

万广, 陈忠辉, 方洪波, 等. 基于特征融合的随机森林模型茶鲜叶分类 [J]. 华南农业大学学报, 2021, 42(4): 125-132.  
WAN Guang, CHEN Zhonghui, FANG Hongbo, et al. Classification of fresh tea leaf based on random forest model by feature fusion[J]. Journal of South China Agricultural University, 2021, 42(4): 125-132.

# 基于特征融合的随机森林模型茶鲜叶分类

万 广<sup>1</sup>, 陈忠辉<sup>1</sup>, 方洪波<sup>1</sup>, 闫建伟<sup>2</sup>, 张文勇<sup>3</sup>, 谢本亮<sup>1</sup>

(1 贵州大学 大数据与信息工程学院, 贵州 贵阳 550025; 2 贵州大学 机械工程学院, 贵州 贵阳 550025;  
3 贵州大学 计算机科学与技术学院, 贵州 贵阳 550025)

**摘要:**【目的】解决机采茶鲜叶中混有不同等级的茶叶,且混杂度高、物理特征分类精确度低的问题。【方法】利用随机森林分类模型,提出一种基于颜色和边缘特征融合的方法。试验采集 3 种不同等级的茶鲜叶,对原始图像进行裁剪、尺寸归一化和去噪等处理,再进行颜色特征和边缘特征提取。通过参数的修改和测试,构建最优的随机森林分类模型,并且同  $K$  最近邻、SVM 分类器进行对比试验。【结果】特征融合之后随机森林模型的分类准确率达到 99.45%,比单一颜色特征和边缘特征的分类准确率分别高 7.14 和 9.34 个百分点;比  $K$  最近邻和 SVM 分类器准确率分别高 15.38 和 5.49 个百分点。【结论】所建立的方法能够对茶鲜叶单芽、一芽一叶、一芽二叶进行精确的分类。

**关键词:** 茶鲜叶分类; 随机森林模型; 颜色特征; 边缘特征; 机器学习  
**中图分类号:** S571.1; TP181      **文献标志码:** A      **文章编号:** 1001-411X(2021)04-0125-08

## Classification of fresh tea leaf based on random forest model by feature fusion

WAN Guang<sup>1</sup>, CHEN Zhonghui<sup>1</sup>, FANG Hongbo<sup>1</sup>, YAN Jianwei<sup>2</sup>, ZHANG Wenyong<sup>3</sup>, XIE Benliang<sup>1</sup>  
(1 School of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China; 2 School of Mechanical Engineering, Guizhou University, Guiyang 550025, China; 3 School of Computer Science and Technology, Guizhou University, Guiyang 550025, China)

**Abstract:** 【Objective】To solve the problems of the machine-picked fresh tea leaves mixing with different grades of tea leaves, high mixing degree and low classification accuracy of physical characteristics. 【Method】Using the random forest classification model, a method based on the fusion of color and edge feature was proposed. We collected three different grades of fresh tea leaves, and processed the original images with cropping, size normalization and denoising, and then extracted the color features and edge features. Through parameter modification and testing, the optimal random forest classification model was constructed, and the comparison experiment was performed with the  $K$ -nearest neighbor and SVM classifier. 【Result】After feature fusion, the classification accuracy of random forest model reached 99.45%, which was 7.14 and 9.34 percentage points higher than those of single color feature and single edge feature, 15.38 and 5.49 percentage points higher than those of  $K$ -nearest neighbor model and SVM classifier respectively.

收稿日期: 2020-12-04      网络首发时间: 2021-04-30 14:41:50  
网络首发地址: <https://kns.cnki.net/kcms/detail/44.1110.S.20210430.1332.006.html>  
作者简介: 万 广 (1995—), 男, 硕士研究生, E-mail: [wanguang19@163.com](mailto:wanguang19@163.com); 通信作者: 谢本亮 (1978—), 男, 副教授, 博士, E-mail: [blxie@gzu.edu.cn](mailto:blxie@gzu.edu.cn)  
基金项目: 国家自然科学基金 (61562009); 贵州大学人才引进科研项目 (贵大人基合字 (2015)29 号); 半导体功率器件教育部工程研究中心开放基金 (ERCMEKFJJ2019-(06)); 贵州科技计划项目 (黔科合成果 [2019]4279 号, 黔科合平台人才 [2019]5616 号)

**【Conclusion】** The established method can accurately separate single bud, one bud and one leaf, and one bud and two leaves of fresh tea leaves.

