

# 基于 Bayes 决策的甘蔗种芽完好性检测与试验

黄亦其, 尹 凯, 黄媚章, 王小波, 罗昭宇

(广西大学机械工程学院, 南宁 530004)

**摘 要:** 由于甘蔗种芽的好坏直接影响来年新植甘蔗的出芽率, 针对目前甘蔗机械化种植过程中无法实现种芽完好性检测的难题, 该文提出一种基于 Bayes 决策的甘蔗种芽完好性分类方法, 将甘蔗种芽分为完好与破损 2 类。首先利用计算机视觉技术, 采集甘蔗种芽图像, 在截取有效的种芽区域图像并提取出包括灰度最大值在内的 5 个分类特征的基础上, 统计种芽区域图像特征值的分布情况; 然后分析特征值的曲线分布图, 利用这些特征的均匀分布区间占整个分布区间的百分比以及完好和破损 2 类种芽相应百分比的差值, 确定能够简化为均匀分布的特征, 并将其作为最终的分类特征; 再结合全概率公式和贝叶斯公式, 将确定的分类特征的条件概率和先验概率转换为后验概率, 通过比较后验概率的大小进行分类; 最后选择常见的 3 个甘蔗品种作为研究对象, 利用 Matlab 对样本进行种芽完好性分类处理。试验结果显示, 3 个品种的种芽完好性分类准确率分别为 92.09%、93.49%和 93.02%, 破损种芽的分类准确率分别达到 98%、97%和 96%, 该分类方法能够实现甘蔗种芽完好性的分类功能。

**关键词:** 无损检测; 甘蔗; 决策; 种芽; 完好性

doi: 10.11975/j.issn.1002-6819.2016.05.008

中图分类号: S126; TP391.41

文献标志码: A

文章编号: 1002-6819(2016)-05-0057-07

黄亦其, 尹 凯, 黄媚章, 王小波, 罗昭宇. 基于 Bayes 决策的甘蔗种芽完好性检测与试验[J]. 农业工程学报, 2016, 32(5): 57—63. doi: 10.11975/j.issn.1002-6819.2016.05.008 http://www.tcsae.org

Huang Yiqi, Yin Kai, Huang Meizhang, Wang Xiaobo, Luo Zhaoyu. Detection and experiment of sugarcane buds integrity based on Bayes decision[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2016, 32(5): 57—63. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.2016.05.008 http://www.tcsae.org

## 0 引 言

广西是中国最大的甘蔗产区, 连续多年产量超过全国总产量的 60%<sup>[1]</sup>。农艺要求表明, 甘蔗种芽完好是新植甘蔗出芽率的基本保障, 要求蔗种的切面光滑, 种节不破裂, 种芽不破损, 在下种时要仔细检查, 将死芽、烂芽去掉<sup>[2-3]</sup>, 但是, 在甘蔗收获、运输、留种和切种过程中, 种芽由于其凸起的形态特征, 容易受到外力的摩擦、挤压和碰撞而导致破损。

因此, 在甘蔗种植过程中需要进行甘蔗种芽的完好性检测, 尽可能多的识别出破损种芽。对于识别分类, 国内外已经有了不少研究, 在农业领域也得到了广泛应用, 主要有支持向量机 (support vector machine, SVM)、人工神经网络和 Bayes 决策理论 3 种, 例如: 基于 SVM 的蜜柚叶部病害识别<sup>[4]</sup>, 稻谷品种鉴别<sup>[5]</sup>, 猪肉颜色分级<sup>[6]</sup>, 淡水鱼种类识别<sup>[7]</sup>等, 基于人工神经网络的普洱茶发酵程度判定<sup>[8]</sup>, 新陈莲子鉴定<sup>[9]</sup>, 冬小麦叶面积指数反演<sup>[10]</sup>等, 基于 Bayes 决策的棉花盲椿象危害的等级识别<sup>[11]</sup>, 树叶边缘识别<sup>[12]</sup>, 杂草种子分类<sup>[13]</sup>, 苹果图像分割<sup>[14]</sup>等。针对甘蔗切种时造成的种芽破损问题, 学者提出了一些蔗种防伤芽方案, 如黄亦其<sup>[15]</sup>, 陆尚平<sup>[16]</sup>, 乔曦<sup>[17-18]</sup>, 唐

书喜<sup>[19]</sup>等利用图像处理技术检测甘蔗茎节, 王小波等<sup>[20-22]</sup>利用传感检测技术检测甘蔗茎节, 其均是利用甘蔗茎节来间接确定甘蔗种芽位置, 然后控制切刀避开种芽, 实现甘蔗切种防伤芽, 虽然这些蔗种防伤芽措施已将平均伤芽率控制在 2% 以下, 但切种前无法检测出已经损坏的种芽。对此陆尚平<sup>[23]</sup>利用图像处理技术对甘蔗种芽进行了分割算法试验, 并提出了改进的 8 点等效最大内切圆分离甘蔗种芽和定位甘蔗种芽的方法, 罗昭宇等<sup>[24]</sup>通过采集甘蔗种芽图像, 并用 Matlab 检测其种芽边缘并与完好种芽边缘特征进行比较, 来区分完好与破损种芽, 但上述研究都没有提出一种明确的种芽完好性分类规则, 在实际生产中无法实现种芽完好性分类, 从而无法实现自动剔除种芽破损的甘蔗种。

综上所述, 为了尽可能多的找出破损种芽, 实现甘蔗种植过程中自动剔除破损种芽的功能, 本文利用 Bayes 决策理论, 通过分析甘蔗种芽图像分类特征的分布规律, 提出一种甘蔗种芽完好性分类方法, 将甘蔗种芽分为完好与破损 2 类 (完好是指甘蔗种芽所有的生理结构完整, 具有正常发芽能力的种芽形态; 破损是指甘蔗种芽表层或内部的生理结构损坏, 无法正常发芽或者不发芽的种芽形态)。

