

不同季相针叶树种高光谱数据识别分析

官 鹏 浦瑞良 郁 彬

(美国伯克利加州大学)

摘 要 利用高分辨率光谱仪在实地测得的光谱数据来识别美国加州的 6 种主要针叶树种。树冠阴面和阳面的高光谱数据分别在 1996 年夏、秋测得。首先对原始光谱数据作简单处理, 然后进行 6 种数据变换: 对数变换、一阶微分变换、对数变换后一阶微分变换、归一化变换、归一化变换后一阶微分变换及归一化后对数变换。采用相邻窄波段逐步加宽的办法, 测试不同波段宽度对树种识别精度的影响。所有的变换方法及波段宽度试验最后均由神经网络算法产生的树种分类精度来评价。试验结果表明对数变换后一阶微分和归一化变换后一阶微分能够获得高于 94% 的平均精度; 归一化变换和微分处理能够限制阴影的影响; 20nm 的波段宽度用于识别此 6 种针叶树种是较为理想的。我们发现太阳高度角变化对树种识别影响不大。

关键词 高光谱数据, 数据变换, 波段宽度, 针叶树种识别

1 引 言

我们曾利用实测高光谱数据对加州塞拉内华达山脉常见的 6 种针叶树种进行识别分析¹⁾。所用分析数据源自 1995 年秋; 分类算法为误差后向传递的神经网络(简称 NN)和判别分析方法(简称 DA)。所得主要结论包括: (1) NN 算法优于 DA, 在训练样本小的情况下尤为如此; (2) 由于数据微分处理能够限制低频背景光谱对目标光谱的影响^[1,2], 利用微分光谱数据能获得较高的分类精度; (3) 可见光波段对树种的识别能力要强于近红外; (4) 一个较小的选择性波段集比起利用全部波段能获得更高的识别精度。

先前的研究表明, 对高光谱数据进行简单的变换(如微分变换)能够有效地改善识别精度。从而驱使我们进行本实验——寻求适于此 6 种针叶树种识别的最佳变换方法。另外, 高光谱数据能被用来识别一些宽波段数据不能或非常困难的目标物(如光谱特性非常相似的针叶树种), 但高光谱数据量大, 会带来计算时间和贮存数据的问题, 同时也会增加参数估计所需的样本数。那么多大的波段宽度变化范围对识别针叶树种较为理想呢? 本文展示了

不同数据变换方法对树种识别影响的实验结果, 并建议使用理想的波段宽窄范围。

2 研究地区及数据采集

2.1 研究地区

位于美国加利福尼亚州塞拉内华达山脉中部地区, 光谱测区落在伯克利加州大学的 Blodgett 林业实验站。植被主要由当地常见的塞拉混交针叶林类型, 主要有 5 种针叶树种: 糖松(SP, *Pinus Lambertiana*), 美国西黄松(PP, *Pinus Ponderosa*), 白冷杉(WF, *Abies concolor*), 花旗松(DF, *Pseudotsuga menziesii*), 和香肖楠(IG, *Calocedrus decurrens*), 及一种硬阔树种——加州黑栎(*Quercus kelloggii*), 另外, 一种属塞拉内华达山脉乡土树种但在 Blodgett 林站并不常见的美国巨杉(GS, *Sequoiadendron giganteum*)。Blodgett 林站的地质地貌、土壤和其它林下草本植物详见 Gong 等¹⁾。

2.2 光谱反射率数据的采集

利用 PSD1000 光谱仪^[3]在野外测量光谱数据。这种光谱仪可以测量波长范围 210—1050nm, 波段宽 < 1nm 的高光谱分辨率反射率数据(实际分辨率

1) Gong, P., R. Pu, B. Yu. Conifer species recognition: exploratory analysis of in situ hyperspectral data. *Remote Sensing of Environment*, in press.