**Key words:** fresh tea leaf classification; random forest model; color feature; edge feature; machine learning

作为茶叶的原产地之一，中国有着历史悠久的茶文化底蕴<sup>[1]</sup>，茶鲜叶质量的优劣基本决定着茶叶品质的高低。随着劳动力的短缺以及劳动成本的提升，茶产业机械化、工业化进程的不断推进，茶鲜叶的机械化采摘技术和装备日趋成熟。但机采茶鲜叶混杂度高，包含有各种等级的茶叶，因此在制茶工艺中，分选成为了一道费工、费时且关键的工序。早期根据茶鲜叶的物理特性，研制出了滚筛、圆筛、风选等分选方法<sup>[2]</sup>，虽然可以去除茶鲜叶中的杂物(如残叶、叶梗等)，但难以对茶鲜叶进行精确的等级划分。

近年来，随着机器视觉和机器学习技术的发展，茶叶的分类技术进入智能化时代。陈全胜等<sup>[3]</sup>利用颜色特征建立支持向量机(Support vector machines, SVM)分类模型，通过色泽来分辨茶叶品质的优劣，模型的平均识别率达到 95% 左右；吴正敏等<sup>[4]</sup>以大红袍为例，通过提取夏秋季节茶叶各种形态特征参数，判断特征权重进行特征选择，并且设置不同的权重比来完成叶和梗的分离，最终识别准确率为 93.8%；高震宇等<sup>[5]</sup>通过建立卷积神经网络，通过局部连接和权值共享等方法提高了网络的训练性能，最终识别的准确率不低于 90%。余洪<sup>[6]</sup>通过 RGB(Red, green, blue) 和 HIS(Hue, intensity, saturation) 颜色模型提取颜色特征，通过统计矩和

灰度共生矩提取纹理特征，共提取到 12 个颜色特征和 22 个纹理特征，建立基于主成分分析、遗传算法和 BP 神经网络的茶叶品质分级模型，识别准确率达到 92.5%。

本研究应用随机森林分类模型，以 3 种不同品质的茶鲜叶图像作为依据，分别提取茶鲜叶样本图像的颜色特征和边缘特征，并且将 2 种特征同时输入到分类模型中，以期实现对 3 种不同品质茶鲜叶的精确识别和分类。

1 材料与方法

1.1 图像采集

试验研究对象是绿茶中的都匀毛尖，于春季在贵州省黔西南州采摘，共采集到茶鲜叶样本叶片 906 个，其中包括单芽 299 个、一芽一叶 302 个、一芽二叶 305 个。拍摄环境为实验室，拍照设备为手机荣耀 20 (后置四摄 4 800 万+1 600 万+200 万+200 万像素)；为了突出茶鲜叶的颜色特征和边缘特征，拍照背景选择 A4 白纸以提高对比度；并用普通白炽灯照射，减弱周围其他光源的影响；拍摄方式为垂直俯拍，并用支架固定手机拍摄位置，手机与样本之间的距离为 30 cm 左右。采集的部分样本图像如图 1 所示。

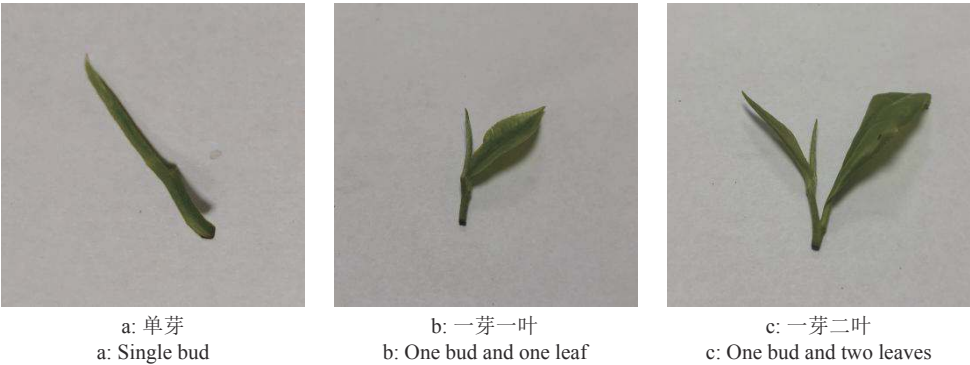


图 1 不同等级茶鲜叶样本图像

Fig. 1 Representative images of different grades of fresh tea leaves

1.2 图像预处理

图像采集过程中避免不了外界因素的干扰，为了便于后续特征提取的方便，需要对获得的图像数据集进行预处理。其中包括区域裁剪、尺寸归一化和噪声去除<sup>[7]</sup>。具体操作流程如图 2 所示。

1.2.1 区域裁剪和尺寸归一化 由于拍摄条件的限制，茶鲜叶在图像中的显示位置会有差异，RGB 图像也不能保持一致。因此需要对茶鲜叶的图像进行区域裁剪。区域裁剪的目的是将研究以外的区域去除，保留图像中心区域的茶鲜叶图像作为感兴趣



图2 图像预处理流程图

Fig. 2 Flow chart of image preprocessing

区域 (Region of interest, ROI); 对区域裁剪处理后的图像进行尺寸归一化, 使各个指标处于同一个数量级, 处理后的图像分辨率为  $256 \times 256$  像素, 如图 2b 所示。