## 1 种芽图像样本采集

### 1.1 图像获取

甘蔗样本来源于广西崇左市扶绥县渠黎镇笃邦村甘蔗种植基地 (金得肥有限公司经营), 从该基地取种植面积较大、最具有代表性的台糖 22 号、桂糖 42 号和粤糖 60 号

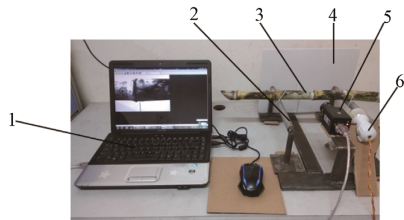
收稿日期: 2015-09-20 修订日期: 2016-01-05

基金项目: 广西自然科学基金资助项目 (2014GXNSFAA118344)

作者简介: 黄亦其, 男, 壮族, 广西隆安人, 博士, 副教授, 从事现代农业装备机电一体化研究。南宁 广西大学机械工程学院, 530004。

Email: hyqgxu@163.com

3 个品种的甘蔗为研究对象。图 1 为种芽图像拍摄平台, 利用 MVC360MF 摄像头自带的拍摄软件 MVC Capture 控制拍照系统对种芽样本进行图像采集, 镜头型号为 Pentax H612A(TH)。甘蔗种植过程中, 切种工作基本在室内进行, 为了方便拍摄, 在室内进行图像采集。拍摄时, 用普通的 40 W 白炽灯进行补光, 并使用白色纸板作为背景板。同时, 调节镜头上的光圈和调焦圈使甘蔗图像亮度合适且成像清晰, 拍摄后得到为 752×480 像素的种芽图像。

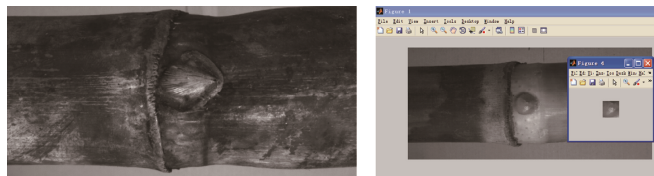


1.计算机 2.支架 3.甘蔗 4.背景板 5.摄像头 6.白炽灯  
1.Computer 2.Bracket 3.Sugar cane 4.Background Board 5.Camera 6.Incandescent lamp

图 1 种芽图像拍摄平台

Fig.1 Shooting platform of bud image

拍摄过程中, 人工将甘蔗样本放置在支架上, 并手动调节甘蔗样本的位置, 确保甘蔗种芽位于甘蔗样本图像中心位置, 每个品种采集种芽完好和种芽破损图像各 100 张。图 2a 为台糖 22 号种芽完好的甘蔗图像样例。



a. 台糖 22 号种芽完好甘蔗图像  
a. Intact sugar cane bud image of Taitang 22  
b. 种芽区域截取效果图  
b. Renderings of bud area cutting

图 2 种芽样本图像

Fig.2 Bud sample image

## 1.2 图像预处理

经过观测发现, 一幅 752×480 像素的甘蔗样本图像, 只有大约 40×40 像素的区域是研究所需要的种芽区域, 为了使研究简化, 利用 MATLAB 截取甘蔗样本图像中 40×40 像素的种芽区域, 得到种芽区域样本图像 (以下简称种芽图像), 如图 2b 所示。在种芽图像样本中, 选出 6 张典型的完好种芽图像样本与 6 张典型的破损种芽图像样本 (包含死芽、烂芽等典型破损情况), 并截取图像样本的种芽区域, 如图 3 所示。对比图 3 中的完好种芽与破损种芽, 发现 2 类种芽图像的颜色和纹理特征的区别都不明显。完好种芽凸起的高度和破损种芽凹陷的深度等结构特征确实存在, 而且区分明显, 但是镜头正对种芽拍摄得到的平面图不能明显地反应出该特征。因此, 无法用颜色、纹理、结构等方面的特征来进行种芽完好性检测。对种芽图像进行进一步分析发现, 完好种芽都包含了某块亮度异常突出的小区域, 而破损种芽则不存在此亮度突出的小区域, 而是包含了部分黑色或较暗的小区域, 因此可初步选定种芽图像的明暗特征作为完好

与破损种芽的区分特征。对种芽图像进行灰度处理, 得到了灰度最大值 (maximum gray scale, Max)、灰度最小值 (minimum gray scale, Min)、灰度平均值 (mean gray scale, Mean)、灰度值标准差 (standard deviation of gray scale, Std) 以及灰度值中位数 (gray scale median, Median, 指将所有灰度值数据排序后, 位置在最中间的灰度值) 5 个甘蔗种芽分类特征。

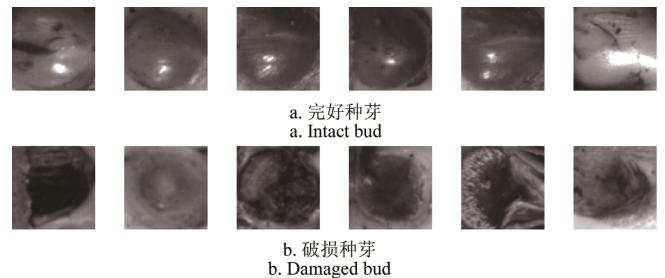


图 3 完好与破损种芽特征对比图

Fig.3 Feature comparison chart of intact and damaged buds

## 2 分类方法

### 2.1 基于最小错误率的 Bayes 决策

SVM、人工神经网络和 Bayes 决策理论这 3 种识别方法均得到了广泛的应用, 但 Bayes 决策理论的时效性更高, 更加适合分类特征已知的场合, 本文已经提取出了种芽的分类特征, 因此选择 Bayes 决策理论作为本文分类方法。

