

引用格式:

赵晨, 王涛, 郭伟雄, 孙光伟, 路晓崇, 宋朝鹏, 陈振国. 基于机器视觉的烤烟烟叶部位的智能识别[J]. 湖南农业大学学报(自然科学版), 2023, 49(4): 405–411.

ZHAO C, WANG T, GUO W X, SUN G W, LU X C, SONG Z P, CHEN Z G. Intelligent recognition of flue-cured tobacco position based on machine vision[J]. Journal of Hunan Agricultural University(Natural Sciences), 2023, 49(4): 405–411.

投稿网址: <http://xb.hunau.edu.cn>



基于机器视觉的烤烟烟叶部位的智能识别

赵晨¹, 王涛², 郭伟雄³, 孙光伟⁴, 路晓崇¹, 宋朝鹏¹, 陈振国^{4*}

(1.河南农业大学烟草学院, 河南 郑州 450002; 2.云南省烟草公司曲靖市公司, 云南 曲靖 655000; 3.广东中烟工业有限责任公司, 广东 广州 510385; 4.湖北省烟草科学研究院, 湖北 武汉 430030)

摘要: 采集烤烟不同叶位烟叶图像并进行预处理, 提取烟叶叶长、叶宽、面积、周长和最小外接矩形面积 5 个叶片绝对形态特征参数, 计算狭长度、矩形度、圆形度、叶宽最大处占比、叶宽轴与质心夹角 5 个相对形态特征; 通过主成分分析筛选出特征向量, 构建基于 K 近邻算法(KNN)、逻辑回归(LR)、基于线性核函数和径向基核函数的支持向量机(SVM)和 BP 神经网络的烟叶部位识别模型, 并对比 5 种模型的识别效果。结果表明: 基于图像轮廓特征所提取的形态特征参数可以较为有效地反映烤烟部位特征; 5 种识别模型中, 基于 BP 神经网络模型的识别效果最好, 识别准确率为 93.75%, 训练集和测试集的模型决定系数均高于 90%。

关键词: 烤烟; 部位; 形态特征; 图像处理; 智能判别

中图分类号: TP391.41

文献标志码: A

文章编号: 1007-1032(2023)04-0405-07

Intelligent recognition of flue-cured tobacco position based on machine vision

ZHAO Chen¹, WANG Tao², GUO Weixiong³, SUN Guangwei⁴,
LU Xiaochong¹, SONG Zhaopeng¹, CHEN Zhenguo^{4*}

(1.College of Tobacco, Henan Agricultural University, Zhengzhou, Henan 450002, China; 2.Qujing Tobacco Company of Yunnan Province, Qujing, Yunnan 655000, China; 3.China Tobacco Guangdong Industry Co. Ltd, Guangzhou, Guangdong 510385, China; 4.Tobacco Research Institute of Hubei Province, Wuhan, Hubei 430030, China)

Abstract: The images of flue-cured tobacco leaves at different leaf positions were collected and preprocessed, to extract 5 absolute shape features of leaf length, leaf width, area, perimeter and minimum external rectangular area, and to calculate 5 relative shape features of narrowness, rectangularity, roundness, percentage of leaf width at the maximum and the angle between leaf width axis and center of mass. The feature vector was filtered out by the main component analysis to construct 5 tobacco leaf position recognition models based on K-nearest neighbor(KNN), logistic regression(LR), support vector machine(SVM) with linear kernel function and radial basis kernel function, and BP neural network, respectively. The recognition effects were compared for the five models. The results showed that the morphological feature parameters extracted based on the image contour features could reflect the characteristics of roasted tobacco positions more effectively. It could be seen that the BP neural network-based model has the best recognition effect with a recognition accuracy of 93.75% among the five recognition models, and the model decision coefficient is above 90% for both the training and test sets.