1.2.2 噪声去除 图像中存在的噪声会对茶鲜叶叶片的特征提取产生不利的影响, 因此需要消除图像中的噪声干扰。采用中值滤波算法对图像进行预处理, 在去除噪声的同时保留完整的叶片信息, 并且能够很好地保护图像的边缘信息, 使图像中茶鲜叶的边缘更加平滑, 便于后续边缘特征的提取。得到的滤波后图像如图 2c 所示。

1.3 图像特征提取

1.3.1 颜色特征提取 颜色作为一种全局特征<sup>[8-10]</sup>, 是图像中最简单直接的一种特征, 直方图作为一种简单有效的基于统计特性的特征描述, 能描述图像中颜色的全局分布。本文将 RGB 颜色空间和 HSV (Hue, saturation, value) 颜色空间<sup>[11]</sup> 配合使用来区分不同等级的茶鲜叶颜色特征。HSV 是根据颜色的直观特性创建的一种颜色空间, 更接近实际人类的视觉特征, 手机拍摄获得的图像为 RGB 格式, 需要转化为 HSV 格式, RGB 坐标系向 HSV 空间转化的公式为:

$$H = \arccos \frac{2R - G - B}{2 \sqrt{(R - G)^2 + (R - B)(G - B)}}, \quad (1)$$

$$S = 1 - \frac{3}{R + G + B} (R, G, B)_{\min}, \quad (2)$$

$$V = \frac{(R + G + B)}{3}, \quad (3)$$

式中:  $(R, G, B)_{\min}$  表示三者中的最小值;  $H$  代表色调, 用角度度量, 取值范围为  $[0^\circ, 360^\circ]$ , 从红色开始按逆时针方向计算, 红色为  $0^\circ$ , 绿色为  $120^\circ$ , 蓝色为  $240^\circ$ ;  $S$  代表饱和度, 表示颜色接近光谱色的程度;  $V$  代表明度, 表示颜色明亮的程度, 通常取值为  $0\%$  (黑)~ $100\%$  (白);  $R$ 、 $G$ 、 $B$  分别表示红、绿、蓝通道。

对经过中值滤波处理后的图像进行颜色特征的提取, 3 种不同等级的茶鲜叶的 RGB 空间的直方图和 HSV 空间的  $H$ 、 $S$ 、 $V$  通道的直方图像素分布情况如图 3~图 6 所示。

对多个茶鲜叶的直方图的像素进行统计得出, 在 RGB 空间中单芽图像的像素主要分布区间为  $[134, 184]$ , 一芽一叶图像的像素主要分布区间为  $[43, 96]$  和  $[147, 186]$ , 一芽二叶图像的像素主要分

