S. Large scale semi-supervised linear SVMs[C]//Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2006: 477-484.
- [156] Singla A., Patra S., Bruzzone L. A novel classification technique based on progressive transductive SVM learning[J]. Pattern Recognition Letters,

2014, 42: 101-106.

- [157] Tarabalka Y., Benediktsson J. A., Chanussot J. Spectral-spatial classification of hyperspectral imagery based on partitional clustering techniques[J]. *Geoscience and Remote Sensing, IEEE Transactions on*, 2009, 47(8): 2973-2987.
- [158] Tarabalka Y., Fauvel M., Chanussot J., et al. SVM-and MRF-based method for accurate classification of hyperspectral images[J]. *Geoscience and Remote Sensing Letters, IEEE*, 2010, 7(4): 736-740.
- [159] Thorp K. R., French A. N., Rango A.. Effect of image spatial and spectral characteristics on mapping semi-arid rangeland vegetation using multiple endmember spectral mixture analysis (MESMA)[J]. *Remote Sensing of Environment*, 2013, 132: 120-130.
- [160] Tu T. M., Chen C. H., Wu J. L., et al. A fast two-stage classification method for high-dimensional remote sensing data[J]. *Geoscience and Remote Sensing, IEEE Transactions on*, 1998, 36(1): 182-191.
- [161] Tuia D., Camps-Valls G. Semisupervised remote sensing image classification with cluster kernels[J]. *Geoscience and Remote Sensing Letters, IEEE*, 2009, 6(2): 224-228.
- [162] Tuia D., Camps-Valls G. Urban image classification with semisupervised multiscale cluster kernels[J]. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 2011, 4(1): 65-74.
- [163] Tuia D., Pasolli E., Emery W. J. Using active learning to adapt remote sensing image classifiers[J]. *Remote Sensing of Environment*, 2011, 115(9): 2232-2242.
- [164] Tuia D., Ratle F., Pacifici F., et al. Active learning methods for remote sensing image classification[J]. *Geoscience and Remote Sensing, IEEE Transactions on*, 2009, 47(7): 2218-2232.
- [165] Tuia D., Volpi M., Copa L., et al. A survey of active learning algorithms

- for supervised remote sensing image classification[J]. *Selected Topics in Signal Processing*, IEEE Journal of, 2011, 5(3): 606-617.
- [166] Vapnik V. Statistical learning theory. 1998[J]. 1998.
- [167] Velasco-Forero S., Manian V. Improving hyperspectral image classification using spatial preprocessing[J]. *Geoscience and Remote Sensing Letters*, IEEE, 2009, 6(2): 297-301.
- [168] Virndts E., De Baerdemaeker J., Ramon H. Weed detection using canopy reflection[J]. *Precision Agriculture*, 2002, 3:63.
- [169] Yang H., Du Q., Ma B. Decision fusion on supervised and unsupervised classifiers for hyperspectral imagery[J]. *Geoscience and Remote Sensing Letters*, IEEE, 2010, 7(4): 875-879.
- [170] Zarco-Tejada P. J., Berjón A., López-Lozano R., et al. Assessing vineyard condition with hyperspectral indices: Leaf and canopy reflectance simulation in a row-structured discontinuous canopy[J]. *Remote Sensing of Environment*, 2005, 99(3): 271-287.
- [171] Zarco-Tejada P. J., Berni J. A. J., Suárez L., et al. Imaging chlorophyll fluorescence with an airborne narrow-band multispectral camera for vegetation stress detection[J]. *Remote Sensing of Environment*, 2009, 113(6): 1262-1275.
- [172] Zarco-Tejada P. J., González-Dugo V., Berni J. A. J. Fluorescence, temperature and narrow-band indices acquired from a UAV platform for water stress detection using a micro-hyperspectral imager and a thermal camera[J]. *Remote Sensing of Environment*, 2012, 117: 322-337.
- [173] Zarco-Tejada P. J., Miller J. R. Land cover mapping at BOREAS using red edge spectral parameters from CASI imagery[J]. *Journal of Geophysical Research: Atmospheres* (1984 – 2012), 1999, 104(D22): 27921-27933.
- [174] Zarco-Tejada P. J., Miller J. R., Morales A., et al. Hyperspectral indices and model simulation for chlorophyll estimation in open-canopy tree crops[J].