Bayes 决策理论包括最小错误率和最小损失两大准则, 分别用于衡量准确率的高低和误判代价的大小。在实际生产中, 种芽完好的甘蔗种因误判而被剔除后, 还可以返回糖厂榨糖, 误判损失较小, 结合本文的研究目的是为了尽可能多的找出破损的种芽, 则本文研究的分类方法应以最小错误率为准则, 优先考虑破损种芽的识别准确率。

$\omega_1$  和  $\omega_2$  分别表示种芽完好和破损的 2 种状态, 由贝叶斯公式<sup>[25]</sup>可得

$$P(\omega_i | x) = \frac{P(x | \omega_i)P(\omega_i)}{\sum_{i=1}^m P(x | \omega_i)P(\omega_i)} \quad (1)$$

式中  $P(\omega_i)$  为  $\omega_i$  类的先验概率, 表示  $\omega_i$  出现的概率;  $P(x | \omega_i)$  是  $x$  的类条件概率, 表示属于  $\omega_i$  类的条件下出现  $x$  的概率;  $P(\omega_i | x)$  为后验概率, 表示将  $x$  放入  $\omega_i$  类, 判断其属于  $\omega_i$  类的概率。

根据 Bayes 决策理论<sup>[14,26]</sup>, 分类规则描述为

若  $P(\omega_i | x_j) = \max_{i=1,2,j=1,2,3,4,5} \{P(\omega_i | x_j)\}$ , 则样本  $x$  属于  $\omega_i$  类 (2)

即将  $x$  归属为后验概率最大的类别。其中,  $j=1, 2, 3, 4, 5$ , 分别表示 Max、Min、Mean、Std 和 Median5 类特征值。

由于  $\sum_{i=1}^m P(x | \omega_i)P(\omega_i)$  为全概率公式, 对于所有类别的  $\omega_i$  都是相等的<sup>[27]</sup>, 则分类规则可改写成

若 $P(x|\omega_1) \cdot P(\omega_1) > P(x|\omega_2) \cdot P(\omega_2)$ ,  $x$ 属于 $\omega_1$ 类,

若 $P(x|\omega_1) \cdot P(\omega_1) < P(x|\omega_2) \cdot P(\omega_2)$ ,  $x$ 属于 $\omega_2$ 类。

(3)

由式(3)可知,先验概率 $P(\omega_i)$ 和类条件概率 $P(x|\omega_i)$ 是 Bayes 决策分类的基础。

2.2 种芽区域特征的先验概率  $P(\omega_i)$

台糖 22 号、桂糖 42 号和粤糖 60 号种芽图像特征

值组成的数据库如表 1, 每一个甘蔗品种种芽图像总数为 200, 记为  $N$ , 其中包括 100 个完好种芽图像和 100 个破损种芽图像, 分别用  $N_1$  和  $N_2$  表示。由先验概率的概念可知

$P(\omega_1)=N_1 / N=50\%$ ,

$P(\omega_2)=N_2 / N=50\%$ 。

(4)

表 1 台糖 22 号、桂糖 42 号、粤糖 60 号种芽图像特征数据表

Table 1 Feature data of bud region image of Taitang 22, Guitang 42 and Yuetang 60

甘蔗品种 Sugarcane varieties	序号 Number	完好种芽 Intact bud					破损种芽 Damaged bud				
		最大值 Max	最小值 Min	平均值 Mean	中位数 Std	标准差 Median	最大值 Max	最小值 Min	平均值 Mean	中位数 Std	标准差 Median
台糖 22 号 Taitang 22	1	0.6431	0.1922	0.3194	0.0721	0.3098	0.4039	0.0627	0.1543	0.0608	0.1451
	2	0.7176	0.2078	0.313	0.0677	0.302	0.4431	0.0627	0.1818	0.0810	0.1647
	3	0.3725	0.1255	0.2222	0.0475	0.2196	0.3922	0.0588	0.1571	0.0645	0.1451
	...	...	...	...	...	...	...	...	...	...	...
	97	0.4902	0.2196	0.3602	0.0482	0.3647	0.6863	0.1765	0.4060	0.0717	0.3961
	98	1.0000	0.2039	0.4627	0.1285	0.4627	0.7098	0.1137	0.3031	0.1074	0.2863
	99	0.8392	0.3098	0.5077	0.0850	0.4980	0.4706	0.1686	0.2946	0.0569	0.2863
桂糖 42 号 Guitang 42	100	1.0000	0.0902	0.4831	0.1280	0.4588	0.6157	0.2392	0.4270	0.0690	0.4353
	1	0.7686	0.1216	0.3674	0.1345	0.3490	0.7216	0.1059	0.3308	0.0759	0.3490
	2	0.7255	0.0863	0.2988	0.1008	0.2724	0.5255	0.1098	0.2799	0.0639	0.2824
	3	1.0000	0.1412	0.4494	0.1427	0.4314	0.3647	0.1373	0.2226	0.0391	0.2196
	...	...	...	...	...	...	...	...	...	...	...
	97	1.0000	0.3882	0.6113	0.1312	0.6000	0.4039	0.1059	0.1903	0.0459	0.1882
	98	1.0000	0.3373	0.5695	0.1371	0.5686	0.4863	0.1255	0.2587	0.0498	0.2510
粤糖 60 号 Yuetang 60	99	1.0000	0.3412	0.5584	0.1621	0.5235	1.0000	0.0863	0.2120	0.1148	0.1882
	100	1.0000	0.1333	0.4354	0.1746	0.3961	0.5020	0.0863	0.2287	0.0960	0.2087
	1	1.0000	0.2392	0.4354	0.1231	0.4196	0.5882	0.1137	0.2721	0.0885	0.2588
	2	1.0000	0.1059	0.4209	0.2282	0.3373	0.3451	0.1137	0.2199	0.0361	0.2235
	3	1.0000	0.1373	0.3896	0.1690	0.3569	0.4510	0.0784	0.1715	0.0774	0.1471
	...	...	...	...	...	...	...	...	...	...	...
	97	1.0000	0.1020	0.3477	0.1285	0.3294	0.2235	0.0627	0.1067	0.0254	0.1020
	98	1.0000	0.0980	0.2909	0.1543	0.2333	0.3137	0.0627	0.1743	0.0496	0.1725
	99	0.8078	0.0980	0.3479	0.1315	0.3490	0.4902	0.0863	0.2090	0.0514	0.2000
	100	1.0000	0.0980	0.3230	0.1270	0.3098	0.5608	0.0667	0.2092	0.0905	0.1882