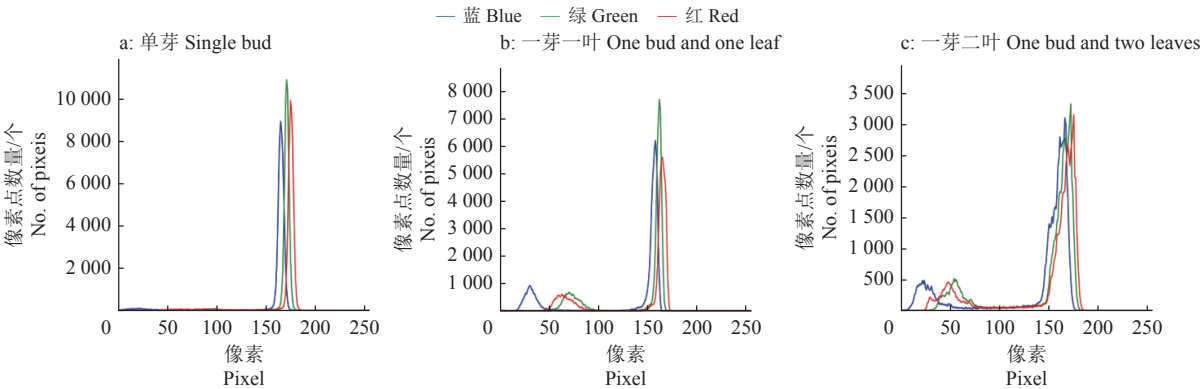


图3 3种等级茶鲜叶的颜色空间直方图

Fig. 3 Color space histogram of three grades of fresh tea leaves

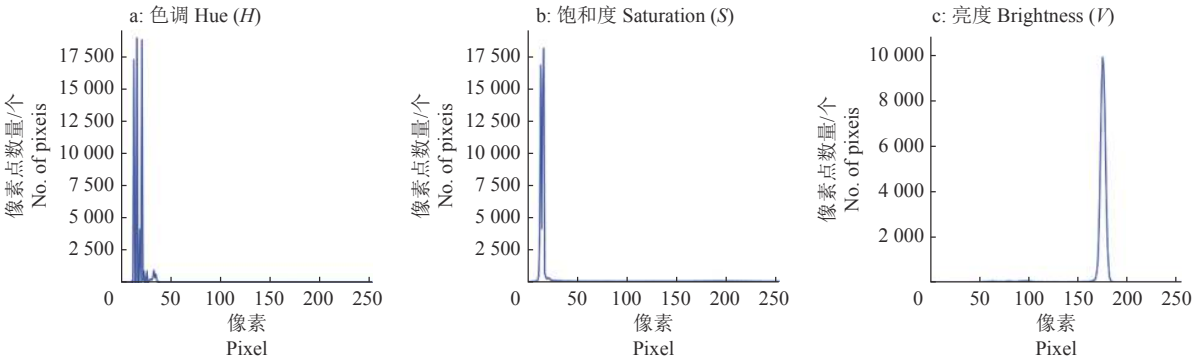


图 4 单芽茶鲜叶的  $H$ 、 $S$ 、 $V$ 通道直方图

Fig. 4  $H, S, V$  channel histogram of fresh tea leaf with single bud

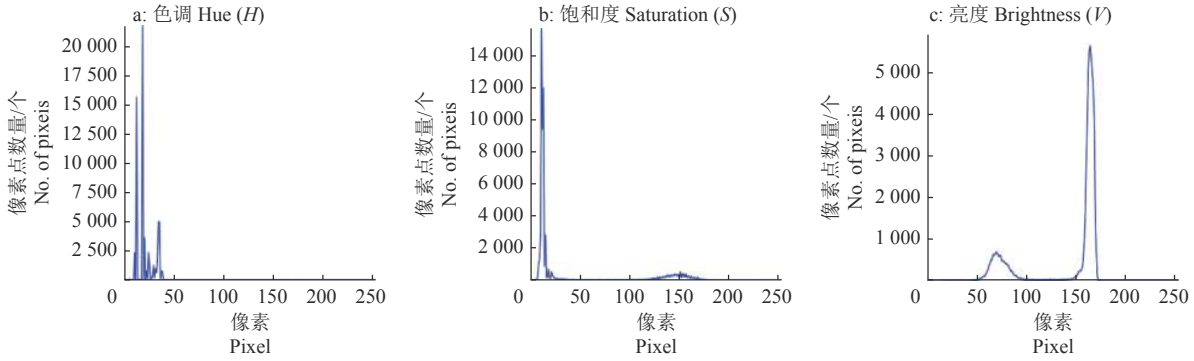


图 5 一芽一叶茶鲜叶的  $H$ 、 $S$ 、 $V$ 通道直方图

Fig. 5  $H, S, V$  channel histogram of fresh tea leaf with one bud and one leaf

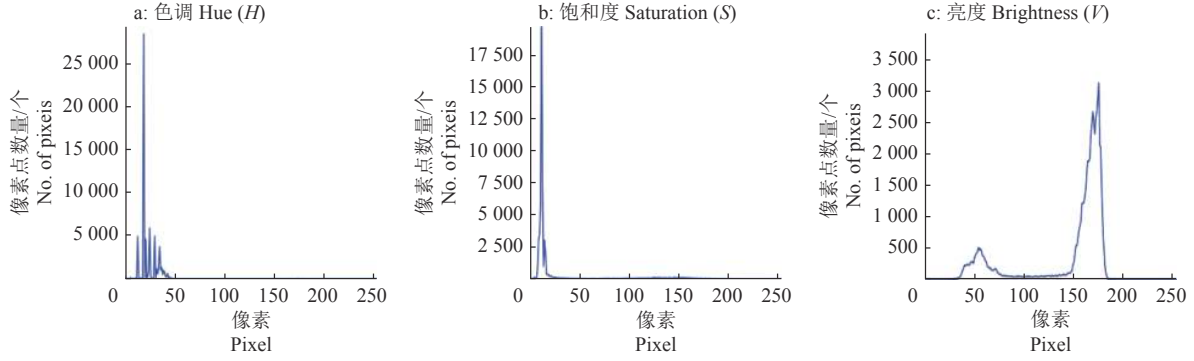


图 6 一芽二叶茶鲜叶的  $H$ 、 $S$ 、 $V$ 通道直方图

Fig. 6  $H, S, V$  channel histogram of fresh tea leaf with one bud and two leaves

布区间为 [41, 197]。在 HSV 空间中 3 种等级的茶鲜叶图像  $H$  通道像素主要分布区间为 [0, 70],  $S$  通道像素主要分布区间为 [0, 200],  $V$  通道像素主要分布区间为 [40, 200]。为了加强不同等级茶鲜叶像素之间的差异, 分别对  $H, S, V$  三通道像素进行非等间隔的量化, 其中色调  $H$  空间分为 8 份, 饱和度  $S$  和亮度  $V$  空间各划分为 3 份, 以提高分类器的构建效率和识别的准确率。

1.3.2 边缘特征提取 图像的边缘特征主要针对物体的外边界<sup>[12-14]</sup>。Canny 算法是一种多级边缘检测算法, 在实际操作中, Canny 算法使用一个低阈值