收稿日期: 2022-04-25

修回日期: 2023-05-08

基金项目: 中国烟草总公司科技重点研发项目(110202102007); 中国烟草总公司湖北省公司项目(027Y2019-006); 中国烟草总公司云南省公司项目(2021530000241036)

作者简介: 赵晨(1998—), 女, 河南平顶山人, 硕士研究生, 主要从事烟草调制技术研究, qchen54533@qq.com; *通信作者, 陈振国, 高级农艺师, 主要从事烟草栽培及调制技术研究, hbskys1@163.com

Keywords: flue-cured tobacco; position; morphological characteristics; image processing; intelligent discrimination

烟叶的着生部位是烟叶分组和分级的重要依据^[1],直接影响着烤后烟叶的外观质量^[2]和化学成分^[3]。在生产实际中,上部薄叶、旱脚叶、杈叶等特殊叶片的实际外观特征与烤烟国家标准所规定的烟叶部位外观特征存在矛盾^[4],依靠采收人员的个人经验对烤烟部位的判断有差异,导致采后烤烟的差异较大。建立不依赖于人工判别的数字化判别方法来客观、准确地判别烤烟部位,对保障烤烟质量有重要意义。

目前,多数研究采用机器视觉技术对植物叶片进行无损检测分类识别。董本志等^[5]基于 Freeman 链码方法对荷兰牡丹、榆树和菊花叶片图像中的拐角点进行检测,叶面积和叶周长的计算精确度达到 90%。董红霞等^[6]基于狭长度、矩形度、球状性等 7 项相对几何特征和纹理特征,利用 BP 神经网络对杨树叶、银杏叶、枫叶等 6 种植物叶片进行分类,识别准确率达到 98.4%。魏蕾等^[7]选取多个特征参数,利用 SVM 分类器对木瓜、三角枫等 4 种植物叶片进行识别,准确率达到 95.8%。LUKIC 等^[8]基于 Flavia 数据集采用 Hu 不变矩和 LBP 算法提取特征,SVM 作为分类器进行植物叶片分类识别,模型识别准确率达到 94.13%。李洋等^[9]提出一种基于形态特征的植物叶片识别算法,采用 KNN-SVM 对三角枫、鹅掌楸等叶片进行分类识别,识别准确率为 94.77%。ZHANG 等^[10]对 10 种分类器进行评估,发现在不同特征的叶片识别中,随机森林和逻辑回归具有较高的准确性和稳定性。

在烟草检测领域,李翠英等^[11]提取了能表征不同部位的烟叶轮廓特征,基于烟叶边缘拟合的二元正交多项式能实现烤后烟叶部位分类。庄珍珠等^[12]采用烟叶形状与颜色融合特征,基于模糊数学方法实现了烤后烟叶分级。汪强等^[13]通过烟叶成熟度与 SPAD 值之间的函数关系,构建基于 HSV 颜色特征值的成熟度判别模型 TMDHSV,对烟叶成熟度进行判定。机器视觉方法在烟草识别领域的研究主要针对烤后烟叶的分级和烟叶成熟度的识别,大多采用烟叶颜色、纹理图像特征^[14],对鲜烟叶部位的识别鲜有报道。笔者提出一种基于烟叶形态特征的鲜烟叶部位识别方法,采用机器视觉技术,对所采集

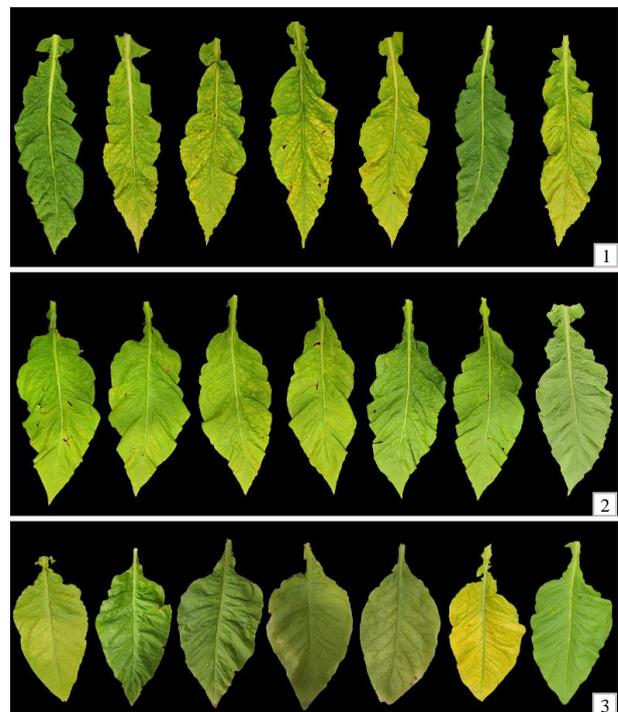
鲜烟叶图像进行预处理,提取烟叶的形态特征参数;采用主成分分析(PCA)进行特征降维,构成烟叶特征数据集;采用 SVM、BP 神经网络、KNN 和 LR 分类器对鲜烟叶部位进行识别分类。现将结果报告如下。