初步分析表 1 中的数据,发现无论是完好种芽还是破损种芽,3 个甘蔗品种的 5 类特征值之间并没有明显的规律,需要对这些特征数据进行进一步的分析。

2.3 种芽分类特征类条件概率  $P(x|\omega_i)$

种芽图像特征值数据库包括 Max、Min、Mean、Std 和 Median 5 类特征值,将表 1 中台糖 22 号种芽图像的每

一类特征值分别按由小到大的顺序排列,并从中取出最小值和最大值,构成种芽图像 5 类特征值的分布区间,同时将个别与相邻数据的差值在分布区间范围 8%以上的数据视为误差值,并剔除,得到了如表 2 所示的完好和破损种芽图像 5 类特征值的分布区间,用 $[A_j,B_j]$ 、 $[a_j,b_j]$ 分别表示完好和破损种芽特征值的分布区间。

表 2 台糖 22 号种芽图像特征分布区间

Table 2 Distribution range of bud image features for Taitang 22

种芽种类 Bud species	最大值 Max.	最小值 Min.	平均值 Mean	中位数 Std.	标准差 Median
完好种芽 Intact bud	[0.3725,1.0000]	[0.0667,0.4078]	[0.1559,0.6088]	[0.0390,0.2331]	[0.1490,0.5666]
破损种芽 Damaged bud	[0.2316,0.7843]	[0.0549,0.3059]	[0.0995,0.4456]	[0.0175,0.1765]	[0.0824,0.4706]

根据表 1 的数据,统计完好种芽灰度最大值的分布情况,得到台糖 22 号完好种芽灰度最大值的统计情况,如表 3 所示。经过分析发现,表 3 的统计数据实际就是在完好种芽的前提下,种芽图像灰度最大值特征的分布情况。按照同样的统计方法,可以得到台糖 22 号完好和

破损种芽图像 5 个特征的分布统计情况,并绘制出分布曲线图,图 4 所示。

由图 4 可知,所有的分布曲线都是由水平直线和倾斜直线首尾连接而成,整体呈现锯齿形。水平直线表示相邻的横坐标所对应的特征值出现的次数相等,锯齿顶

点则是某个特征值出现的次数与左右相邻的特征值的次数都不相同,引起了特征值的分布波动。图 4 显示,仅有图 4c 和图 4d 的分布曲线较为稳定,主要由水平直线构成,其余特征分布曲线都属于上、下波动的锯齿型曲线。不管是水平直线还是锯齿型曲线,都与正态分布的“倒钟型”密度函数曲线相差较远,常见的二次函数、三次函数和指数函数等连续函数也无法与特征分布曲线匹配。分段函数可以完美拟合分布曲线,但是波动最频繁的图 4e 曲线,将被分成 34 段函数,分布最稳定的图 4c 曲线也需分成 10 段,如此多的分段数量,将对分类规则的制定和分类决策带来复杂的计算和判断。

表 3 台糖 22 号完好种芽图像灰度最大值分布统计

Table 3 Distribution statistics of intact bud image gray scale of Taitang 22

特征值 Eigenvalue	个数 Number	特征值 Eigenvalue	个数 Number	特征值 Eigenvalue	个数 Number
0.3725	2	0.6863	2	0.7725	1
0.8588	1	0.9373	3	0.4196	4
0.6902	1	0.7922	2	0.8745	2
0.9451	1	0.4863	1	0.7059	1
0.8000	2	0.8824	1	0.9490	2
0.4902	1	0.7176	2	0.8196	1
0.8863	1	0.9686	1	0.5216	1
0.7255	1	0.8235	2	0.9020	2
0.9765	1	0.5725	1	0.7373	2
0.8353	1	0.9084	1	0.9922	1
0.6157	1	0.7451	2	0.8392	2
0.9176	1	1.0000	35	0.6275	1
0.7529	1	0.8431	1	0.9216	1
0.6431	1	0.7569	1	0.8510	1
0.9255	1	0.6824	1	0.7608	1
0.8549	1	0.9294	3		

注:特征值表示台糖 22 号完好种芽灰度最大值的数值,个数则表示该特征值在表 1 中出现的次数。

Note: Eigenvalue indicates the numerical value of maximum gray value of Taitang 22, and number indicates the time of this eigenvalue in Table 1.

