

支持向量机复合核函数的高光谱显微成像木材树种分类

赵 鹏*, 唐艳慧, 李振宇

东北林业大学信息与计算机工程学院, 黑龙江 哈尔滨 150040

摘 要 采用体视显微高光谱成像方法, 构建木材树种分类识别模型。利用 SOC710VP 体视显微高光谱图像采集系统获取可见光/近红外(372.53~1 038.57 nm)波段内的木材高光谱图像。首先, 采用 ENVI 软件提取木材样本感兴趣区域(ROI)的平均光谱, 分别采用连续投影算法(SPA)和竞争性自适应重加权算法(CARS)对光谱数据进行降维。再利用支持向量机(SVM)分别建立木材样本采集波段和特征波长下的分类模型。然后, 在空间维采用第一主成分图像, 计算基于灰度共生矩阵(GLCM)的木材纹理特征。在 0°, 45°, 90°和 135°四个方向计算能量、熵、惯性矩、相关性等 16 个特征参数后输入 SVM 进行木材树种分类处理。最后, 采用四个复合核函数 SVM 进行光谱维和空间维的特征融合及分类识别。20 个树种的分类实验结果表明, CARS 的特征波长选择效果和运行速度较好一些, 采用普通 SVM 进行木材光谱维特征分类处理时, 测试集分类准确率达到 92.166 7%。采用基于 GLCM 的木材空间维纹理特征时, 采用普通 SVM 的测试集分类准确率是 60.333 0%, 具有较低的分类精度。在将光谱维和空间维纹理特征进行数据融合及分类处理时, 采用复合核函数 SVM 分类具有更好的效果。采用第二个复合核函数的 SVM 分类精度最高, 测试集分类正确率是 94.166 7%, 运行时间为 0.254 7 s。另外, 采用第一个和第三个复合核函数的 SVM 的测试集分类准确率分别是 93.333 3%和 92.610 0%, 运行时间分别为 0.180 0 和 0.260 2 s。可以看出, 采用这 3 种复合核函数的 SVM 进行木材树种分类, 分类精度都高于采用普通 SVM 的光谱维或者空间维的分类识别精度。因此, 利用体视显微高光谱成像和复合核函数 SVM 可以提高木材树种分类精度, 为木材树种快速分类提供了参考。

关键词 木材树种识别; 高光谱成像; 复合核函数; SVM; 特征融合

中图分类号: TP391 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)12-3776-07

引 言

在木材贸易活动中, 对木材树种的识别、树种等级的判定等历来是一项较为艰辛而又需要实际操作经验的工作, 不同树种的木材其木材材性、物理性能及价格差别很大。基于计算机分析处理的木材树种识别技术主要分为两大类, 即微观的木材组织结构分析处理法^[1-2]和宏观的木材表面特征分析处理法^[3-4]。从使用的技术手段来看, 也可以分成图像颜色特征分类法、图像纹理特征分类法^[1-2]和光谱曲线特征分类法^[3-4]。例如, Yusof 的研究组使用木材样本图像的管孔纹理分布特征, 开展了 50 余种热带木材树种的分类识别研究。他们使用 Basic Grey-Level Aura Matrix (BGLAM) 来计算管孔分布纹理特征, 再应用模糊分类或者模糊推理方法进行两

级的层次化的分类处理^[1-2]。Yusof 等还使用了核遗传算法提取了管孔纹理的非线性特征, 进一步用于木材树种分类处理^[5]。

高光谱成像拥有波段多、分辨率高和图谱合一的优点, 集光谱维信息和空间维信息于一体, 已经在远程遥感分类领域和样本组织成分检测领域得到应用。例如, Liu 等利用高光谱成像和偏最小二乘回归模型进行全蛋粉掺假检测处理^[6]。Wang 等利用高光谱成像, 对比线性逐步判别分析法和非线性偏最小二乘支持向量机进行建模, 完成对腐烂、病害及正常梨枣的分类^[7]。Sun 等对原始光谱做特征提取和特征选择, 建立支持向量机模型对掺假大米进行检测^[8]。Deng 等融合高光谱图像光谱、纹理和形态特征对水稻种子单粒进行品种快速鉴别^[9]。又如, 远程遥感中人们使用高光谱图像中的空间维和光谱维信息进行数据融合, 提高了地物分类

收稿日期: 2019-03-18, 修订日期: 2019-07-24

基金项目: 国家自然科学基金面上项目(31670717), 中央高校基本科研业务费专项基金项目(2572017EB09), 黑龙江省自然科学基金面上项目(C2016011)资助