收稿日期: 1997-10-27; 收到修改稿日期: 1998-03-03

大约 2.6nm)。仪器的视场角约 22°。我们采用与先前同样的方法¹⁾在 Blodgett 林验站的两块小斑——放牧区和围篱区,使用 PSD1000 光谱仪分别测量 6

种针叶树 (DF, GS, IC, PP, SP 和 WF) 的光谱 (表 1)。光谱分别在 1996 年 6 月 2—3 日和 1996 年 11 月 1—2 日测得。为测量方便,我们仅测 4—7 年生

表 1 本研究使用的高光谱数据样本
Table 1 Hyperspectral samples used in this study

树种	1996年6月				1996年11月				合 计	
	放牧区		围篱区		放牧区		围篱区		阳面	阴面
	阳面	阴面	阳面	阴面	阳面	阴面	阳面	阴面		
DE	10	10	12	12	10	10	10	10	42	42
GS	10	10	12	12	10	10	10	10	42	42
IC	10	10	12	12	10	10	10	10	42	42
PP	10	10	12	12	10	10	10	10	42	42
SP	10	10	12	12	10	10	10	10	42	42
WF	10	10	12	12	10	10	10	10	42	42
合计	60	60	72	72	60	60	60	60	252	252

的幼树,高度一般在 4m 之内。测点高度低于 1.5m。为了分析不同光照条件对光谱值的影响,连续测量树冠的阴面和阳面。测量中始终保持探头铅垂向下。测量条件:温度 20℃—30℃,干燥、微风,除 1996 年 11 月测围篱区有变化的薄云外,其余测区和时期均为晴空。测量中标准板白光和暗光谱每隔 5—10min 重复一次,以减少可能的光照变化引起树冠光谱的测量误差。当天的测量时间被控制在 11:00—13:00。

3 方 法

3.1 高光谱数据变换

为了减少数据量和限制随机噪声水平,我们先对波长覆盖 300—900nm 的 489 个原始波段合并到波长宽约 3.4nm 的 163 个波段 (简称简单处理)。没有使用信号弱且噪声大的 300—900nm 以外的测量值。我们将简单处理的光谱用 $R = (r_1, r_2, \dots, r_n)$, $n = 163$ 表示。相对于 R 的 6 种变换方法包括:

$$(1) \text{ 对 } R \text{ 的一阶微分变换, } d(R) = \left(\frac{(r_3 - r_1)}{\Delta\lambda}, \frac{(r_4 - r_2)}{\Delta\lambda}, \dots, \frac{(r_n - r_{n-2})}{\Delta\lambda} \right), \Delta\lambda \text{ 为两倍波段宽;}$$

$$(2) \text{ 对 } R \text{ 的对数变换, } \log(R) = (\log(r_1), \log(r_2),$$

$\dots, \log(r_n))$;

$$(3) \text{ 对 } R \text{ 的归一化变换, } N(R) = \left(\frac{nr_1}{\sum r_i}, \frac{nr_2}{\sum r_i}, \dots, \frac{nr_n}{\sum r_i} \right), \text{ (这里使用了各波段光谱之平均值,以免变换后的值太小,);}$$

$$(4) \text{ 对 } \log(R) \text{ 的一阶微分变换, } d(\log(R));$$

$$(5) \text{ 对 } N(R) \text{ 的对数变换, } \log(N(R));$$

$$(6) \text{ 对 } N(R) \text{ 的一阶微分变换, } d(N(R)).$$

利用 1996 年 11 月放牧区数据作成图 1(a)—(f) 以示上述变换算法 (除 (5))。由图 1(a) 可见,在那些不同树种间曲线斜率相近的区段,其一阶微分值非常接近 (图 1(b))。斜率接近零的区段,其微分值接近零。这有利于限制低频背景光谱 (通常是土壤、凋落物及死亡植物光谱) 对目标光谱的影响^[1,4]。光谱反射值经对数变换后,趋向于减少因光照条件变化的乘性因素影响。然而,仅对光谱做对数变换是不够的,还需做微分处理,这样才能取得较好效果 (方法 (4))。从理论上讲归一化处理 (图 1(e)) 可完全消除光照条件差异的乘性因素影响。对归一化变换结果做微分处理 (方法 (6)) 能取得与变换方法 (4) 的相似效果。这是因为不仅去除因光照条件变化的乘性因素的影响,而且减弱了耐加的低频噪声的影响。

1) Gong, P., R. Pu, B. Yu. Conifer species recognition: exploratory analysis of in situ hyperspectral data. *Remote Sensing of Environment*, in press.