Remote sensing of environment, 2004, 90(4): 463-476.

- [175] Zomer R. J., Trabucco A., Ustin S. L. Building spectral libraries for wetlands land cover classification and hyperspectral remote sensing[J]. Journal of Environmental Management, 2009, 90(7): 2170-2177.

个人情况简介

博士期间发表文章情况

已发表或接收文章

- [1] Zhang X, **Shang K**. Estimating ecological indicators of karst rocky desertification by using linear spectral unmixing method[J]. International Journal of Applied Earth Observation and Geoinformation, 2014, 31: 86-94. (SCI)
- [2] **Shang K**, Zhang X, Dong Z, et al. Information extraction method on coastal wetland using TM data: A case study in Dongying, Shandong, China[C]//Earth Observation and Remote Sensing Applications (EORSA), 2012 Second International Workshop on. IEEE, 2012: 77-81. (EI)
- [3] **Shang K**, Zhang X, Zhang L, et al. Evaluation of hyperspectral classification methods based on FISS data[C]//Seventh International Symposium on Multispectral Image Processing and Pattern Recognition (MIPPR2011). International Society for Optics and Photonics, 2011: 80020L-80020L-8. (EI)
- [4] **Shang K**, Zhang X. A novel spectral spatial filtering approach for hyperspectral image classification[C]//Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS). (Accepted)(EI)
- [5] 尚坤, 李培军, 程涛. 基于合成核支持向量机的高光谱土地覆盖分类[J]. 北京大学学报: 自然科学版, 2011, 47(001): 109-114. (中文核心)
- [6] Xie Y S, Wang J N, **Shang K**. An improved approach based on Moment Matching to Destriping for Hyperion data[J]. Procedia Environmental Sciences, 2011, 10: 319-324. (EI)
- [7] Zhang X, Zhuang Z, Shuai T, **Shang K**, Sun Y. Improved GM (1, 1) model for sea level change prediction[C]// Geoscience and Remote Sensing Symposium, 2014. IGARSS'14. (Accepted) (EI)

- [8] Shuai T, Zhang X, Wang S, Zhang L, **Shang K**, et al. A Spectral Angle Distance-Weighting Reconstruction Method for Filled Pixels of the MODIS Land Surface Temperature Product[J]. Geoscience and Remote Sensing Letters, IEEE, 2014, 11(9): 1514-1518. (SCI)
- [9] 张旭凯, 张霞, 杨邦会, 庄智, **尚坤**. 结合海岸类型和潮位校正的海岸线遥感提取[J]. 国土资源遥感, 2013, 25(4): 91-97. (CSCD)

在投或在审文章

- [1] **Shang K**. Monitoring ecological environmental changes of wetlands in coastal areas in the Yellow River delta from 1987 to 2010 using remote sensing techniques[J]. Remote Sensing. (Major Revision) (SCI)
- [2] **尚坤**. 基于植被特征库构建与优化的高光谱植被精细分类研究[J]. 光谱学与光谱分析. (Under Review) (SCI)

博士期间参与专著编写

- [1] 参与修订由张良培、张立福编著的普通高等教育“十一五”国家级规划教材《高光谱遥感》，测绘出版社，2012（负责第六章的修订与增补）。

博士期间参加学术交流

- [1] The Seventh International Symposium on Multispectral Image Processing & Pattern Recognition (MIPPR), Guilin, China, 2011. (Oral Presentation)
- [2] The Second International Workshop on Earth Observation and Remote Sensing Applications (EORSA), Shanghai, China, 2012. (Oral Presentation)
- [3] 第五届全国地理信息系统博士生学术论坛, 南京, 中国, 2012. (口头报告)