作者简介: 赵 鹏, 1972 年生, 东北林业大学信息与计算机工程学院教授 e-mail: bit_zhao@aliyun.com * 通讯联系人

(例如沙地、灌木、耕地、林地、草地等类别)识别的精度^[10-12]。

当前,国内外对于木材树种光谱分类的研究主要集中在近红外波段的光谱反射率或者透射率特征进行分类研究^[13-14]。本文利用高光谱成像对 20 种木材树种进行分类,使用了 SVM 的复合核函数,实现了木材样本高光谱图像的光谱维和空间维特征的数据融合,提高了木材树种的分类识别精度。

1 实验部分

1.1 木材树种样本

采用 5 种东北常见的杨木、桦木、樟子松、白松、落叶松和 15 种其他树种分别为巴西花梨、红花梨、印尼菠萝格、美国红橡、非洲卡斯拉、缅甸柚木、海棠木、克隆木、金丝柚木、漆木、水煮柚木、南美柚木、唐木、香樟木、白蜡木作为实验样本。将其加工成木块。每个树种采集 80 幅高光谱图像,则共有 1 600 幅高光谱图像作为实验样本。

1.2 高光谱图像采集

图 1 为显微高光谱成像系统示意图,该系统主要是由便携式高光谱成像仪(SOC710VP, USA)和 CCD 相机两部分构成,成像波段为可见光/近红外波段,光谱分辨率为 5.2 nm,光谱范围 372.53~1 038.57 nm。采集木材横切面的显微高光谱影像,所采集的高光谱图像尺寸是 520×696×128。台式机电脑配置为 i7 处理器,4G 运行内存。

这里强调一点,考虑到仪器的价格,我们使用了价格较低的便携式高光谱成像仪,它的成像波段在可见光和少部分近红外波段内。因此,实验时应该保持外界环境例如温度、湿度、样本保存时间等因素尽量恒定,尽量确保各个木材树种样本的颜色保持不变。否则,可能造成可见光波段内的光谱图像变化,影响后续的特征提取和分类识别精度。

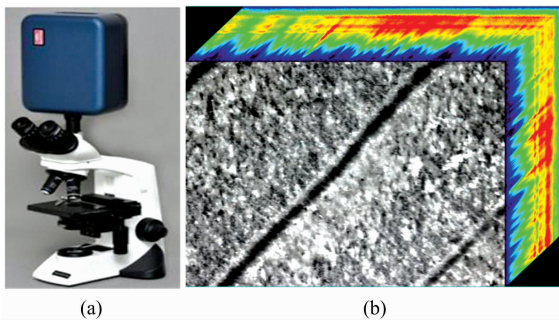


图 1 木材体视显微高光谱成像系统

(a): 高光谱成像系统; (b): 采集的木材显微高光谱图像立方体(木材树种是印尼菠萝格)

Fig. 1 A wood stereomicroscopic hyper-spectral imaging system

(a): The hyper-spectral imaging system;

(b): The wood hyper-spectral image cube for wood species "Merbau"

由于 CCD 相机暗电流及不均匀照明强度会影响高光谱图像质量,应校准原始图像。校正公式如式(1)

$$R = \frac{I_O - I_B}{I_w - I_B} \quad (1)$$

式(1)中, I_O 为原始未修正的高光谱图像, I_B 为扫描得到的黑板校正图像, I_w 为白板校正图像,从而得到校正后的图像 R 。

1.3 高光谱图像处理

1.3.1 ROI 的选择

采用 ENVI5.1 软件中 region of interest 工具进行高光谱图像感兴趣区域(region of interest, ROI)的提取,以此增加实验样本数量。本实验在样本中心区域手动选取像素的正方形区域作为 ROI,每个图像选取 3 个 ROI,共提取 4 800 个 ROI,然后计算 ROI 内所有像素点光谱反射率的平均值。

1.3.2 光谱预处理

为了提取高光谱图像中光谱有效信息,提高建模稳健性和预测精度,需要对光谱数据进行预处理来减少光谱信号中的光散射、高频噪声等干扰信息。常用的预处理方法有中值滤波平滑、变量标准化校正(standard normal variate, SNV)、多元散射校正(multiplicative scatter correction, MSC)、Savitsky-Golay 平滑算法(Savitzky-Golay smoothing, SG)、一阶导数(first derivative, FD)等。本工作采用 SNV, MSC 和 SG 三种预处理方法,根据模型的预测精度选择最佳的预处理方法进行建模。

1.3.3 光谱特征选择

所获得的高光谱图像样本具有波段多、数据量大等特点,存在大量冗余和多重共线性问题,在采集波段光谱中很多波长与分类预测无关,因此删除这些不相关波长可以提高运行速度。采用连续投影算法(successive projection algorithm, SPA)和竞争性自适应重加权算法(competitive adaptive reweighted sampling, CARS)进行特征波长选择。