和一个高阈值来确定哪些点属于轮廓, 低阈值主要包括所有属于明显图像轮廓的边缘像素; 高阈值是定义所有重要轮廓的边缘, 最后组合低阈值和高阈值两幅边缘图生成最优的轮廓图。图像的边缘可以指向不同的方向, 因此经典 Canny 算法用 4 个梯度算子来计算水平、垂直和对角线方向的梯度<sup>[15]</sup>。本文采用梯度算子中的 Sobel 算子计算水平和垂直方向的差分  $G_x$  和  $G_y$ , 由这 2 个条件便可计算梯度模和方向, 如式 (4)、式 (5) 所示。

$$G = \sqrt{G_x^2 + G_y^2}, \tag{4}$$

$$\theta = \arctan(G_y/G_x), \tag{5}$$

式中, 角度 $\theta$ 范围为  $[-\pi, \pi]$ ,  $G_x$ 、 $G_y$  是 1 对卷积阵列, 梯度方向近似到 4 个可能的角度 ( $0^\circ$ 、 $45^\circ$ 、 $90^\circ$ 、 $135^\circ$ )。经过 Canny 算法处理后的边缘特征如图 7 所示。

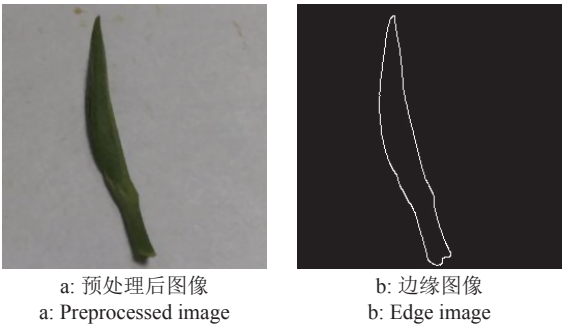


图 7 边缘特征图像  
Fig. 7 Edge feature image

**1.4 随机森林算法**

**1.4.1 算法原理** 随机森林 (Random forest, RF)<sup>[16-19]</sup> 属于并行集成学习中 Bagging(Bootstrap AGGREGatING) 算法的一种扩展变体, 是在以决策树为基学习器构建 Bagging 集成的基础上, 进一步在决策树的训练过程中引入随机属性选择。它将多种弱分类器集成, 形成新的分类器模型, 决策树决定了输出分类, 无需特征筛选也能得到较高的正确率, 对特征具有较好的鲁棒性。与传统的分类器相比, 随机森林需要很少的参数调整, 并且在准确率方面有很大的优势。随机森林拥有其独特的特点, 能够处理很高维度 (特征较多) 的数据, 不需要降维, 训练速度较快, 容易做成并行化方法, 由于随机性的引入, 使得其很少出现过拟合的现象, 随机森林算法流程如图 8 所示。

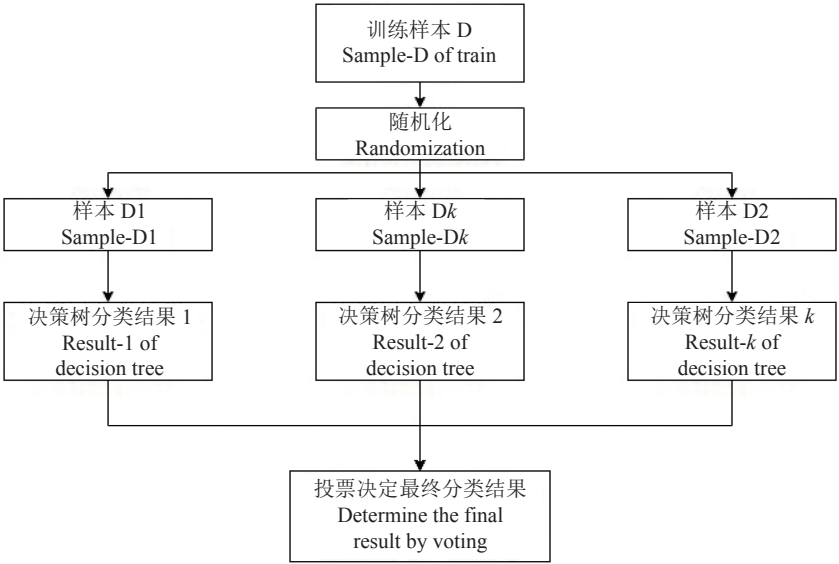


图 8 随机森林流程图  
Fig. 8 Flow chart of random forest

**1.4.2 随机森林的训练过程** 随机森林的模型训练是通过随机采样 (Bootstrap) 方法随机有放回地抽取  $k$  个样本, 并进行  $N_{\text{tree}}$  次采样, 生成  $N_{\text{tree}}$  个训练集, 分别训练  $N_{\text{tree}}$  个决策树模型的过程。对于单个决策树模型, 每次分裂时根据信息增益和信息熵选择最好的特征进行分裂。对没有抽中的样本作为袋外数据 (Out of bag, OOB)。