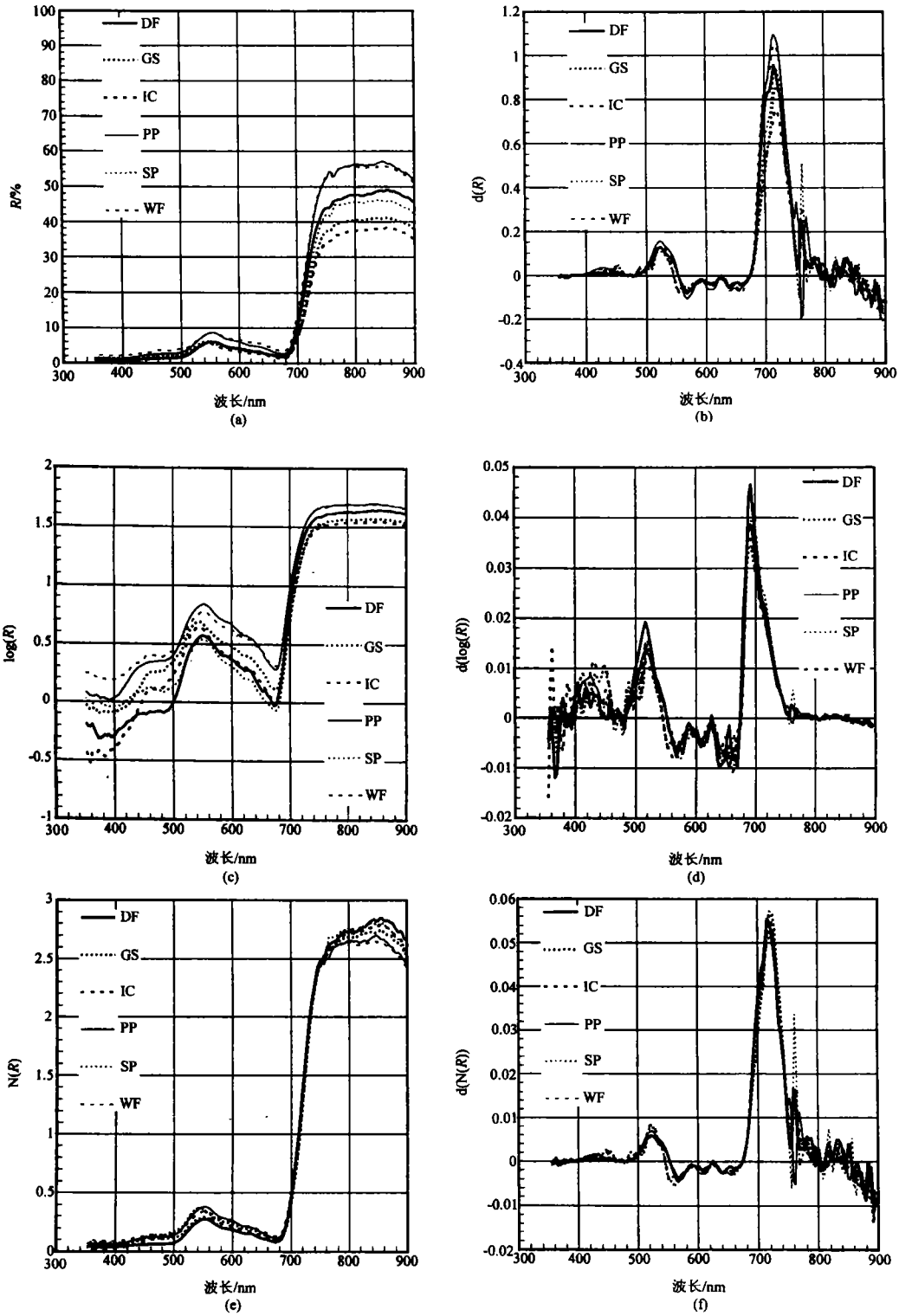


图1 1996年11月放牧区数据经5种数据变换和简单处理产生的结果

Fig.1 Results of hyperspectral data transformation with measurements of Grazed site of Nov. 1996