1.3.4 纹理特征提取

高光谱图像具有图谱合一的特点,除了对高光谱图像的光谱维进行特征选择外,还需要对其空间维进行纹理计算。这里采用经典的灰度共生矩阵(gray level co-occurrence matrix, GLCM),GLCM 基于像素灰度的空间相关性表示纹理特征。首先对高光谱图像样本进行主成分变换,最后选择第一主成分图像,距离参数值设为 1。对第一主成分图像基于 GLCM 分别提取图像中的 0° , 45° , 90° 和 135° 方向上的能量、熵、惯性矩、相关性 16 个特征值作为纹理特征。

1.4 SVM 复合核分类器

1.4.1 支持向量机

支持向量机(support vector machine, SVM)是由 Vapnik^[15]等提出,是一种以结构风险最小化原则为基础的模式识别算法,克服了在传统机器学习中维数灾难问题,实验表明它在小样本数据集中的分类问题上有显著优势。SVM 求解最优分类超平面问题等价求解如下方程式(C, ϵ_i 是相应参数, ϕ 是非线性映射函数, $y_i \in [-1, 1]$)

$$\begin{aligned} \min_{w, b, \epsilon_i} & \left[\frac{1}{2} \|w\|^2 + C \sum_i \epsilon_i \right] \\ \text{s. t. } & y_i (\langle \phi(x_i), w \rangle + b) \geq 1 - \epsilon_i, \\ & \epsilon_i \geq 0, \forall i = 1, \dots, n \end{aligned} \quad (2)$$

引入 Lagrange 乘子 α_i 可以推导出二次规划问题

$$\begin{aligned} \max_{\alpha_i} & \left[\sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \right] \\ \text{s. t.} & \sum_i \alpha_i y_i = 0, \\ & 0 \leq \alpha_i \leq C, i = 1, 2, \dots, n \end{aligned} \quad (3)$$

其中 $K(\cdot)$ 是满足 Mercer 条件的核函数, 相应的 SVM 判别函数为

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

1.4.2 复合核函数

为了提升 SVM 分类器的性能, 选择合适的核函数至关重要。因此, Gustavo 等提出了基于复合核函数的新型 SVM 分类器^[16], 将其应用于高光谱遥感图像的地物分类中。在遥感图像的地物分类中, 处理的基本数据单元是像素, Gustavo 等重新定义一个新的像素点特征矢量 $x_i = \{x_i^s, x_i^w\}$, 它是由光谱维特征 $x_i^w \in R^{N_w}$ 和空间维特征 $x_i^s \in R^{N_s}$ 组成, 其中 N_w 代表光谱维特征个数, N_s 代表空间维特征个数。相应的, K_s 代表空间核函数, K_w 代表光谱核函数, 用它们或者复合核函数替换式(3)和式(4)中原始的核函数 K , 具体的 4 种复合核函数表示如下^[16]。

(1) Stacked Features Approach. 这里重新定义一个像素点特征向量 $x_i = \{x_i^s, x_i^w\}$, 它是由该像素点空间维特征和光谱维特征直接连结, 相应复合核函数表达式为

$$K_{(s,w)} \equiv K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle \quad (5)$$

(2) Direct Summation Kernel. 在希尔伯特空间 H 中定义 $\varphi_1(\cdot)$ 和 $\varphi_2(\cdot)$ 两个非线性变换, 再生成新的变换组合 $\phi(x_i) = \{\varphi_1(x_i^s), \varphi_2(x_i^w)\}$, 进而构造复合核函数, 这里 $\dim(K) = \dim(K_s) = \dim(K_w) = n \times n$ 。

$$\begin{aligned} K(x_i, x_j) &= \langle \phi(x_i), \phi(x_j) \rangle = \\ & \langle \{\varphi_1(x_i^s), \varphi_2(x_i^w)\}, \{\varphi_1(x_j^s), \varphi_2(x_j^w)\} \rangle = \\ & K_s(x_i^s, x_j^s) + K_w(x_i^w, x_j^w) \end{aligned} \quad (6)$$

(3) Weighted Summation Kernel. 这种情况可以表示为 B 的加权形式, 其中 $0 \leq \mu \leq 1$, μ 为权因子

$$K(x_i, x_j) = \mu K_s(x_i^s, x_j^s) + (1 - \mu) K_w(x_i^w, x_j^w) \quad (7)$$

(4) Cross-Information Kernel. 它考虑光谱维信息和空间维信息的交叉关系, 在希尔伯特空间 H 中定义一个非线性映射 $\varphi(\cdot)$ 和从 H 到 H_k 三个线性变换 A_k ($k=1, 2, 3$), 从而对光谱维和空间维的交互信息进行融合, 得到如下基于交互信息的复合核函数表达式。