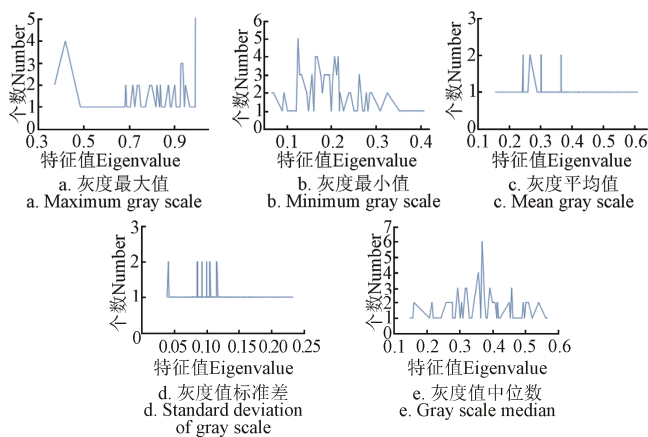


图 4 完好种芽图像特征分布情况曲线图

Fig.4 Distribution curve of intact bud image features

进一步观察图 4 发现,分布曲线中表示非均匀分布的倾斜直线在横轴上的投影占整个特征区间的比例较小,可忽略不计。根据连续分布概率计算公式(5),一个定点的概率为 0,则锯齿峰值点也可以略去,最后只剩下水平直线对应的特征值,且这些特征值出现的次数相等,可看成是均匀分布。

$$P(a \leq x \leq b) = \int_a^b f(x) = F(b) - F(a)。(5)$$

将表 3 中出现次数相同且连续的特征值组成的区间称为均匀分布区间。观察图 4 发现,组成均匀分布区间的特征次数有 1、2 和 3 次,但次数为 1 的均匀分布区间占到了 92.72%,为了简化均匀分布函数,将它们统一归为次数为 1 的均匀分布,然后进行统计,得到台糖 22 号完好种芽图像灰度最大值均匀分布区间宽度,并以同样的方法得到台糖 22 号完好与破损种芽图像 5 个特征的均匀分布区间宽度。根据表 2 的数据,计算得到台糖 22 号完好和破损种芽图像 5 个特征的分布区间宽度。将台糖 22 号完好和破损种芽图像 5 个特征的均匀分布区间宽度与特征分布区间宽度的数据对应相除,得到 5 个特征的均匀分布区间占整个特征区间的百分比,并计算两类种芽相应百分比的差值,结果如表 4 所示。

表 4 台糖 22 号种芽区域图像特征分布

Table 4 Bud image features' uniform distribution of Taitang 22

项目 Item	种类 Species	最大值 Max	最小值 Min	平均值 Mean	中位数 Std	标准差 Median
均匀分布区间宽度 Uniform distribution interval width	完好 Intact	0.3230	0.1136	0.4031	0.1841	0.1205
	破损 Damaged	0.2819	0.1137	0.3254	0.1481	0.1632
分布区间宽度 Distribution interval width	完好 Intact	0.6275	0.3411	0.4529	0.1941	0.4176
	破损 Damaged	0.5527	0.2510	0.3461	0.1590	0.3882
均匀分布比例 Uniform distribution proportion/%	完好 Intact	51.47	33.30	89.00	94.85	28.86
	破损 Damaged	51.00	45.30	94.02	93.14	42.04
	差值	0.47	12.00	5.02	1.71	13.18

表 4 数据显示,Min 和 Median 2 个特征的均匀分布比例较低,在 28.86%~45.30%之间,而且 2 类种芽相应的比例差值较大,达到了 12.00%和 13.1%。而 Max、Mean 和 Std 3 个特征的均匀分布比例较高,均超过了 50%,Mean 和 Std 的比例达到了 90%左右,且 2 类种芽相应的比例差值较小。虽然 Mean 的比例差值为 5.02%,但是相对其 91.51%的均匀分布平均比例来说较小,可以忽略。因此,将 Max、Mean 和 Std 3 个特征的分布简化为连续均匀分布,其概率密度函数<sup>[28]</sup>为

$$f(x) = \frac{1}{b-a}, a \leq x \leq b。(6)$$

由表 2 数据可得类条件概率密度函数  $p(x_j | \omega_i)$  为

$$\text{完好种芽 } p(x_j | \omega_1) = \frac{1}{B_j - A_j},(7)$$

$$\text{破损种芽 } p(x_j | \omega_2) = \frac{1}{b_j - a_j}。$$

类条件概率  $P(x_j | \omega_i)$  为

$$\text{完好种芽 } P(x_j | \omega_1) = \int_{A_j}^x \frac{1}{B_j - A_j} = \frac{x_j - A_j}{B_j - A_j},(8)$$

$$\text{破损种芽 } P(x_j | \omega_2) = \int_x^{b_j} \frac{1}{b_j - a_j} = \frac{b_j - x_j}{b_j - a_j}。$$

按照台糖 22 号种芽图像特征的分析方法, 将桂糖 42 号种芽图像的 Max、Min、Mean 和 Std 特征, 粤糖 60 号种芽图像的 Max、Mean、Std 和 Median 特征的分布情况简化为均匀分布。

#### 2.4 种芽完好性分类规则的制定

联立式 (3)、式 (4)、式 (8), 有

$$P(x|\omega_1) - P(x|\omega_2) \begin{cases} > 0, \text{则 } x \in \omega_1 \\ < 0, \text{则 } x \in \omega_2 \end{cases}, \quad (9)$$

台糖 22 号、桂糖 42 号和粤糖 60 号 3 个品种的种芽图像中能够简化为均匀分布的特征均包含 Max、Mean 和 Std, 而种芽完好性检测过程中并不需要检验该种芽属于哪一个品种的甘蔗, 因此, 可以将 Max、Mean 和 Std 3 个特征视为 3 个甘蔗品种的种芽完好性分类特征, 对分类规则进行简化。