3.2 波段宽度变化

虽然高光谱数据能俘获丰富的光谱信息,但它

有两个不足之处:数据量庞大及传感器系统制作困难。因此有必要平衡波段宽度和满足非常专用的高光谱遥感系统的应用要求。为了测试波段宽度变化

对树种识别精度的影响,我们采用逐步合并相邻窄波段并取平均值的方法来达到波段宽度变化的目的。我们采用 3 套波段长范围:一是仅取 SPOT 波段 1, 2 和 3 覆盖的波长范围的波段作逐步合并平均处理,基本波段数为 79;二是仅取 TM 波段 1, 2, 3 和 4 覆盖的波段作相应的合并和均值处理,波段数为 106;三是利用全部波段(163 个,覆盖 300—900nm)作相邻逐步合并平均处理。

3.3 人工神经网络算法的应用

我们采用人工神经网络(NN)算法识别树种。NN 训练机制是误差后向传递算法^[5]。本研究采用的 NN 程序是根据 Pao 改编的^[6]。该种神经网络要求一部分样本被用来训练 NN,当训练成熟(用最小误差标准或取最佳测试精度衡量)后,可用于对测试样

本进行识别分类。许多研究人员已用 NN 算法进行遥感数据和其它数据进行分类^[7-9]。

在本次实验中,三分之一作为训练样本,剩余的三分之二样本作为测试样本。测试样本的分类精度被用做某种数据变换方法的优劣及波段宽窄的指标。

4 结果与分析

4.1 不同数据变换方法的效果

表 2 列出了用 NN 算法计算的 7 套数据的测试精度(用 % 表示)。所采用的 NN 结构参数相似于 Gong 等论文^[1]中的(即学习系数 η 为 0.2 或 0.3, 动量系数 α 为 0.6 或 0.7, 一层中间层神经元数为 50)。表 2 右端列出了阳面、阴面及阴阳面合起来的各自的平

表 2 不同变换方法的树种识别精度

Table 2 Summary of species recognition accuracies

(%)

变换方法	1996年6月,放牧区		1996年6月,围篱区		1996年11月,放牧区		1996年11月,围篱区		平均		
	阳面	阴面	阳面	阴面	阳面	阴面	阳面	阴面	阳面	阴面	总平均
R	75.8	65.3	81.9	90.3	74.2	60.8	75.0	73.3	76.7	72.4	74.6
d(R)	90.8	81.7	93.1	89.6	95.8	97.6	94.2	90.0	93.5	89.7	91.6
log(R)	78.3	60.0	77.8	89.6	60.1	65.7	60.8	70.8	69.3	71.5	70.4
N(R)	72.5	66.7	84.0	91.7	79.8	78.4	76.7	80.8	78.3	79.4	78.8
d(log(R))	93.3	92.5	97.2	97.9	96.6	99.2	93.3	96.7	95.1	96.6	95.8
log(N(R))	71.7	65.0	85.4	93.8	76.5	71.8	75.0	79.2	77.2	77.4	77.3
d(N(R))	91.7	85.0	95.1	95.1	98.3	97.6	97.5	96.7	95.7	93.6	94.6

均分类精度。从这些结果中,我们可以看出对数变换后再取一阶微分的方法能取得最佳效果,不管是从树冠阳面还是阴面数据得到的精度都超过 90%,总平均为 95.8%。比之稍差的是由正归化后再取一阶微分的方法,它的平均精度为 94.6%,仅低于前者 1% 左右。直接对简单处理数据进行一阶微分处理亦取得 91.6% 的较高平均精度。与未经过任何变换(简单变换)的 74.6% 精度相比,最高的 3 种变换分别超过其 21.2%, 20.0% 和 17.0%。3 种能产生最高平均分类精度的方法均进行了微分变换。这充分说明高光谱数据的微分处理的确能产生较高的精度。大部分由微分变换产生的精度高于简单处理的 15%—30%,仅对简单处理数据采用对数变换并不能改善分类效果,反而下降了 4.2%,而单独进行正归化变换由于限制了因光照条件变化引起的乘性因素