$$\phi(x_i) = \{A_1 \varphi(x_i^s), A_2 \varphi(x_i^w), A_3 (\varphi(x_i^s) + \varphi(x_i^w))\} \quad (8)$$

$$K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle = \varphi(x_i^s)^T R_1 \varphi(x_j^s) +$$

$$\varphi(x_i^w)^T R_2 \varphi(x_j^w) + \varphi(x_i^s)^T R_3 \varphi(x_j^w) + \varphi(x_i^w)^T R_3 \varphi(x_j^s) \quad (9)$$

$$R_1 = A_1^T A_1 + A_3^T A_3, R_2 = A_2^T A_2 + A_3^T A_3, R_3 = A_3^T A_3 \quad (10)$$

写成复合核函数的形式为

$$\begin{aligned} K(x_i, x_j) &= K_s(x_i^s, x_j^s) + K_w(x_i^w, x_j^w) + \\ & K_{sw}(x_i^s, x_j^w) + K_{ws}(x_i^w, x_j^s) \end{aligned} \quad (11)$$

这里的复合核函数需要光谱维信息和空间维信息同维度才能使用, 即 $N_w = N_s$ 。根据选取的非线性映射函数 $\phi(\cdot)$ 的

不同, 一般有以下 3 种常见的核即线性核、多项式核及径向基函数核。

$$\begin{cases} K(x_i, x_j) = \langle x_i, x_j \rangle \\ K(x_i, x_j) = (\langle x_i, x_j \rangle + 1)^d \\ K(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / 2\sigma^2) \end{cases} \quad (12)$$

SVM 采用这些复合核函数以后, 可以明显提高高光谱遥感图像的地物分类精度^[10]。但是, 这类方法并不能直接应用到高光谱显微图像的木材树种分类识别中。这是因为, 在遥感影像地物分类中, 基本任务是将每个像素分类到合适的地物类别中(例如沙地、河流、耕地、林地、建筑物等类别), 它处理的基本数据单元是像素。在这种情况下, 单个像素的光谱维特征是各个波长下的光谱反射率值, 而其空间维特征是该像素邻域窗口(例如 3×3 , 9×9 邻域)的均值和标准差。

1.5 模型构建与评价

为了将基于复合核函数的 SVM 应用到木材树种显微高光谱图像的分类识别任务中, 需要作以下的转换工作。首先, 这里处理的基本数据单元是手动选取的 ROI 区域, 具体步骤参见 1.3.1 节。其次, 对于 ROI 区域的光谱维特征提取, 采用了光谱降维和计算 ROI 区域平均光谱值的方法, 参见 1.3.1 节和 1.3.3 节。最后, 对于 ROI 区域的空间维特征提取, 采用了基于 GLCM 的纹理特征参数计算方法, 具体步骤参见 1.3.4 节。

为了准确评价分类模型的精度, 从 4 800 个样本中, 随机挑选 3 000 个样本作为训练集, 1 800 个样本作为测试集。分类器评价指标是由训练集分类准确率、测试集分类准确率、KAPPA 系数(KC)和测试时间四个指标为评价标准。

2 结果与讨论

2.1 光谱维分类

2.1.1 原始光谱

利用 Matlab2016b 软件对光谱进行建模和分类, 图 2 所示是 20 种木材树种样本的平均光谱曲线图。每种木材几乎都和其他树种类别有交叉, 光谱重叠严重, 而且部分光谱走势相似, 这对树种分类造成困难, 故必须对木材光谱数据做

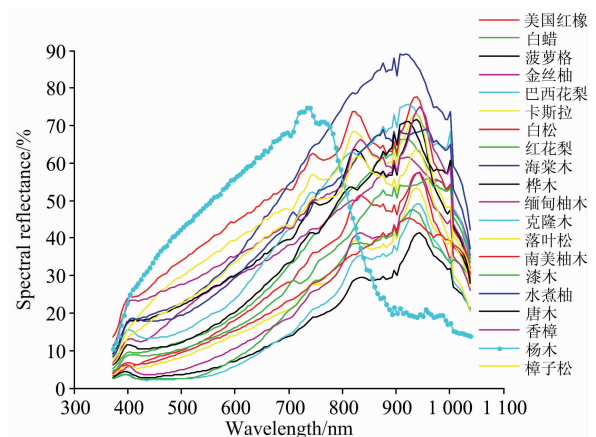


图 2 20 种木材树种样本的平均光谱曲线

Fig. 2 Wood spectral curves of 20 wood species

进一步处理。

2.1.2 采集波段建模分类

首先,对采集波段光谱数据进行建模分类,采用 SVM 和 BP 神经网络建立判别模型,分类准确率如表 1。由表中可知,BP 神经网络的测试时间少于 SVM 的测试时间,但分类准确率和 SVM 相比较低。SVM 方法训练集和测试集分类准确率都在 90% 以上,说明采集波段范围内利用 SVM 模型对木材树种分类是可行的。SVM 核函数采用了径向基核函数,最终利用交叉验证法和网格搜索法确定惩罚因子 C 和径向基核函数中的参数 σ 的值。