博士期间参与项目情况

- [1] 遥感与数字地球研究所“一三五”规划突破三：建立“新型对地观测模拟系统实施方案”
- [2] 国家 863 计划《宽幅高光谱小卫星载荷关键技术研究》项目
 - a) 子课题 2《高光谱数据定标与星载热红外高光谱数据模拟技术研究》（课题编号 2008AA121102-2）
 - b) 子课题 3《热红外高光谱遥感应用技术研究》（课题编号 2008AA121103-3）
- [3] 国家航天局航天遥感论证中心项目《HJ 高光谱载荷典型目标特征提取关键技术研究》（专题编号 A01A04）
- [4] 国家自然科学基金资助项目（编号 40471088）专著《高光谱遥感》
- [5] 高分民用应用系统重大行业应用示范项目《环境保护遥感动态监测信息服务系统先期攻关》子课题《高分矿山环境遥感监测应用示范关键技术研究》（课题编号 E0203/1112-01A-03）
- [6] 海洋公益性行业项目《基于生态系统的环渤海区开发集约用海研究》子课题《环渤海区域开发集约用海卫星遥感监测与评价》（项目编号 201005009）
- [7] 国家科技支撑计划项目《青海湖流域生态环境综合监测系统》课题 3 子课题《青海湖流域土壤与水环境参量反演模块研发》（课题编号 2012BAH31B03-02）
- [8] 基础地质调查与研究项目《全国尾矿地球化学调查与分析》子课题《基于高光谱遥感的尾矿监测系统研究》（课题编号 Y0H0240039）
- [9] 中国环境科学研究院项目《环境保护卫星遥感重点领域学科发展调研与评估》

致谢

2009 年到 2014 年，五年的时间如白驹过隙，转瞬即逝，这五年里，我从一个刚走出象牙塔、懵懂不知自己想要什么的小姑娘，成长为了一个随时可以走向工作岗位的有志青年。这些年，我有过迷茫，有过无助，有过纠结，但是没有后悔过自己的选择。一直觉得自己是极其幸运的，五年的时光一路走来，得到了和蔼可亲的老师们的耐心指导，得到了古道热肠的师兄师姐们的大力帮助，得到了意气风发的同学们的相互鼓励，得到了活泼可爱的师弟师妹们的陪伴支持。

衷心感谢我的导师、我国高光谱遥感领域的开拓者---童庆禧院士，能成为您的学生是我一生的荣幸！清晰的记得那是 2005 年 10 月的一天，在北大地学楼的一间教室里，您为我们介绍遥感、高光谱遥感、“奖状”飞机，那时起，我开始对高光谱遥感产生了浓厚的兴趣。在随后的学习、科研中，您的渊博、严谨、儒雅无时无刻不在感染着我。您不仅是一位举世闻名的学者，对学生我而言，您更是一位平易近人的恩师，“身教胜于言传”，在您的指引下，我能感觉到自己的成长与进步。在今后的科研生活中，我必将以您为榜样，继续开展高光谱遥感方面的研究，为我国的高光谱遥感事业贡献自己的一份力量！

衷心感谢我的导师张霞研究员。张霞老师在学术上严谨求实、思维敏锐，每次与霞老师交流，我都受益匪浅。从本论文的选题、组织结构、数据选择，到算法改进、论文撰写及最终成稿，均得到了霞老师的耐心指导与鼓励，感谢霞老师对我论文倾注的大量时间与精力。同时，作为我的导师，霞老师一直非常清楚什么时候该推我一把什么时候该扶我一下。正是您的督促，使我不断提高对自己的要求、提升自己的科研水平；而在我无助的时候，您及时的点拨与帮助使我渡过一个又一个难关！学生在此向霞老师表达衷心的感谢和深深的祝福！