则分类规则描述为: 任意输入某个种芽图像的 Max、Mean 和 Std 3 个特征的特征值  $x_1$ 、 $x_3$  和  $x_4$ , 根据式 (8) 得 6 个概率, 分别为  $P(x_1|\omega_1)$ 、 $P(x_1|\omega_2)$ 、 $P(x_3|\omega_1)$ 、 $P(x_3|\omega_2)$ 、 $P(x_4|\omega_1)$  和  $P(x_4|\omega_2)$ , 比较它们的大小, 找出

最大的概率, 根据此最大概率所对应的种类  $\omega$ , 判断该种芽图像的完好性。

综上所述, 分类规则, 可用式 (10) 表示

$$\text{若 } P(x_j|\omega_i) = \max_{i=1,2,j=1,3,4} \{P(x_j|\omega_i)\}, \text{ 则 } x \in \omega_i \text{ 类。} \quad (10)$$

### 3 种芽完好性分类试验及结果

#### 3.1 试验

为了验证本文提出的种芽完好性分类方法, 使用图 1 所示的种芽图像拍摄平台获取试验样本并进行预处理, 截取出  $40 \times 40$  像素的种芽图像, 得到了 3 个甘蔗品种各 215 个种芽试验样本, 其中包括 115 个完好种芽样本和 100 个破损种芽样本, 破损种芽样本中自然破损和人为模拟破损的种芽样本分别占 60% 和 40%。

试验在 Matlab2012b 环境下进行, 通过编写程序读入种芽试验样本图像, 提取出 Max、Mean、Std3 个灰度特征值, 根据式 (8) 计算  $P(x_j|\omega_i)$ , 按照式 (10) 的分类规则进行分类判断, 试验结果如表 5 所示。

表 5 种芽完好性分类准确率  
Table 5 Accuracy of bud integrity classification

品种 Varieties	完好种芽 Intact bud			破损种芽 Damaged bud			平均准确率 Average accuracy/%
	试验样本数 Test number	准确分类数 Exact number	分类准确率 Classification accuracy/%	试验样本数 Test number	准确分类数 Exact number	分类准确率 Classification accuracy/%	
台糖 22 号 Taitang 22	115	100	86.96	100	98	98.00	92.09
桂糖 42 号 Guitang 42	115	104	90.43	100	97	97.00	93.49
粤糖 60 号 Yuetang 60	115	104	90.43	100	96	96.00	93.02
平均值 Average value	115	102.67	89.27	100	97	97.00	92.87

#### 3.2 试验结果分析

表 5 显示, 3 个品种的种芽完好性分类试验都获得较高的分类准确率, 分别为 92.09%、93.49% 和 93.02%, 整体平均准确率为 92.87%。其中, 完好种芽分类准确率为 86.96%、90.43% 和 90.43%, 平均准确率为 89.27%; 破损种芽的分类准确率分别达到 98%、97% 和 96%, 平均准确率为 97%。

虽然破损种芽平均分类准确率已达 97%, 但仍有少数破损种芽识别失败, 原因如下: 1) 在获取甘蔗样本图像时采用的是人工将甘蔗放置在支架上, 手动控制甘蔗种芽位于甘蔗样本图像的中心位置的方式, 这种方式不能保证种芽位置的精度; 2) 采用普通的白炽灯补光, 而非专业光源, 会对图像的真实性产生一定影响; 3) 在截取  $40 \times 40$  像素的种芽区域时, 有一部分的非种芽区域也包含在截取后的种芽图像中, 这部分非种芽区域包含的信息会对种芽区域的灰度值信息产生一定的干扰。

完好种芽的平均分类准确率比破损种芽低 7.73%, 但本文研究的目的是为了尽可能多的找出破损种芽 (即破损种芽最小误判原则), 试验数据显示破损种芽分类准确率已达到 97%, 证明本文提出的分类方法是有效的, 能够代替人工挑选, 成功找出破损的种芽, 为自动剔除破损种芽提供关键信息。

### 4 结 论

以尽可能多的检测出破损种芽为目的, 提出以 Bayes 决策理论中最小错误率为准则的甘蔗种芽完好性分类方法, 将甘蔗种芽分为完好与破损两类。

1) 种芽图像分类特征的分布曲线无法用正态分布函数、二次函数、三次函数和指数函数等连续函数拟合, 虽然分段函数可以完美拟合分布曲线, 但由于分段数量过多, 会对分类规则的制定和分类决策带来复杂的计算和判断。进一步分析分布曲线, 发现将曲线中所占比例较小的倾斜直线与概率为 0 锯齿峰值忽略, 剩下的水平直线部分对应的特征值可简化为均匀分布。

2) 综合考虑灰度特征值的均匀分布区间占整个分布区间的百分比以及完好和破损两类种芽相应特征百分比的差值, 将台糖 22 号的 Max、Mean、和 Std 特征, 桂糖 42 号的 Max、Min、Mean 和 Std 特征和粤糖 60 号的 Max、Mean、Std 和 Median 特征的分布情况简化为均匀分布, 并作为各自甘蔗品种的有效分类特征。取 3 个品种共有的 Max、Mean 和 Std3 个有效分类特征, 将其作为 3 个甘蔗品种的种芽完好性分类特征。

3) 试验表明: 破损种芽分类平均准确率为 97%, 完



好种芽分类平均准确率为 89.27%，总体平均准确率为 92.87%，证明本文提出的分类方法有效。

本文的研究是以台糖 22 号、桂糖 42 号和粤糖 60 号为对象，但本文提出的种芽完好性分类方法也适用于其他品种的甘蔗，可为甘蔗种植机械化过程中自动剔除破损种芽提供关键信号。