表 1 采集波段光谱建模分类对比
Table 1 Spectral classification comparisons for overall spectral band

Model	Training set/%	Testing set/%	KC	Testing time/s
SVM	98.800 0	92.222 2	0.918 1	0.405 2
BP	90.330 0	85.780 0	0.850 3	0.013 6

2.1.3 采集波段预处理建模分类

采用 3 种预处理方式对原始光谱进行预处理,然后基于各种预处理光谱建立 SVM 模型。从表 2 中可以看出,通过不同预处理,分类模型的精度得到了不同程度的提升。采用 SNV 预处理的训练集和测试集分类准确率最高,精度达到 98.766 7% 和 93.000 0%,而且所需测试时间也最短,因此后续建模都采用这种方式。

表 2 不同预处理下采集波段光谱 SVM 模型分类结果
Table 2 Spectral classification comparisons for overall spectral band and different preprocessing schemes

Pretreatment	SVM			
	Training set/%	Testing set/%	KC	Testing time/s
NONE	98.800 0	92.222 2	0.918 1	0.405 2
SNV	98.766 7	93.000 0	0.926 3	0.356 5
MSC	67.866 7	52.500 0	0.500 0	1.051 3
SG	98.900 0	92.277 8	0.918 7	0.379 0

2.1.4 特征波长选取及建模分类

连续投影算法(SPA)是由 Araujo 等在以多元校正的背景下提出^[17],通过对向量的投影分析,确定投影向量最大的为待选波长,再基于校正模型选择最终特征波长。本实验在 Matlab2016b 软件中对经过 SNV 预处理后的样本进行特征波长挑选,根据 RMSE 最小原则选取特征波长的集合。如图 3 所示,波长数目为 19 时 RMSE=0.456 56 接近最小值。因此我们挑选 19 个波长,分别是 372, 382, 387, 392, 397, 412, 417, 432, 457, 580, 637, 699, 725, 762, 778, 800, 816, 961 和 1 000 nm。

竞争性自适应重加权算法(CARS)^[18]是效仿达尔文进化论中“适者生存”的原则提出的变量优选方法。对采样次数反复迭代并寻找每次采样的 RMSECV 最小值,所对应的变量

视为优选出来的变量子集。图 4(a)表示选取变量数与运行次数的关系,由图可知呈递减趋势。图 4(b)表示 RMSECV 的变化趋势,当 RMSECV 变大表示剔除了有效信息;当 RMSECV 变小表示剔除了无效信息。图 4(c)表示当运行次数为 30 时(中位线的位置),RMSECV 最小。反复进行参数选取调试,筛选中,蒙特卡洛采样次数设为 100 次。最终 CARS 选择的特征波长分别是 392, 397, 427, 432, 437, 447, 452, 462, 467, 549, 554, 564, 569, 590, 595, 605, 611, 616, 637, 642, 668, 673, 678, 683, 710, 741, 773, 842, 869, 885, 891, 902, 907, 912, 934, 950, 956 和 989 nm 波长。表 3 给出了采用这些特征波长进行 SVM 分类识别的结果。综合对比两种特征波长选择方法,CARS 的特征波长选择效果和运行速度较好一些。