针对分类问题, 随机森林的输出采用多数投票法。利用随机森林模型对测试集样本进行分类判别, 过程就是让每棵决策树进行投票抉择, 最终输出最多的那个类别作为分类结果, 输出判别式如式 (6) 所示。

$$H(x) = \arg \max_Y \sum_{i=1}^k I(h_i(x) = Y), \quad (6)$$

式中:  $\arg \max$  表示函数取得最大值时的参数值;  $H(x)$  表示随机森林的最终分类结果;  $h_i(x)$  表示单一决策树模型分类结果;  $I()$  为示性函数 (所谓示性函数是指一个函数使得当集合内有此数时值为 1, 当集合内无此数时值为 0)<sup>[18]</sup>;  $Y$  表示输出变量 (或称目标变量)。

2 结果与分析

2.1 试验平台和参数

本试验在 Pycharm 环境下进行操作, 所用电脑操作系统为 Win10(64 位), 运行内存 4 G, 处理器为酷睿 i5-6200u, 主频 2.30 GHz, 进行图像处理为 OpenCV2 库。通过 sklearn.ensemble 调用 Random-ForestClassifier 分类模块。随机森林中决策树的数目

( $N_{tree}$ ) 对最终分类结果有着及其重要的影响, Random-ForestClassifier 模块中通过  $n\_estimators$  进行调节。

多次试验证明, 当 $N_{tree}$ 取值较小时, 随机森林的分类误差会较大, 当 $N_{tree}$ 的值逐渐增大时, 随机森林分类的精确度会有明显的提升, 但最终会趋于稳定, 甚至会有所下降。但当 $N_{tree}$ 的数量越大时, 占用的内存与训练和预测的时间也会相应增加, 且边际效益是递减的, 所以要在可承受范围内尽可能地选取合适的数量。

为了进一步选择合适的决策树数量, 在固定其他参数不变、仅改变  $n\_estimators$  参数的情况下, 对茶鲜叶数据集进行多次分类试验, 观察分类精确度随着决策树数目的变化。

图 9 显示了分类准确率随着 $N_{tree}$ 变化而变化的曲线。考虑到内存和训练时间, 本试验的 $N_{tree}$ 选择为 35。

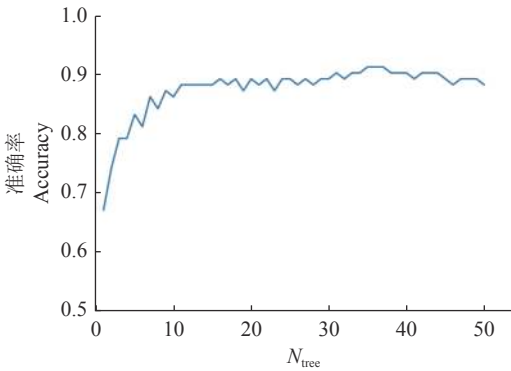


图 9 分类准确率随决策树数目 ( $N_{tree}$ ) 的变化  
Fig. 9 Classification accuracy varied with the number of decision trees ( $N_{tree}$ )

2.2 试验结果

试验时针对上述构建好的随机森林分类器模型, 将预处理之后得到的 906 张图像 (单芽图像 299 张, 一芽一叶图像 302 张, 一芽二叶图像 305 张) 按照 4:1 的比例划分后进行试验, 其中训练集 724 张, 测试集 182 张。识别准确率 (Accuracy) 是分类器性能判断的重要指标之一, 另外, 精确率 (Precision)、召回率 (Recall) 以及综合评价指标 (F1-score) 在最终分类阶段也作为常用的指标。4 种评价指标公式如式 (7)~式 (10) 所示。

准确率 =  $\frac{TP + TN}{TP + FP + FN + TN} \times 100\%$ , (7)

精确率 =  $\frac{TP}{TP + FP} \times 100\%$ , (8)

召回率 =  $\frac{TP}{TP + FN} \times 100\%$ , (9)

综合评价指标 =  $\frac{2TP}{2TP + FP + FN} \times 100\%$ , (10)

式中: TP(True positive) 为真阳性, 表示实际类别为正, 算法输出类别也为正; FP(False positive) 为假阳性, 表示实际类别为负, 样本输出类别为正; TN (False negative) 为假阴性, 表示实际类别为负, 算法输出类别也为负; FN(False negative) 为假阴性, 表示实际类别为正, 算法输出类别为负。

为了验证本文方法的有效性, 分别将 2 个单一特征和融合特征输入到随机森林分类模型中, 根据式 (7)~(10), 得到准确率、精确率、召回率和综合评价指标等结果数值, 分别列于表 1~表 3。