1 烤烟叶片图像的处理

1.1 图像采集及预处理

2020—2021 年,选取河南农业大学试验基地规范管理烟田的代表性烟株,上部叶采收 15~18 叶位、中部叶采收 8~14 叶位、下部叶采收 5~7 叶位的烟叶。

将烟叶平铺放置在黑色背景布中央,将奥林巴斯 XZ-1 CCD 相机固定在三脚架上,距地面约 1.5 m,利用水平仪调节镜头与地面垂直,采集烟叶图像,如图 1 所示。共采集 480 个烟叶样本,其中上部叶 164 个、中部叶 168 个、下部叶 148 个。烤烟烟叶原始图像为 24 位真彩色图像,分辨率为 3648 像素×2736 像素。按 4:1 的比例随机选取训练集和测试集。



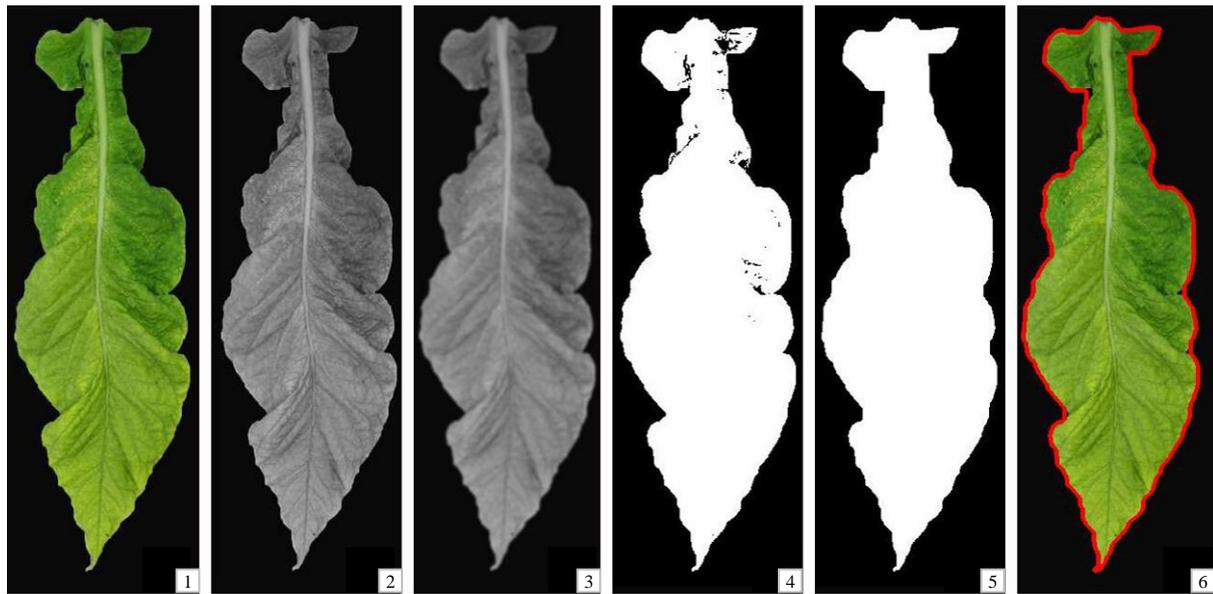
1、2、3 分别为上部叶、中部叶和下部叶。

图 1 不同部位的烟叶图像

Fig.1 Tobacco leaf images at different positions

参照文献^[15]的图像处理方法,将原始图像转化为灰度图像,通过高斯滤波、最大类间方差法和形

态学处理，对图像中叶片与背景的分离等进行预处理(图 2)，采用 Canny 算子提取烤烟轮廓边缘曲线。



1 原始图像; 2 灰度化; 3 高斯滤波; 4 二值化; 5 形态学处理; 6 轮廓提取。

图 2 烤烟烟叶图像的预处理结果

Fig.2 Pre-processing results of tobacco leaf images

1.2 烟叶特征的提取

1.2.1 绝对形态特征参数的提取

为准确识别烤烟烟叶部位，对烤烟烟叶的形态特征参数^[16]进行提取。对烟叶图像预处理后提取的轮廓信息构建叶片轮廓的外接矩形^[17]，提取叶片的绝对形态特征参数(图 3)。