影响而亦能改善精度 4.2%。

4.2 阴影影响

表 2 中 6 种变换方法除单独对数变换方法外,其余 5 种对减少阴影对分类精度的影响均有效果。有的方法已使平均阴面数据精度超过平均阳面精度。因此采用这些变换方法对减少阴影影响是有效的。其中以归一化处理和对数处理变换后的一阶微分处理尤佳。

4.3 太阳高度角的影响

本实验分析的数据分别在 6 月初(夏季)和 11 月初(秋季)采集。太阳高度角相对于 Blodgett 林验站的地理位置在此期间内变化了 20° 以上。虽然 11 月的数据可能会加大阴影对分类精度的影响(阴面阳

1) Gong, P., R. Pu, B. Yu. Conifer species recognition: exploratory analysis of in situ hyperspectral data. *Remote Sensing of Environment*, in press.

面的平均精度差异在 6 月是 1.1%，而在 11 月是 7.6%)。但经过上述 6 种变换方法后(除方法(2))，不同季节采集的数据能取得相似的分类精度。因此太阳高度角的变化引起的树种识别精度的差异已不明显。

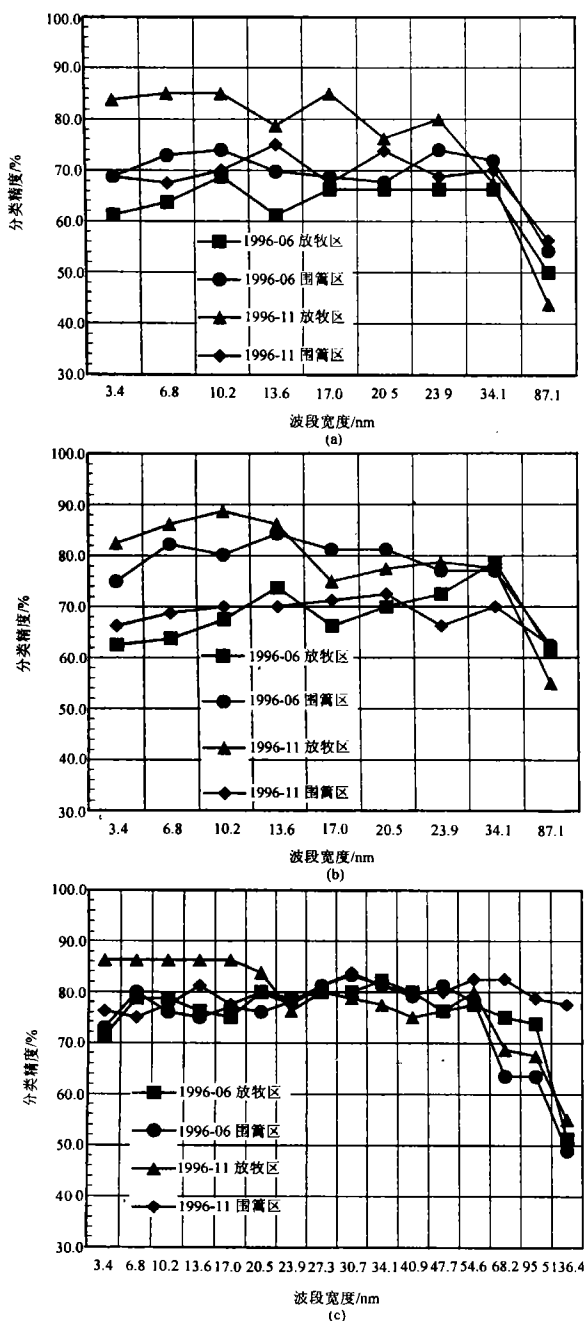


图2 由不同波段宽度用简单处理的高光谱数据产生的树种识别精度

(a)由覆盖 SPOT 3个波段范围产生的; (b)由覆盖 TM 4个波段范围产生的; (c)由利用全部波段产生的

Fig.2 Species recognition accuracies obtained from the hyperspectral data as a function of band width