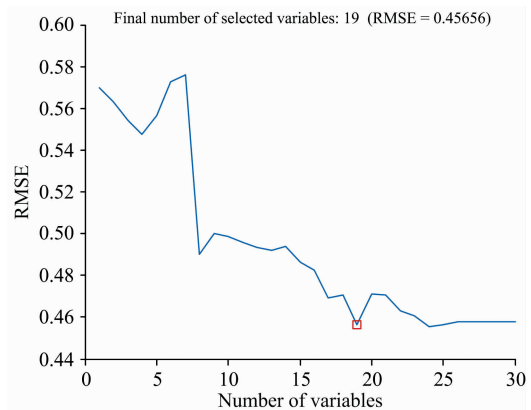


图 3 SPA 波长数量与均方根误差关系图

Fig. 3 The relationship between wavelength number and RMSE

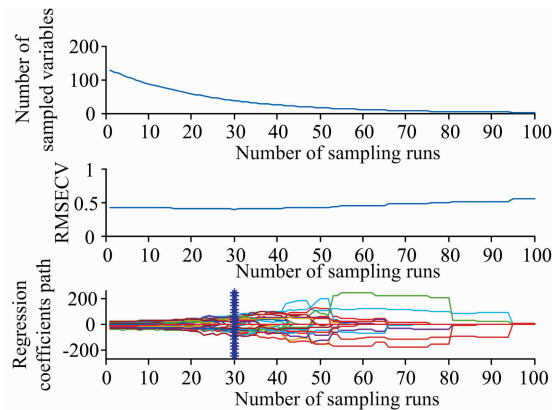


图 4 CARS 方法波长筛选过程

(a): 选取变量数与运行次数的关系; (b): RMSECV 的变化走势图; (c): RMSECV 的最小值选取

Fig. 4 Spectral wavelength selection procedure with CARS scheme

(a): Relationship between wavelength number and running times; (b): Varying trend chart for RMSECV; (c): Minimal value selection for RMSECV

2.2 空间维分类

高光谱数据可表示为 3D 立方体,不仅包含丰富的光谱

表 3 采用特征波长 SVM 模型分类结果

Table 3 Classification results of SVM using feature wavelength

Pretreatment	Methods	SVM			
		Training set/%	Testing set/%	KC	Testing time/s
SNV	SPA	98.300 0	91.944 4	0.915 2	0.135 4
	CARS	98.966 7	92.166 7	0.917 5	0.126 0

表 4 纹理建模分类比较

Table 4 Classification comparisons with wood texture features

Model	Training set/%	Testing set/%	KC	Testing time/s
SVM	99.800 0	60.333 0	0.577 2	0.128 8
BP	82.830 0	54.670 0	0.522 8	0.040 5

维信息,同时包含丰富的空间维/纹理信息。采用灰度共生矩阵计算纹理特征。先对高光谱图像进行主成分变换,选取前 3 个主成分图像 PC1—PC3,PC1 图像保留很多原始图像信息,而 PC2 和 PC3 的木材管孔射线等特征显示不清晰。因此我们采用 PC1 对 GLCM 的纹理特征进行提取,分别在 0, 45, 90 和 135 四个方向上的能量、熵、惯性矩、相关性总共 16 个特征参数输入到 SVM 和 BP 神经网络中。SVM 的训练集和测试集分类精度最高达到 99.800 0% 和 60.333 0%,但测试时间却稍高于 BP 神经网络。为了能准确对木材树种分类,采用 SVM 对木材高光谱图像纹理进行建模分类。

2.3 光谱维空间维特征融合及 SVM 复合核函数建模分类

将木材高光谱成像样本的 38 个特征波长下光谱值与 16

个纹理特征参数进行融合,融合前对光谱和纹理特征做归一化处理,融合后的特征参数作为输入数据进行复合核函数 SVM 建模分类,识别结果如表 5。

表 5 光谱和纹理信息融合分类结果

Table 5 Wood spectral-spatial fused classification comparisons

Methods	Composite kernels	SVM			
		Training set/%	Testing set/%	KC	Testing time/s
CARS+纹理	复合核 1	100.000 0	93.333 3	0.929 8	0.180 0
CARS+纹理	复合核 2	95.522 2	94.166 7	0.938 6	0.254 7
CARS+纹理	复合核 3	93.766 7	92.610 0	0.922 2	0.260 2
PCA+纹理	复合核 4	89.466 7	82.160 0	0.812 3	0.430 6

从表 5 可知采用四种复合核函数后 SVM 建模分类正确率有所上升,这是由于综合考虑了光谱维和空间维的分类特征。第四个复合核函数分类精度较低,这是因为第四核函数要求光谱与纹理同维限制,为满足同维要求,最终我们采用 PCA 方法,因为通过对比 CARS, SPA 和 PCA 这 3 种降维方法选取 16 维光谱信息,PCA 方法对本数据集的处理效率较高,精度较好。第二个复合核函数训练集和测试集分类效果最好可达到 95.522 2% 和 94.166 7%。综合分析,本实验中 CARS 能够较好地用于可见光/近红外高光谱的选择,该方法有效去除光谱中的无信息变量同时也能对共线性变量有效的剔除,所以 CARS+纹理模型优于其他模型,同时也说明在高光谱图像上融合光谱和纹理特征对木材树种进行识别是可行的。整体实验过程的流程图参见图 5。