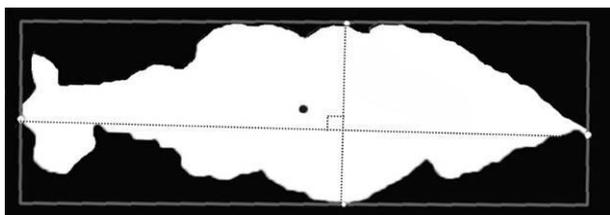


图 3 烤烟叶片绝对形态特征参数

Fig.3 Extraction of absolute morphological parameters for tobacco leaf

1) 叶长。叶片轮廓上端和下端距离最大的 2 点之间连线的像素点个数。

2) 叶宽。叶片轮廓上与叶长 2 点连线垂直且距离最远的 2 点连线的像素点个数。

3) 面积。叶片轮廓包含的像素点个数。

4) 周长。叶片轮廓边缘的像素点个数。

5) 最小外接矩形面积。叶片轮廓的最小外接矩形面积的像素点个数。

1.2.2 相对形态特征参数的提取

在样本图像处理过程中，缩放、旋转和平移等

操作对绝对形态特征参数的数值结果影响较大，因此烤烟的叶长、叶宽、面积等绝对形态指标不足以作为分类依据。依据文献[18]，选取烟叶狭长度、矩形度、圆形度、叶宽最大处占比、叶宽轴与质心夹角弧度 5 项相对几何特征，表征烟叶形态特征。

1) 狭长度。叶长和叶宽的比值，用于描述叶片圆形程度。

2) 矩形度。烤烟面积与其外接矩形的面积之比，反映叶片对其外接矩形的充满程度。

3) 圆形度。周长的平方与面积的比值，描述叶片边界复杂程度。

4) 叶宽最大处占比。叶宽最大处距叶基的距离与叶长的比值，反映叶片宽度变化程度。

5) 叶宽轴与质心夹角弧度。叶宽最大处两点与质心连线所成夹角对应的弧度值，描述叶片宽度特征。

由图 4 可知，烤烟不同部位叶片的狭长度、矩形度和叶宽 3 个特征参数的差异较大，圆形度、叶宽轴与质心夹角弧度的差异性较小，叶长、叶面积、叶周长、外接矩形面积和叶宽最大处位置这 5 个特征参数的分散程度较高，差异不显著。烤烟形态特征参数的描述性统计结果(表1)表明，烟叶狭长度、矩形度、叶长和叶周长的变异系数均小于 10%，叶面积和外接矩形面积的变异系数达到 15%，总体来看，所提取烤烟