4.4 波段宽窄的影响

图 2 和图 3 分别展示了简单处理和一阶微分数

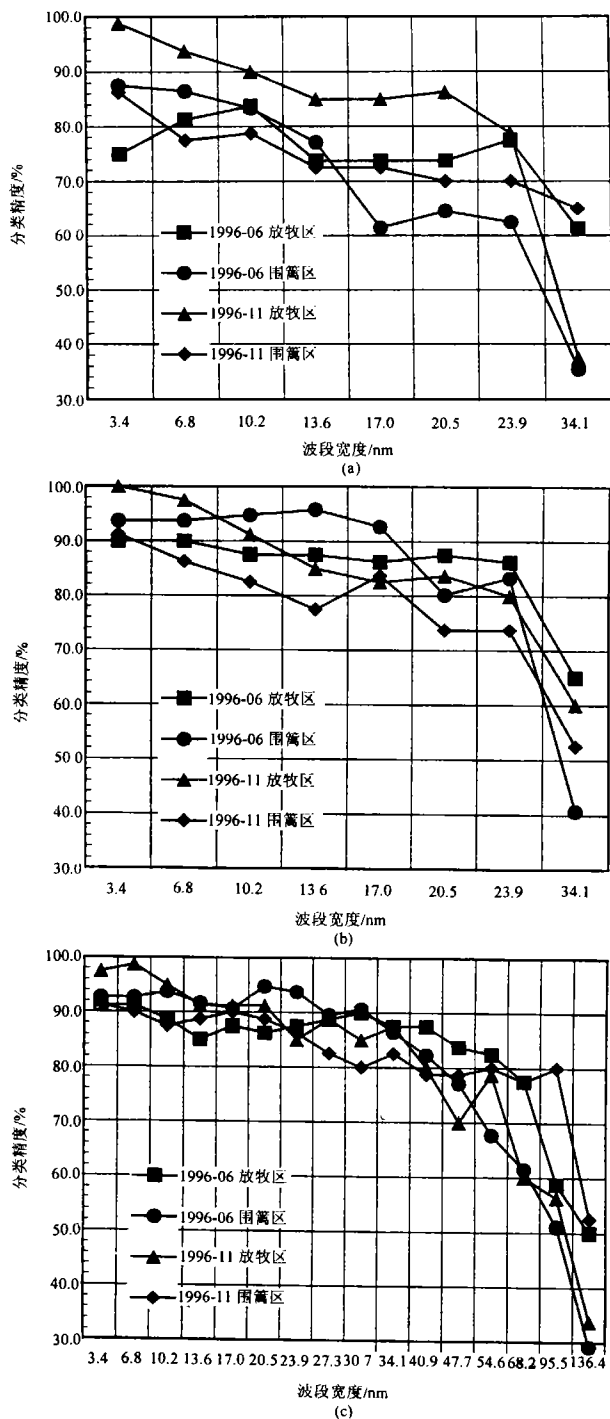


图3 由不同波段宽度用一阶微分数数据产生的树种识别精度

(a)由覆盖 SPOT 3个波段范围产生的; (b)由覆盖 TM 4个波段范围产生的; (c)由利用全部波段产生的

Fig.3 Species recognition accuracies obtained from the derivative hyperspectral data as a function of band width

据因使用不同波段宽度而产生的分类精度变化。分类精度已综合了阴阳面的效果。图 2(a)–(c)分别是对应于 3 个 SPOT 波段范围、4 个 TM 波段范围和全部波段参与的结果。同样地,图 3(a)–(c)是由使用一阶微分变换的数据产生的相应 3 种波段范围的结果。由图 2 可见:最窄的波段宽(3.4nm)并未导致最高分类精度;精度的上限对于图 2(a)–(c)是相似的;几乎每条曲线(图 2)有相似的变化趋势,即在一定的波段宽度之前,曲线在 10% 之内变化,但过了此波段宽度就突然下降;此波段宽度对于图 2(a)和 (b)是近乎 34.1nm,而对于图 2(c)是 47–54nm。因此,对于本次识别树种而言,假如直接使用简单处理后的数据,波段宽度选在 30nm 左右就够了。