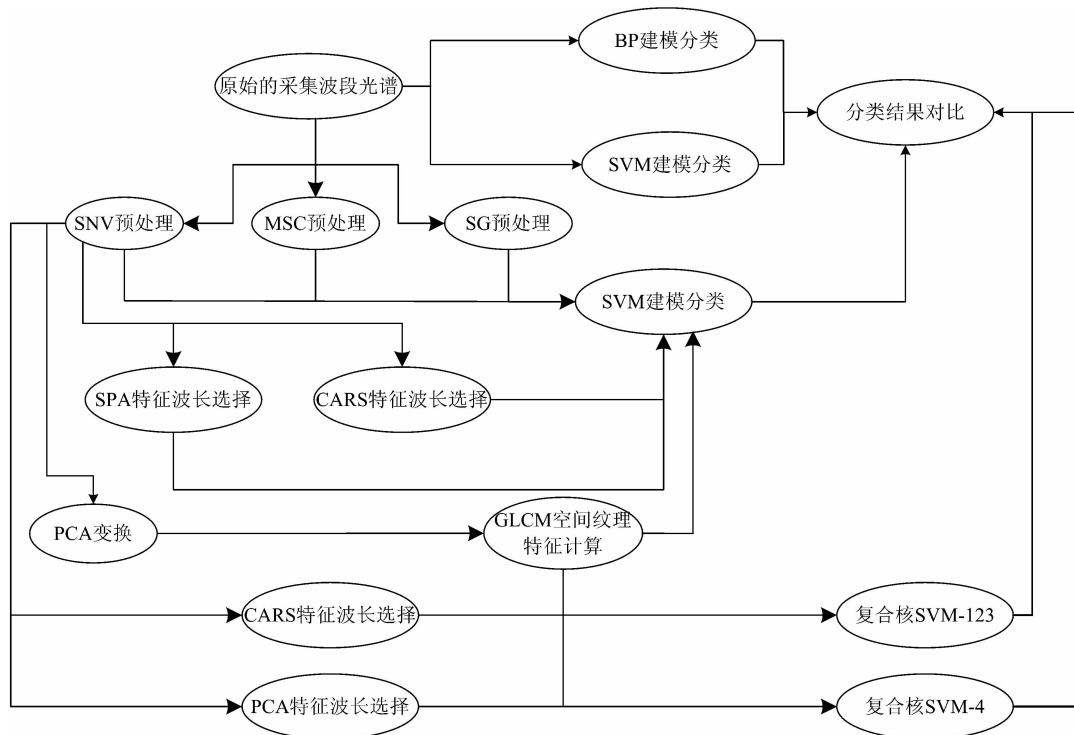


图 5 实验过程的流程图

Fig. 5 Experimental flow graph

3 结 论

高光谱成像技术具有图谱合一的特点, 将其应用在木材树种分类识别中。使用 SVM 的 4 种复合核函数, 构造了融合光谱与纹理特征的 SVM 分类器。对比了不同的光谱预处理方法和特征波长选择算法的实际效果, 验证了使用复合核函数 SVM 能够在一定程度上提高木材树种分类精度。特别是采用第二个复合核函数的 SVM 训练集分类准确率达到

95.522 2%, 测试集分类准确率达到 94.166 7%, KC 为 0.938 6, 测试时间为 0.254 7 s。因此, 本方法相比单独使用光谱特征或纹理特征进行的木材树种分类准确率有一定程度的提升, 这种情况在遥感领域的高光谱成像地物分类识别中也有类似的体现。因此, 如果采集的不同树种木材样本图像的纹理特征和光谱特征具有互补性的特点, 采用复合核函数 SVM 就能够有效的融合这两类特征并且提高木材树种的分类精度。

References

- [1] Yusof R, Khalid M, Khairuddin A S M. *Machine Vision and Applications*, 2013, 24: 1589.
- [2] Ibrahim I, Khairuddin A S M, Talip M S A, et al. *Wood Science and Technology*, 2017, 51(3): 431.
- [3] Piuri V, Scotti F. *IEEE Trans SMC-Part C*, 2010, 40(3): 358.
- [4] Yuan W, Shuai S, Nan Z, et al. *Bioresources*, 2019, 14(1): 1033.
- [5] Yusof R, Khalid M, Khairuddin A S M. *Computers and Electronics in Agriculture*, 2013, 93: 68.
- [6] LIU Ping, MA Mei-hu(刘 平, 马美湖). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2018, 38(1): 246.
- [7] WANG Bin, XUE Jian-xin, ZHANG Shu-juan(王 斌, 薛建新, 张淑娟). *Transactions of the Chinese Society for Agricultural Machinery(农业机械学报)*, 2013, 44(s1): 205.
- [8] SUN Jun, JIN Xia-ming, MAO Han-ping, et al(孙 俊, 金夏明, 毛罕平, 等). *Transactions of the Chinese Society for Agricultural Engineering(农业工程学报)*, 2014, 30(21): 301.
- [9] DENG Xiao-qin, ZHU Qi-bing, HUANG Min(邓小琴, 朱启兵, 黄 敏). *Laser & Optoelectronics Progress(激光与光电子学进展)*, 2015, 52(2): 122.
- [10] Fauvel M, Tarabalka Y, Benediktsson J A, et al. *Proceedings of the IEEE*, 2013, 101(3): 652.
- [11] Dong K X, Tao L S, Benediktsson J A. *IEEE Transactions on Geoscience & Remote Sensing*, 2014, 52(5): 2666.
- [12] Hong L, Song Y, Chen C L P. *IEEE Transactions on Geoscience & Remote Sensing*, 2017, 55(9): 5302.
- [13] Cui X, Wang Q Q, Zhao Y, et al. *Applied Physics B: Lasers and Optics*, 2019, 125(4): 56.
- [14] Dawson A B, Adedipe O E. *Wood Science and Technology*, 2012, 46(6): 1193.
- [15] Cortes C, Vapnik V. *Machine Learning*, 1995, 20(3): 273.
- [16] Valls G C, Chova L G, Mari J M, et al. *IEEE Geoscience and Remote Sensing Letters*, 2006, 3(1): 93.
- [17] Pontes M J C, Galvo R K H, Araujo M C U, et al. *Chemometrics and Intelligent Laboratory Systems*, 2005, 78(1-2): 11.
- [18] Li H, Liang Y, Xu Q, et al. *Analytica Chimica Acta*, 2009, 648(1): 77.