的形态特征变异系数均小于 16%，采用这 10 个形态

特征参数进行鲜烤烟部位识别有效。

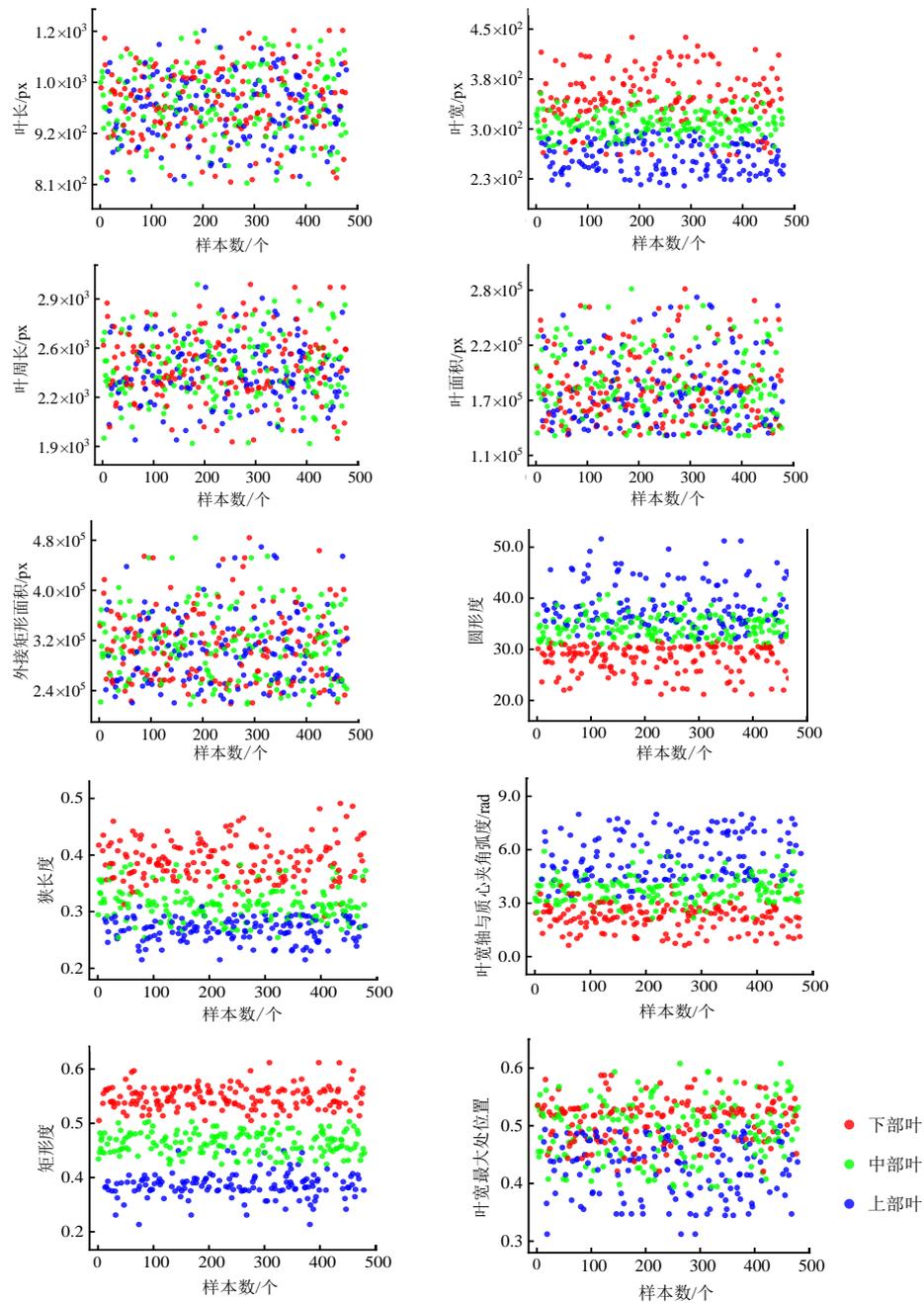


图 4 烤烟不同部位烟叶的形态特征分布

Fig.4 Morphological characteristics distribution of tobacco leaves at different positions

表 1 烤烟不同部位烟叶形态特征参数的描述性统计

Table 1 Descriptive statistic results of morphological characteristics of tobacco leaves at different positions

烟叶的形态特征	部位	极小值	极大值	均值	标准差	偏度	峰度	变异系数/%
圆形度	下部叶	21.19	31.56	28.10	2.71	-0.95	-0.14	9.66
	中部叶	31.04	40.72	34.44	2.22	0.67	0.17	6.44
	上部叶	32.04	51.62	38.71	4.55	0.81	-0.03	11.77
狭长度	下部叶	0.29	0.49	0.39	0.04	0.28	0.08	9.20
	中部叶	0.25	0.38	0.31	0.03	0.41	-0.19	9.29
	上部叶	0.22	0.31	0.27	0.02	-0.43	-0.28	7.58

表 1(续)