#### [参 考 文 献]

- [1] 国家统计局. 2014 中国统计年鉴[M]. 北京: 中国统计出版社, 2014.
- [2] 王静, 郭教业. 甘蔗栽培管理技术[J]. 农技服务, 2013, 30(4): 321—322.
- [3] 杨荣芬. 甘蔗种植技术要点分析与应用[J]. 北京农业, 2014(18): 46—47.
- [4] 张建华, 孔繁涛, 李哲敏, 等. 基于最优二叉树支持向量机的蜜柚叶部病害识别[J]. 农业工程学报, 2014, 30(19): 222—231.  
Zhang Jianhua, Kong Fantao, Li Zhemin, et al. Recognition of honey pomelo leaf diseases based on optimal binary tree support vector machine[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2014, 30(19): 222—231. (in Chinese with English abstract)
- [5] 刘伟, 刘长虹, 郑磊. 基于支持向量机的多光谱成像稻谷品种鉴别[J]. 农业工程学报, 2014, 30(10): 145—151.  
Liu Wei, Liu Changhong, Zheng Lei. Discrimination in varieties of rice seeds with multispectral imaging using support vector machine[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2014, 30(10): 145—151. (in Chinese with English abstract)
- [6] 贾渊, 李振江, 彭增起. 结合 LLE 流形学习和支持向量机的猪肉颜色分级[J]. 农业工程学报, 2012, 28(9): 147—152.  
Jia Yuan, Li Zhenjiang, Peng Zengqi. Pork color grading based on LLE manifold learning and support vector machine[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2012, 28(9): 147—152. (in Chinese with English abstract)
- [7] 吴一全, 殷骏, 戴一冕, 等. 基于蜂群优化多核支持向量机的淡水鱼种类识别[J]. 农业工程学报, 2014, 30(16): 312—319.  
Wu Yiquan, Yin Jun, Dai Yimian, et al. Identification method of freshwater fish species using multi-kernel support vector machine with bee colony optimization[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2014, 30(16): 312—319. (in Chinese with English abstract)
- [8] 宁井铭, 宛晓春, 张正竹, 等. 近红外光谱技术结合人工神经网络判别普洱茶发酵程度[J]. 农业工程学报, 2013, 29(11): 255—260.  
Ning Jingming, Wan Xiaochun, Zhang Zhengzhu, et al. Discriminating fermentation degree of Pu'er tea based on NIR spectroscopy and artificial neural network[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2013, 29(11): 255—260. (in Chinese with English abstract)
- [9] 罗丽萍, 赵占锋, 戴喜末, 等. 表面解吸常压化学电离质谱结合人工神经网络鉴别新陈莲子[J]. 农业工程学报, 2013, 29(7): 261—266.  
Luo Liping, Zhao Zhanfeng, Dai Ximo, et al. Surface desorption atmospheric pressure chemical ionization mass spectrometry for identification of lotus seeds freshness based on PCA and BP-ANN[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2013, 29(7): 261—266. (in Chinese with English abstract)
- [10] 马茵驰, 阎广建, 丁文, 等. 基于人工神经网络方法的冬小麦叶面积指数反演[J]. 农业工程学报, 2009, 25(12): 187—192.  
Ma Yinchu, Yan Guangjian, Ding Wen, et al. Leaf area index retrieval of winter wheat using artificial neural network[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2009, 25(12): 187—192. (in Chinese with English abstract)
- [11] 翟治芬, 徐哲, 周新群, 等. 基于朴素贝叶斯分类器的棉花盲椿象危害等级识别[J]. 农业工程学报, 2015, 31(1): 204—211.  
Zhai Zhifen, Xu Zhe, Zhou Xinqun, et al. Recognition of hazard grade for cotton blind stinkbug based on Naive Bayesian classifier[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2015, 31(1): 204—211. (in Chinese with English abstract)
- [12] Chih-Ying Gwo, Chia-Hung Wei, Yue Li. Rotary matching of edge features for leaf recognition[J]. Computers and Electronics in Agriculture, 2013, 91(2): 124—134.
- [13] Pablo M Granitto, Hugo D Navone, Pablo F Verdes, et al. Weed seeds identification by machine vision[J]. Computers and Electronics in Agriculture, 2002, 33(2): 91—103.
- [14] 包晓敏, 汪亚明. 基于最小错误率贝叶斯决策的苹果图像分割[J]. 农业工程学报, 2006, 22(5): 122—124.  
Bao Xiaomin, Wang Yaming. Apple image segmentation based on the minimum error Bayes decision[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2006, 22(5): 122—124. (in Chinese with English abstract)
- [15] 黄亦其, 乔曦, 唐书喜, 等. 基于 Matlab 的甘蔗茎节特征分布定位与试验[J]. 农业机械学报, 2013, 44(10): 93—97.  
Huang Yiqi, Qiao Xi, Tang Shuxi, et al. Localization and test of characteristics distribution for sugarcane internode based on Matlab[J]. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(10): 93—97. (in Chinese with English abstract)
- [16] 陆尚平, 文友先, 葛维, 等. 基于机器视觉的甘蔗茎节特征提取与识别[J]. 农业机械学报, 2010, 41(10): 190—194.  
Lu Shangping, Wen Youxian, Ge Wei, et al. Recognition and features extraction of sugarcane nodes based on machine vision[J]. Transactions of the Chinese Society for Agricultural Machinery, 2010, 41(10): 190—194. (in Chinese with English abstract)
- [17] 广西大学. 甘蔗切割防伤芽系统: 201210123151.7[P]. 2013-03-20.
- [18] 乔曦. 基于计算机视觉的甘蔗种茎切割防伤芽系统的研究[D]. 南宁: 广西大学, 2013.  
Qiao Xi. System of Sugarcane Cutting Device to Protect the Bud Based on Computer Vision[D]. Nanning: Guangxi University, 2013. (in Chinese with English abstract)
- [19] 唐书喜, 罗昭宇, 黄亦其, 等. 甘蔗种切割防伤芽装置设计[J]. 农业开发与装备, 2011(5): 12—14.
- [20] 广西大学. 一种基于电阻式应变片的甘蔗种芽计数系统: 201320816269.8[P]. 2014-07-09.