衷心感谢张立福研究员。福老师治学态度严谨、思维发散，总能在学术方面提出建设性意见，并为我们创造了一个积极向上而又宽松温暖的科研环境。生活中，福老师亦师亦友，提醒我们一定要找到并坚持自己的研究兴趣与研究方向，在求职过程中，福老师也为我提供了无私的帮助。感谢福老师一直以来的关心和帮助，衷心祝福您桃李满天下！

衷心感谢王晋年研究员，王老师有着活跃的学术思想和非常敏锐的洞察力，豁达干练、风趣幽默、充满激情，在繁重的科技工作之余多次参加我们的学术交流，对于我们的学习和科研提出了很多宝贵建议，在此向王老师致以诚挚的敬意与由衷的感谢！

特别感谢郑兰芬研究员。感谢郑老师五年来在科研和生活中给予我的关心和指导，您像亲人一样在方方面面为大家着想，每次见面、无论您多忙都会亲切的询问一下我的近况，希望这篇论文没有辜负您对我的殷切期望！

特别感谢我的本科导师、北京大学的李培军教授。感谢李老师在为我在高光谱遥感分类方面打下的良好基础，和九年来持续给予的指导与帮助，同时在本文的开题与论文撰写过程中，也得到了李老师宝贵的建议，谢谢您！

特别感谢王树东老师在科研、生活中给予的帮助；特别感谢方俊永老师与刘学老师对我生活上的关心和航飞实验中的指导；感谢吴太夏、岑奕、卫征、李儒、杨杭、明涛、吕婷婷、刘佳、付东杰、张红明、张东彦、郭娜、高英倩等师兄师姐在科研、学习中给予的指导；特别感谢赵冬、王潇、张晓红三位师兄师姐在科研和生活中对我一如既往的关心和帮助；感谢刘波、张学文、刘海霞、宋鹏飞、李颖、陈小平、胡顺石、王倩、房丛卉、朱文君、吴松等师兄师姐的帮助与支持；特别感谢孙韬与黄长平两位师兄在求职方面的建议。

感谢 Hasan、段依妮、刘凯、孙雪剑、赵恒谦、姜高珍、甄明、庄智、李雪珂、余晓君、王晶、张良中、孙艳丽、李瑶、石婷婷、王桂珍、童晓伟、张楠、韩冰、李亮、张鹏、林红磊、冯淑娜、孙伟超、石宁卓、王强、陈浩等同门与师弟师妹的陪伴与帮助；感谢覃环虎、张明、兰穹穹、吴向楠、李晶宜、张旭凯、丰明博、朱海涛、王力彦、杨苏新在野外实验中的并肩奋斗；特别感谢帅通与庄智在整个毕业论文撰写、以及小论文修改中给予的支持与鼓励。

感谢高光谱研究室小周姐、龚正姐与王畅三位秘书的默默付出；感谢遥感地球所研究生处吴晓清、杨硕和刘华梅三位老师在学生工作方面给予的关心与帮助；感谢我的九年同窗段祺坤、姚尧、张淼，和多年的好室友蔡丹路、刘爽，以及王静、李霞、刘珠妹、刘一良、彭菁菁、覃帮勇、马胜、曾甜等好友，是你们让我五年的博士生活如此精彩，祝福你们今后在各自的工作岗位中一帆风顺！

感谢我亲爱的爸爸、妈妈、爷爷，和在天国保佑我的奶奶，感谢你们的养育之恩、以及二十多年来对我的无限理解、支持与信任，无论我走在哪里，家都是我最温馨的港湾。感谢我所有的家人一直以来对我的关心与呵护！

感谢我的男友谢一淞，谢谢你对我的默默付出和包容体贴，无论是学术还是生活，总能为我提出非常好的建议。很幸运有你做我最坚实的后盾，让我有勇气追求自己想要的生活，祝你早日梦想成真！

最后，衷心的感谢本文的评阅专家与参与答辩的各位老师，谢谢你们提出的宝贵意见！

尚坤

2014 年 5 月

于天地科学园区