烟叶的形态特征	部位	极小值	极大值	均值	标准差	偏度	峰度	变异系数/%
矩形度	下部叶	0.49	0.61	0.54	0.03	0.49	0.59	4.66
	中部叶	0.39	0.48	0.44	0.02	-0.05	-0.90	5.53
	上部叶	0.26	0.42	0.34	0.03	0.01	1.71	8.25
叶宽轴与质心	下部叶	0.63	3.97	2.22	0.31	-0.14	-0.20	13.96
	中部叶	2.17	5.91	3.75	0.53	0.53	0.58	14.13
夹角弧度/rad	上部叶	3.33	7.99	5.61	0.68	0.20	-1.15	12.12
	下部叶	0.42	0.59	0.51	0.04	-0.17	-0.38	7.32
	中部叶	0.39	0.61	0.49	0.05	-0.08	-0.71	10.46
叶宽最大处位置	上部叶	0.31	0.50	0.43	0.05	-0.38	-0.95	11.60
	下部叶	261.09	437.67	348.54	41.03	-0.23	-0.11	11.77
	中部叶	272.94	353.31	308.10	18.53	0.26	-0.30	6.02
叶宽/px	上部叶	214.45	308.21	257.23	23.19	0.11	-1.07	9.01
	下部叶	822.74	1152.18	994.45	78.72	-0.27	-0.29	7.92
	中部叶	805.84	1134.57	974.79	88.50	-0.17	-1.02	9.08
叶长/px	上部叶	885.41	1134.15	979.38	56.53	0.41	-0.37	5.77
	下部叶	2021.63	2974.34	2512.55	233.48	-0.23	-0.46	9.29
	中部叶	1938.42	2852.60	2416.27	211.76	-0.32	-0.24	8.76
叶周长/px	上部叶	2105.56	2601.08	2342.74	120.31	0.21	-0.58	5.14
	下部叶	1.37×10 ⁵	2.70×10 ⁵	2.21×10 ⁵	3.41×10 ⁴	-0.32	-0.30	15.08
	中部叶	1.29×10 ⁵	2.44×10 ⁵	1.76×10 ⁵	2.64×10 ⁴	0.68	0.39	15.02
外接矩形面积/px	上部叶	1.29×10 ⁵	1.86×10 ⁵	1.53×10 ⁵	1.61×10 ⁴	0.27	-1.03	10.50
	下部叶	2.55×10 ⁵	4.83×10 ⁵	3.56×10 ⁵	5.38×10 ⁴	0.18	-0.16	15.09
	中部叶	2.17×10 ⁵	3.95×10 ⁵	3.01×10 ⁵	4.38×10 ⁴	0.07	-0.55	14.58
外接矩形面积/px	上部叶	2.21×10 ⁵	3.45×10 ⁵	2.74×10 ⁵	3.23×10 ⁴	0.39	-1.11	11.83

2 烟叶部位识别模型的建立

为避免维数过高导致的数据冗余对分类模型效率和精确度的影响,对烤烟烟叶的 10 个形态特征参数进行主成分分析。由图 5 可知,前 2 个主成分变量可以解释 10 个形态特征数据所包含 90.2%(PC1 66.2%、PC2 24.0%)的信息,烟叶 3 个部位都表现出较好的聚类效果,说明形态特征对于烤烟烟叶部位分类预测具有较好的效果。根据 10 个形态特征在前 2 个主成分变量的载荷值,剔除与部位分类相关度较低的形态特征,最终选取对部位分级贡献率较大的前 3 个形态特征参数即烟叶狭长度、矩形度和圆形度以及 PCA 筛选的前 2 个主成分变量作为特征变量,构建基于 K 近邻算法(KNN)、逻辑回归(LR)、基于线性核函数的支持向量机(SVM-linear)、基于径向基核函数的支持向量机(SVM-rbf)和基于 BP 神经网络的识别模型,比较这 5 种模型的识别效果。

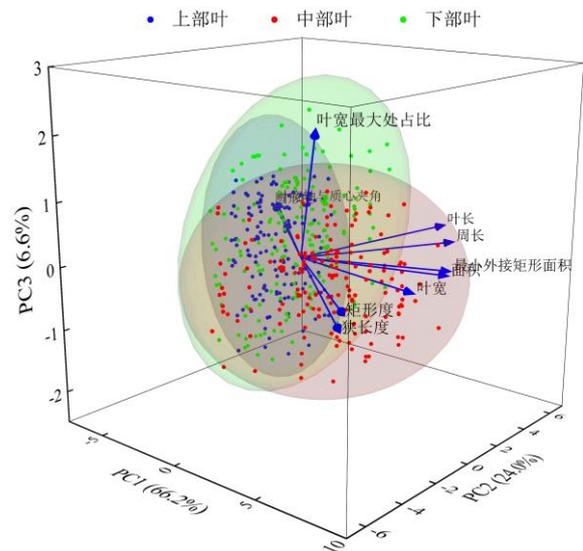


图 5 鲜烤烟的形态特征主成分分析

Fig.5 Analysis of principal component for tobacco leaves shape characteristics

选择模型的决定系数和识别准确度作为评价指标^[19],评价模型效果。

3 模型的评估与验证

采用所构建的5种模型对不同部位烤烟烟叶进行识别, LR模型的正则化系数、SVM模型的核函数参数以及BP神经网络模型的隐含层网络节点个数,均采用遍历优化法^[20]进行筛选;KNN模型中 k 值采用5折交叉验证法^[20]进行筛选,最终,SVM模型中linear核函数的惩罚系数为2.0,rbf核函数的核函数方差为0.4,惩罚系数为1.5;LR模型采用的L2正则化系数为10;BP神经网络模型的输入层层数为5,输出层层数为3,隐含层层数为5;KNN模型中所取近邻样本的个数为6。