与图 2 相比,图 3 表明较窄的波段宽度能取得较高的精度。图 3 虽然显示随着波段宽度的增加,其分类精度是递减的,但精度在 10–20nm 波段范围内下降缓慢。这说明比起简单处理后的光谱数据,微分光谱要求更窄的波段宽度。如要设计一个传感器系统例如数字相机^[4],并想获取光谱分辨率窄于 10nm 是相当困难的。假如能维持光谱分辨率在 20nm 左右是很重要的。因为用一般的数字相机经过滤机制

等处理容易实现这一目标。由于微分光谱用于识别树种比非微分光谱能取得更高的精度,因此建议波段宽为 20nm 或更窄一点可用于相似于本实验的应用目的。

4.5 训练样本的大小

上述讨论的实验是基于训练样本占总样本的三分之一(40 个左右),测试样本用剩余的三分之二。那么训练样本代表性如何呢?为了回答这个问题,我们将 7 套数据的每套总样本随机分成 3 份,第 1 份用作训练,第 2,3 份作测试样本,记下测试样本的分类精度;然后将第 2 份作训练样本,第 1,3 份作测试样本,记下测试样本的分类精度;最后将第 3 份作训练样本,第 1,2 份作测试样本并记下其分类精度。这样可以得到 3 套互不重叠的训练样本所产生分类精度。图 4(a)和 (b)分别用 1996 年 6 月和 11 月围篱区数据得出 3 套测试样本的分类精度。图中 T1, T2 和 T3 分别表示由第 1 份样本、第 2 份样本和第 3 份样本作训练样本训练 NN 而产生的 3 套测试精度。由图 4 可见,大部分不同训练样本产生的测试样本精度在 10% 内变化,由数据变换方法 $d(\log(R))$ 和

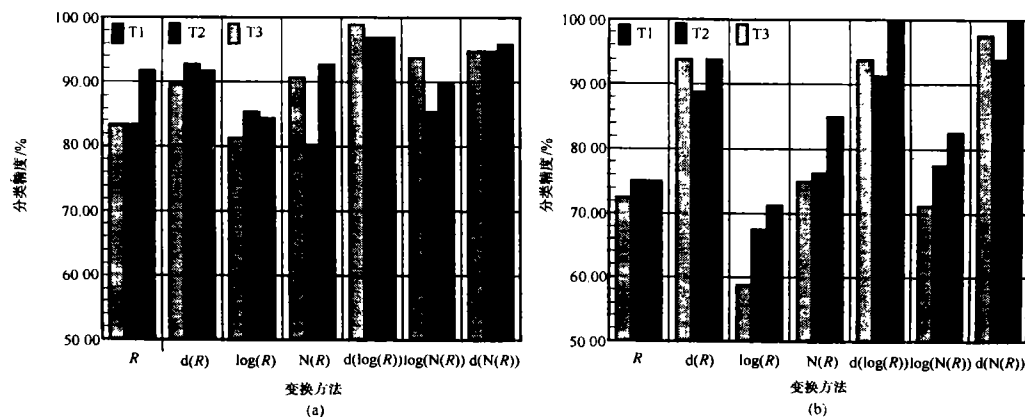


图4 利用围篱区光谱数据由相互独立的3套训练样本产生的测试精度

(a) 1996年6月测; (b) 1996年11月测

Fig.4 Species recognition accuracies obtained from the three sets of non-overlapping training samples collected at the Fenced site

$d(N(R))$ 产生的 3 套测试精度均高于 90%。类似的结果也在放牧区得到。这些结果说明,基于本实验的样本特征,取每套数据样本的 40 个左右作训练并不嫌少。

5 结 论

基于我们的实验结果及分析,可以得出如下结

论:

(1) 对野外高光谱分辨率仪器测得的 6 种针叶树种数据仅作简单的平滑和波段合并处理,然后利用误差后向传递的神经网络算法来识别此 6 种树种,效果并不理想。一些高光谱数据的变换方法能够改善树种的识别精度。两种最理想的变换方法是先对简单处理的数据进行对数变换,然后再计算其一阶微分值;或在简单变换数据作一阶微分之前先对其作