表 1 颜色特征分类结果  
Table 1 Color feature classification results %

| 茶鲜叶等级<br>Grade of fresh tea leaf | 精确率<br>Precision | 召回率<br>Recall | 综合评价指标<br>F1-score | 准确率<br>Accuracy |
|----------------------------------|------------------|---------------|--------------------|-----------------|
| 单芽 Single bud                    | 96.23            | 96.23         | 96.23              |                 |
| 一芽一叶 One bud and one leaf        | 92.06            | 87.88         | 89.92              |                 |
| 一芽二叶 One bud and two leaves      | 89.39            | 93.65         | 91.67              |                 |
| 平均值 Average                      | 92.56            | 92.59         | 92.61              | 92.31           |

表 2 边缘特征分类结果  
Table 2 Edge feature classification results %

| 茶鲜叶等级<br>Grade of fresh tea leaf | 精确率<br>Precision | 召回率<br>Recall | 综合评价指标<br>F1-score | 准确率<br>Accuracy |
|----------------------------------|------------------|---------------|--------------------|-----------------|
| 单芽 Single bud                    | 100              | 95.59         | 97.74              |                 |
| 一芽一叶 One bud and one leaf        | 81.36            | 87.27         | 84.21              |                 |
| 一芽二叶 One bud and two leaves      | 87.93            | 86.44         | 87.18              |                 |
| 平均值 Average                      | 89.96            | 89.77         | 89.71              | 90.11           |

表 3 融合特征分类结果

Table 3 Fusion feature classification results

%

| 茶鲜叶等级<br>Grade of fresh tea leaf | 精确率<br>Precision | 召回率<br>Recall | 综合评价指标<br>F1-score | 准确率<br>Accuracy |
|----------------------------------|------------------|---------------|--------------------|-----------------|
| 单芽 Single bud                    | 100              | 100           | 100                |                 |
| 一芽一叶 One bud and one leaf        | 98.21            | 100           | 99.10              |                 |
| 一芽二叶 One bud and two leaves      | 100              | 98.31         | 99.15              |                 |
| 平均值 Average                      | 99.40            | 99.44         | 99.42              | 99.45           |

由表 1 可知, 颜色特征的识别准确率为 92.31%, 对单芽的分类结果最好, 精确率、召回率和综合评价指标都为 96.23%, 主要原因在于单芽的直方图像素区间较小, 比较集中; 一芽一叶和一芽二叶的颜色特征综合评价指标为 90% 左右, 因为大多数一芽一叶的直方图像素区间为 2 个, 但 2 个区间的像素数量并不是绝对为 0, 所以一芽一叶和一芽二叶容易造成混淆。

由表 2 可知, 边缘特征的识别准确率为 90.11%, 比颜色特征低 2.2 个百分点, 主要原因也在于一芽一叶和一芽二叶之间的混淆, 例如: 测试集混淆矩阵中, 一芽一叶 55 张, 有 7 张识别为一芽二叶; 一芽二叶 59 张, 有 8 张识别为一芽一叶。

由表 3 可知, 颜色和边缘特征融合识别的准确率为 99.45%, 比颜色特征和边缘特征的识别准确率分别高 7.14, 9.34 个百分点, 该方法取 2 个特征识别时的交集, 避免了单一特征识别时的局限性, 提高了识别准确率, 证明了该方法的可行性。

2.3 不同分类模型结果对比

为了进一步说明本文方法的有效性, 将本文方法同传统的机器学习方法  $K$  最近邻 ( $K$ -nearest neighbor, KNN)<sup>[20]</sup> 和 SVM 分类器<sup>[21-22]</sup> 进行对照试验。其中  $K$  最近邻算法中  $K$  设置为 3, SVM 分类器中惩罚系数  $C$  设置为 10, 核函数系数  $\gamma$  为 0.000 1, 试验结果如表 4 所示。

表 4 不同模型的平均分类结果

Table 4 Average classification results of different models

%

| 模型<br>Model        | 精确率<br>Precision | 召回率<br>Recall | 综合评价指标<br>F1-score | 准确率<br>Accuracy |
|--------------------|------------------|---------------|--------------------|-----------------|
| $K$ 最近邻 KNN        | 83.05            | 83.00         | 83.02              | 84.07           |
| SVM                | 93.81            | 93.90         | 93.82              | 93.96           |
| 随机森林 Random forest | 99.40            | 99.44         | 99.42              | 99.45           |

由表 4 可以看出, 针对茶鲜叶等级的分类问题, 随机森林模型明显优于  $K$  最近邻和 SVM 分类器, 在准确率上分别高出 15.38 和 5.49 个百分点, 精确率、召回率和综合评价指标也有明显的优势。以上结果表明, 在茶鲜叶的分类中, 随机森林模型的分性能最优, SVM 次之,  $K$  最近邻效果最差。作为机器学习中最常用的一种算法, SVM 需要对参数进行不断的优化, 以提高分类的准确率和防止过拟合问题的出现, 相较于随机森林模型需要耗费更多的时间和精力。作为最简单的算法之一,  $K$  最近邻算法原理简单, 容易理解, 但需要对样本的特征进行量化, 才能获得较好的分类结果。