Wood Species Recognition with Microscopic Hyper-Spectral Imaging and Composite Kernel SVM

ZHAO Peng*, TANG Yan-hui, LI Zhen-yu

College of Information and Computer Engineering, Northeast Forestry University, Harbin 150040, China

Abstract In this paper, a stereomicroscopic hyper-spectral imaging scheme is used for wood species recognition. The SOC710VP hyper-spectral imaging system is used to pick up the wood images in visible and near-infrared spectral band (i. e., 372.53~1 038.57 nm). First, the ENVI software is used to pick up the mean spectra of wood sample's Region of Interest (ROI). The Successive Projection Algorithm (SPA) and Competitive Adaptive Reweighted Sampling algorithm (CARS) are used for spectral dimension reduction. Second, a Support Vector Machine (SVM) is used to classify the wood samples in full spectral band and in feature wavelengths. Third, in the spatial dimension the 1st principal component image (PC1) is used to compute the wood texture features based on Gray Level Co-occurrence Matrix (GLCM). In the 4 directions of 0°, 45°, 90°, 135° the 16 feature parameters such as energy, entropy, inertia moment and so on are calculated and are put into SVM for wood species recognition. Lastly, the 4 composite kernels SVM are used to fuse the spatial-spectral features for wood species recognition. Experiments on 20 wood species classification indicate that CARS is a better choice in view of the feature wavelength selection

and running speed and the classification accuracy for testing set reaches to 92.1667% if the ordinary SVM is used for wood spectral classification. If the wood texture features based on GLCM are used, the classification accuracy for testing set reaches to 60.3330% if the ordinary SVM is used. When the wood spectral and texture features are fused for classifications, the composite kernel SVM has the best classification accuracy. Especially, the classification accuracy of the 2nd composite kernel SVM is the highest with 94.1667% for testing set and a processing speed of 0.2547 s. Moreover, the classification accuracy of the 1st or 3rd composite kernel SVM reaches to 93.3333% or 92.6100% with a running speed of 0.1800 or 0.2602 s. Therefore, wood species classification accuracy can be improved by use of hyper-spectral imaging and composite kernel SVM, which may be applied in the practical wood species classification system.

Keywords Wood species classification; Hyper-spectral imaging; Composite kernel SVM; Feature fusion

(Received Mar. 18, 2019; accepted Jul. 24, 2019)

* Corresponding author

关于《光谱学与光谱分析》调整审稿费收费标准的通知

尊敬的《光谱学与光谱分析》广大作者、读者：我刊自 2018 年 7 月 1 日以后登记的稿件向投稿作者收取审稿费 200 元/篇，在您投稿之前，为免受经济损失，请您必须考虑：

1. 没有创新的一般性稿件，请您不要投稿。
2. 没有国家级基金资助的稿件，请您不要投稿。
3. 不是光谱专业的稿件，请您不要投稿。
4. 与其他文章重合率超过 10% 的稿件，请您不要投稿。

所投稿件经初审通过后，作者会收到缴纳审稿费的通知。请作者及时从我刊网站(<http://www.gpxygpfx.com>)查询稿件是否处于交审稿费状态，在收到通知后，请及时缴纳审稿费；如在 10 天之内没有收到您的审稿费，被视为自动放弃，我刊不再受理。交费后我刊开据增值税电子普通发票，并传至作者提供的电子邮箱，作者可自行打印。

联系电话：010-62181070，62182998

电子邮箱：chngpaxygpfx@vip.sina.com

感谢您多年来对《光谱学与光谱分析》的支持和厚爱！

《光谱学与光谱分析》期刊社

2018 年 6 月 30 日