归一化处理。

(2) 测自阴面的高光谱数据比起阳面的数据可能会产生较低的树种识别精度。然而,应用归一化处理或对简单处理的数据先作对数运算后再作一阶微分处理则能够减少二者的精度差异。

(3) 太阳高度角变化引起的光谱变化经过上述处理后对树种识别的精度影响不大。

(4) 识别针叶树种最好利用波段宽为 20nm 或更窄一点的光谱数据。

参 考 文 献

- 1 浦瑞良, 宫鹏, 约翰 R. 米勒. 美国西部黄松叶面积指数与高光谱分辨率 CASI 数据的相关分析. 环境遥感, 1993, 8(2): 112—124.
- 2 张良培, 郑兰芬, 童庆禧等. 利用高光谱对生物变量进行估计. 遥感学报, 1997, 1(2): 111—114.
- 3 ANCAL, Inc. C-SPEC Data Acquisition and Manipulation Program. Users Guide & Operating Instructions, Version 1.5, 1995.
- 4 宫鹏, 史培军, 浦瑞良, 郭华东. 对地观测技术与地球系统科学. 北京: 科学出版社, 1996.
- 5 Rumelhart, D. E., G. E. Hinton, R. J. Williams, Learning internal representations by error propagation. In Parallel Distributed Processing—Explorations in the Microstructure of

- Cognition. Vol. 1: Edited by D. E. Rumelhart and J. L. McClelland. The MIT Press: Massachusetts. 1986, 318—362.
- 6 Pao, Y. Adaptive Pattern Recognition and Neural Networks. Addison and Wesley, New York. 1989, 113—299.
 - 7 Gong, P. Integrated analysis of spatial data from multiple sources: using evidential reasoning and an artificial neural network for geological mapping. *Photogrammetric Engineering and Remote Sensing*, 1996, 62(5): 513—523.
 - 8 Gong, P., R. Pu, J. Chen. Mapping ecological land systems and classification uncertainties from digital elevation and forest cover data using neural networks. *Photogrammetric Engineering and Remote Sensing*, 1996, 62(11): 1249—1260.
 - 9 Hepner, G. F., T. Logan, N. Ritter, N. Bryant, Artificial neural network classification using minimal training set. *Photogrammetric Engineering and Remote Sensing*, 1990, 56: 469—473.

作 者 简 介

宫鹏, 1965年生。1986年毕业于南京大学, 1990年获加拿大滑铁卢大学博士学位。主要研究有关遥感生态测量学、遥感地方病流行环境、地理信息系统应用。发表论文 100 余篇。著有《对地观测技术与地球系统科学》和《实用地理信息系统》, 由科学出版社出版。现担任《地理信息科学》主编和《国际遥感杂志》美国编辑。

Conifer Species Recognition with Seasonal Hyperspectral Data

Gong Peng Pu Ruiliang Yu Bin

(University of California at Berkeley, USA)

Abstract In situ hyperspectral data obtained with a high spectral resolution radiometer were analyzed for identification of six conifer species. Hyperspectral data were measured in the summer and late fall seasons from both the sunlit and shaded sides of canopies. An artificial neural network algorithm was applied for the identification purpose. Six types of transformation were applied to the hyperspectral data R preprocessed with a simple smoothing followed by band merging. These include $\log(R)$, first derivative of R , first derivative of $\log(R)$, normalized R , first derivative of normalized R , and $\log(N(R))$.

First derivative of $\log(R)$ and first derivative of normalized R resulted in best species recognition accuracies with greater than 94% average accuracies. The effect of hyperspectral data taken from the shade sides of tree canopies can be minimized by applying normalization or by taking derivative after applying logarithm to the preprocessed data. We found that a big difference in solar angle due to seasonality did not cause noticeable difference in accuracies of species recognition. A band width of 20nm or narrower is recommended for the recognition of the six species.

Key words Hyperspectral data, Data transformation, Band width, Conifer species recognition