3 结论

本文以春季的都匀毛尖作为研究对象, 利用随机森林分类模型, 提出了一种颜色特征和边缘特征

融合的方法, 对 3 种不同等级的茶鲜叶进行识别和分类。根据准确率、精确率、召回率和综合评价指标进行判定。首先对茶鲜叶图像进行区域裁剪、尺寸归一化和噪声去除等预处理, 提取 RGB 空间的彩色直方图并转移到 HSV 空间中, 计算  $H$ 、 $S$ 、 $V$  三通道像素区间, 获得茶鲜叶图像的颜色特征。利用 Canny 边缘检测算法提取图像的边缘特征, 将 2 种特征融合作为茶鲜叶的识别特征。为了判断本文方法的有效性, 将单一特征和融合特征后的分类结果进行对照, 并且将随机森林模型同  $K$  最近邻和 SVM 分类器进行对照试验。试验结果表明, 特征融合随机森林模型的分类准确率、精确率、召回率和综合评价指标分别达 99.45%、99.40%、99.44% 和 99.42%, 明显高于单一特征的分类结果, 并且随机森林模型的分性能最优。该方法也能快速应用到其他类型的茶鲜叶分选, 为茶鲜叶分选智能化进程提供了一定的依据。

参考文献:

[1] 黄藩, 刘飞, 王云, 等. 计算机视觉技术在茶叶领域中的应用现状及展望[J]. 茶叶科学, 2019, 39(1): 81-87.

[2] 唐小林, 李文萃, 范起业. 机采茶鲜叶分类分级技术及相关设备研究进展[J]. 中国茶叶加工, 2015(2): 5-8.

[3] 陈全胜, 赵杰文, 蔡健荣, 等. 支持向量机在机器视觉识别茶叶中的应用研究[J]. 仪器仪表学报, 2006(12): 1704-1706.

[4] 吴正敏, 曹成茂, 王二锐, 等. 基于形态特征参数的茶叶精选方法[J]. 农业工程学报, 2019, 35(11): 315-321.

[5] 高震宇, 王安, 刘勇, 等. 基于卷积神经网络的鲜茶叶智能分选系统研究[J]. 农业机械学报, 2017, 48(7): 53-58.

[6] 余洪. 基于计算机视觉技术的茶叶品质分级研究[D]. 南昌: 江西农业大学, 2017.

[7] 金良海, 李德华. 基于噪声检测的图像去噪算法[J]. 模式识别与人工智能, 2008, 21(3): 298-302.

[8] 孙卫红, 黄志鹏, 梁曼, 等. 基于颜色特征和支持向量机的蚕茧分类方法研究[J]. 蚕业科学, 2020, 46(1): 86-95.

[9] 杨倩, 高晓阳, 武季玲, 等. 基于颜色和纹理特征的大麦主要病害识别研究[J]. 中国农业大学学报, 2013, 18(5): 129-135.

[10] 员伟康, 木拉提·哈密提, 严传波, 等. 基于颜色和纹理特征的新疆维吾尔医植物药材图像特征提取与判别分析[J]. 中国中医药信息杂志, 2016, 23(1): 78-81.

[11] 刘永波, 雷波, 胡亮, 等. 机器视觉在 HSV 颜色空间下稻瘟病病程分级判定研究[J]. 农学学报, 2020, 10(10): 83-90.

[12] 苗荣慧, 杨华, 武锦龙, 等. 基于图像分块及重构的菠菜重叠叶片与杂草识别[J]. 农业工程学报, 2020, 36(4): 178-184.

[13] 谢文涌, 柴琴琴, 甘永辉, 等. 基于多特征提取和 Stacking 集成学习的金线莲品系分类[J]. 农业工程学报, 2020, 36(14): 203-210.

[14] 宋彦, 谢汉垒, 宁井铭, 等. 基于机器视觉形状特征参数的祁门红茶等级识别[J]. 农业工程学报, 2018, 34(23): 279-286.

[15] 段红燕, 邵豪, 张淑珍, 等. 一种基于 Canny 算子的图像边缘检测改进算法[J]. 上海交通大学学报, 2016, 50(12): 1861-1865.

[16] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.

[17] 曹正凤. 随机森林算法优化研究[D]. 北京: 首都经济贸易大学, 2014.

[18] 陈进, 毛先成, 刘占坤, 等. 基于随机森林算法的大尹格庄金矿床三维成矿预测[J]. 大地构造与成矿学, 2020, 44(2): 231-241.

[19] 王猛, 张新长, 王家耀, 等. 结合随机森林面向对象的森林资源分类[J]. 测绘学报, 2020, 49(2): 235-244.

[20] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.

[21] 王珏. 基于 SVM 和组合特征的分类算法研究[D]. 大连: 大连理工大学, 2017.

[22] 赵立新, 侯发东, 吕正超, 等. 基于迁移学习的棉花叶部病虫害图像识别[J]. 农业工程学报, 2020, 36(7): 184-191.

【责任编辑 李晓卉】