对3个部位烟叶的识别混淆矩阵如表2所示。可以看出,KNN模型的误识次数最少,LR模型的误识次数最多,基于径向基核函数的SVM模型误识次数略低于基于线性和函数的SVM模型的误识次数。不同模型的误识都发生在相邻部位,其中中部叶错分概率最高,可能是相邻部位烤烟叶片特征存在相似性,目标区分度较小的缘故。

表2 不同模型对烤烟烟叶部位识别结果的混淆矩阵
Table 2 Confusion matrix of different tobacco leaf position identified by different models

识别模型	烟叶部位	识别结果		
		上部叶	中部叶	下部叶
LR	上部叶	26	4	0
	中部叶	12	6	12
	下部叶	0	0	36
SVM-linear	上部叶	24	6	0
	中部叶	5	25	0
	下部叶	0	0	36
SVM-rbf	上部叶	25	5	0
	中部叶	5	25	0
	下部叶	0	0	36
BP	上部叶	27	3	0
	中部叶	2	27	1
	下部叶	0	1	36
KNN	上部叶	25	5	0
	中部叶	4	26	0
	下部叶	0	0	36

基于LR、KNN、SVM和BP神经网络的模型验证结果如表3所示。LR模型的识别准确度仅有70.83%;SVM模型的识别结果优于LR模型的,其中采用线性核函数的模型识别准确度为88.54%,径

向基核函数的模型识别准确度为89.58%;KNN模型的识别效果优于SVM模型的,模型识别准确度达到了90.63%,模型决定系数达到了85%以上;基于BP神经网络模型的识别准确度为93.75%,训练集和测试集的模型决定系数均高于90%,识别效果最好。

综合不同模型的验证结果以及混淆矩阵,发现不同模型识别错误都发生在烤烟相邻部位烟叶,且中部叶误分概率高。可能是SVM、LR模型都通过非线性函数映射,对形成的超平面进行分割,相邻两部位鲜烤烟的外观特征参数数值分布差异较小,且存在不同烟株的同一部位烤烟外观特征差异较大现象,因此处于烟株部位分界点叶位的鲜烤烟其部位识别难度较大,这也是导致识别错误集中在相邻部位的主要原因。KNN算法通过寻找相邻样本进行分类,针对区域交叉或重叠较多的样本集具有更高的识别精确度,在样本数据量小的情况下也不容易出现过拟合现象,能够较好地对待烤烟烟叶的所属部位进行判定。BP神经网络算法的预测准确度更高,但所需样本量较大,在数据量较小的情况下容易出现过拟合。在实际烘烤中,同一烤房采摘鲜烤烟的形态特征相似度较高,使用BP神经网络和KNN算法可以有效提升识别精度,在智能烘烤体系构建过程中,可以运用此分类模型对烤烟烟叶部位进行智能无损识别。

表3 不同预测模型对鲜烤烟部位的识别效果

Table 3 Identification effects of different model for tobacco positions

模型	测试集识别准确度/%	测试集 R^2	训练集 R^2
SVM-linear	88.54	84.76	84.47
SVM-rbf	89.58	84.98	86.92
LR	70.83	62.57	74.43
BP	93.75	91.88	95.43
KNN	90.63	85.71	90.19

4 结论

针对不同部位烤烟烟叶的识别,采用机器视觉技术,提出基于形态特征的识别方法。结果表明,采用烤烟烟叶的狭长度、矩形度、圆形度等形态特征描述叶片形态,实现不同部位烤烟烟叶识别是可行的。处于烟株部位分界点叶位的鲜烤烟形态特征参数数值差异较小,在识别过程中易发生误分现