- [21] 黄亦其, 王小波, 李晓峰, 等. 基于电阻式应变片的甘蔗种切割防伤芽系统[J]. 农机化研究, 2015, 37(2): 214—217.  
Huang Yiqi, Wang Xiaobo, Li Xiaofeng, et al. A sugarcane bud cutting injury prevention system based on the resistance strain gauge[J]. Journal of Agricultural Mechanization Research, 2015, 37(2): 214—217.(in Chinese with English abstract)
- [22] 黄亦其, 王小波, 尹凯, 等. 基于感应计数的甘蔗种切割防伤芽系统的设计与试验[J]. 农业工程学报, 2015, 31(18): 41—47.  
Huang Yiqi, Wang Xiaobo, Yin Kai, et al. Design and experiments of buds-injury-prevention system based on induction-counting in sugarcane-seeds cutting[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2015, 31(18): 41—47. (in Chinese with English abstract)
- [23] 陆尚平. 基于机器视觉的甘蔗茎节识别与蔗芽检测研究[D]. 武汉: 华中农业大学, 2011.  
Lu Shangping. Research on Sugarcane Internodes and Sugarcane Buds Identification Based on Machine Vision[D]. Wuhan: Huazhong University of Science and Technology, 2011.(in Chinese with English abstract)
- [24] 广西大学. 甘蔗种芽完好性检测系统及方法: 201410323245.8[P]. 2014-07-08.
- [25] Han J, Kamber M, Pei J. Data Mining: Concepts and Techniques[M]. Morgan Kaufmann, 2006.
- [26] 朱晓丹. 朴素贝叶斯分类模型的改进研究[D]. 厦门: 厦门大学, 2014.  
Zhu Xiaodan. Research on Improving Naive Bayes Classification Model[D]. Xiamen: Xiamen University, 2014. (in Chinese with English abstract)
- [27] 李静梅, 孙丽华, 张巧荣, 等. 一种文本处理中的朴素贝叶斯分类器[J]. 哈尔滨工程大学学报, 2003, 24(1): 71—74.  
Li Jingmei, Sun Lihua, Zhang Qiaorong, et al. Application of native Bayes classifier to text classification[J]. 2003, 24(1): 71—74. (in Chinese with English abstract)
- [28] 盛骤, 谢式千, 潘承毅. 概率论与数理统计[M]. 北京: 高等教育出版社, 2001: 23—24.

## Detection and experiment of sugarcane buds integrity based on Bayes decision

Huang Yiqi, Yin Kai, Huang Meizhang, Wang Xiaobo, Luo Zhaoyu  
(College of Mechanical Engineering, Guangxi University, Nanning 530004, China)

**Abstract:** Guangxi, a province in China, contributes more than 60% of China's sugarcane production per year. The sprouting rate of newly planted sugarcane is directly affected by viability of sugarcane buds. In sugarcane mechanized production, the detection of sugarcane bud viability for planting is still an issue and there is no a good classification method for differentiating viable sugarcane buds from the poor ones. To solve the problem, we proposed a classification method based on the minimum error rate of Bayes decision to detect the damaged buds as accurate as possible. Sugarcane cultivars of Taitang 22, Guitang 42 and Yuetang 60 were chosen for this study because they were representative cultivars of sugarcane in the area. We obtained sample images of sugarcane via image acquisition equipment, cut out the effective bud region images with computer vision technology, and extracted five features and eigenvalues including maximum gray scale. Then, a sample database was established, in which the image eigenvalues of the three sugarcane breeds was contained. Based on the data, prior probabilities of intact buds and damaged ones were calculated, from which eigenvalues' statistics of bud region images were obtained, and curve distribution diagrams of eigenvalues were described. By analyzing these statistics and diagrams, we found that all the distribution curves were connected end to end by horizontal lines and inclinable lines like zigzags. So these curves could not be fitted by normal distribution function, quadratic function, cubic function or exponential functions etc. Piecewise function could fit the distribution curve perfectly, but it would make the process of calculation and judgment more complicated when setting classification rules because of the excessive number of segments. Further observation showed that portions of small-proportion inclined lines and the saw tooth peak whose probability was zero could both be neglected. Then, the corresponding eigenvalues of the rest horizontal lines could be simplified to a uniform distribution. Calculating the even-distributed interval width and the entire interval width of the eigenvalues based on the database, the percentage between the two interval widths and the differences of percentage of the corresponding eigenvalues between the intact buds and damaged ones could be obtained. Using the percentage and the difference, even-distributed features could be confirmed. At the meantime, the conditional probability density was calculated by the probability density function of uniform distribution. Three features, including maximum gray scale, average gray scale and standard deviation of gray scale, which were the same even-distributed features of the three sugarcane cultivars, were chosen as the final classification features to simplify the classification rules. According to the full probability and Bayes formula, the prior probability and the conditional probability density of the final classification features were transformed to posterior probability, and then it could be classified by comparing the value of posterior probability. Matlab 2012b could be used to distinguish whether the bud was intact or not. The result of experiments showed that bud integrity classification accuracy of the three sugarcane cultivars were 92.09%, 93.49% and 93.02%, respectively. And the classification accuracy rate of damaged species had reached 98%, 97% and 96%, respectively. It proved that this classification method was feasible, which meant that it could detect damaged buds basically and provided a signal for eliminating the damaged buds automatically.

**Keywords:** nondestructive examination; sugar cane; decision making; bud; integrity