象, 所构建的鲜烤烟部位预测模型表现为基于 BP 神经网络模型的识别效果最好, 识别准确度达到 93.75%, 基于 KNN 算法的模型识别准确度仅次于 BP 神经网络模型, 为 90.63%, LR 模型的识别准确度最低。在不同部位烤烟烟叶识别中, BP 神经网络算法的预测准确度更高, 但模型训练所需样本量较大, 在样本量小的情况下, 采用 KNN 算法可以较好地对比烟叶的部位进行判定。

参考文献:

- [1] 邓云龙, 崔国民, 孔光辉, 等. 品种、部位和成熟度对烟叶淀粉含量及呼吸质量的影响[J]. 中国烟草科学, 2006, 27(4): 18–23.
- [2] 蔡杰, 朱国蕾, 孙晓燕, 等. 基于不同部位烟叶外观特征与内在质量分析研究[J]. 农业与技术, 2019, 39(13): 27–28.
- [3] 陈敏, 郭吉兆, 郑赛晶, 等. 烟叶部位、产地与卷烟主流烟气 7 种有害成分释放量关系研究[J]. 中国烟草学报, 2012, 18(5): 16–22.
- [4] 杨尚明, 李宜健, 肖云峰, 等. 特殊外观特征烟叶部位的识别判定研究[J]. 安徽农业科学, 2014, 42(5): 1534–1535.
- [5] 董本志, 康欣, 任洪娥. 植物叶片轮廓特征提取方法研究[J]. 计算机工程与应用, 2015, 51(8): 143–147.
- [6] 董红霞, 郭斯羽. 一种结合形状与纹理特征的植物叶片分类方法[J]. 计算机工程与应用, 2014, 50(23): 185–188.
- [7] 魏蕾, 何东健, 乔永亮. 基于图像处理和 SVM 的植物叶片分类研究[J]. 农机化研究, 2013, 35(5): 12–15.
- [8] LUKIC M, TUBA E, TUBA M L. Leaf recognition algorithm using support vector machine with Hu moments and local binary patterns[C]//2017 IEEE 15th International Symposium on Applied Machine Intelligence and Informatics(SAMI). Herlany, Slovakia: IEEE, 2017: 485–490.
- [9] 李洋, 李岳阳, 罗海驰, 等. 基于形状特征的植物叶片在线识别方法[J]. 计算机工程与应用, 2017, 53(2): 162–165.
- [10] ZHANG Q, ZENG S J, ZHANG B. Initial investigation of different classifiers for plant leaf classification using multiple features [C]// International Conference on Digital Image Processing. Guangzhou, China: SPIE, 2019: 1117922.
- [11] 李翠英, 贺立源, 马文杰, 等. 采用轮廓特征的烟叶部位组分类研究[J]. 计算机工程与应用, 2009, 45(26): 236–239.
- [12] 庄珍珍, 祝诗平, 孙雪剑, 等. 基于机器视觉的烟叶自动分组方法[J]. 西南师范大学学报(自然科学版), 2016, 41(4): 122–129.
- [13] 汪强, 席磊, 任艳娜, 等. 基于计算机视觉技术的烟叶成熟度判定方法[J]. 农业工程学报, 2012, 28(4): 175–179.
- [14] 杨尚明, 李宜健, 肖云峰, 等. 特殊外观特征烟叶部位的识别判定研究[J]. 安徽农业科学, 2014, 42(5): 1534–1535.
- [15] 朱静, 田兴军, 陈彬, 等. 植物叶形的计算机识别系统[J]. 植物学通报, 2005, 40(5): 599–604.
- [16] SABZI S, ABBASPOUR-GILANDEH Y, JAVADIKIA H. Machine vision system for the automatic segmentation of plants under different lighting conditions[J]. Biosystems Engineering, 2017, 161(12): 157–173.
- [17] GUYER D E, MILES G E, GAULTNEY L D, et al. Application of machine vision to shape analysis in leaf and plant identification[J]. Transaction of the ASABE, 1993, 36(1): 163–171.
- [18] 高良, 闫民, 赵方. 基于多特征融合的植物叶片识别研究[J]. 浙江农业学报, 2017, 29(4): 668–675.
- [19] 刘赛可, 何晓群, 夏利宇. 不平衡数据下模型评价指标的有效性探讨[J]. 统计与决策, 2022, 38(19): 5–9.
- [20] 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.

责任编辑: 罗慧敏

英文编辑: 吴志立