基于机器视觉的花生种子外观品质检测与分类方法研究

马佳佳,王克强*,郑奕雄,蔡 肯,林钦永 (仲恺农业工程学院,广东广州 510225)

摘要 [目的]花生种子的有效分类是花生产业中选育良种的重要环节,为降低花生产业对人工的依赖程度,简化选种过程,提出了一种基于机器视觉的花生种子外观品质检测与分类方法。[方法]在相同环境下采集单粒花生种子图像,建立花生种子对象在图像中像素数与其实际质量的回归模型,以花生种子尺寸和外观颜色作为主要特征,采用支持向量机分类模型完成分类任务。[结果]使用该方法完成12个类别的分类,对批量花生种子的分类准确率达86%,符合实际生产中花生种子初步分类要求。[结论]该方法对花生种子图像代表性特征的选择和识别样本的简化使得分类系统更符合实际生产需要,对同品种花生种子的不同品质分类以及不同品种花生种子的直接分类有着积极意义。

关键词 花生种子;机器视觉;图像识别;支持向量机
中图分类号 S126 **文献标识码** A
文章编号 0517-6611(2021)10-0225-03
doi:10.3969/j.issn.0517-6611.2021.10.057
开放科学(资源服务)标识码(OSID):

Research on the Appearance Quality Detection and Classification of Peanut Seeds Based on Machine Vision

MA Jia-jia, WANG Ke-qiang, ZHENG Yi-xiong et al (Zhongkai University of Agriculture and Engineering, Guangzhou, Guangdong 510225)

Abstract [Objective]The effective classification of peanut seeds was an important link in peanut breeding. In order to reduce the dependence of peanut industry on labor and simplify the seed selection process, a method of peanut seed appearance quality detection and classification based on machine vision was proposed. [Method] The image of single peanut seed was collected in the same environment, and the regression model between the image prime number and the actual quality of peanut seed object was established. The classification task was completed by using support vector machine (SVM) classification model with the main characteristics of peanut seed size and appearance color. [Result]The classification accuracy of 12 categories was 86%, which met the preliminary classification requirements of peanut seeds in actual production. [Conclusion]The selection of representative features of peanut seed image and the simplification of recognition samples make the classification system more in line with the actual production needs. It has positive significance for different quality classification of the same peanut seed and direct classification of different peanut seeds.

Key words Peanut seeds;Machine vision;Image recognition;Support vector machine

花生是中国产量丰富、食用广泛的一种坚果,同时也是中国为数不多的具有强劲国际竞争力的出口创汇农产品[1]。花生良种的选择与培育是花生种植的关键,因此花生种子的精选分类也是我国花生高产稳产的首要保证。分类是对农作物果实的大小、形状、颜色、营养物质含量、表面缺陷等性状进行检测和识别后,根据行业标准划分类别的操作过程[2]。通过机器视觉对花生种子进行有效分类有利于花生种植中选育良种,降低产业对人工的依赖程度,提高相关产业竞争力,对我国花生产业升级也具有重要意义。机器视觉技术可以无损、快速地获取样本表面特征信息,随着计算机技术和信息科学的不断发展,国内外机器视觉水平也不断到达新的高度。目前机器视觉已广泛应用于多个领域,在农业中可用于路径导航、杂草识别以及各种瓜果蔬菜的检测分级等[3-8]。基于机器视觉的分类方法具有检测精度高、效率快、能实现无损操作等优势,同时排除了人工主观因素的干扰。基于机器视觉技术,笔者研究了一种花生种子外观品质检测与分类方法,以花生种子图像的对象像素个数和颜色作为主要特征实现对花生种子的初步分类。需要说明的是,利用机器视觉进行初步分类只是花生种子选育过程中的第一步,后续还有许多问题需要通过其他技术方法来研究、解决,目前相关工作正逐步开展。笔者在相同环境下采集单粒花生种子图像,建立花生种子对象在图像中像素数与其实际质量的回归模型,以花生种子尺寸和外观颜色作为主要特征,采用支持向量机分类模型完成分类任务。

1 图像采集与预处理

1.1 试验准备 选择贺油133X品种花生作为试验材料。在同一杂交后代中随机选取脱壳后花生种子共400颗作为试验对象,使用电子天平精确测量得每颗花生种子的质量,同时经过专家辨别得到花生种子颜色,以质量与颜色作为分类标准先确定该花生种子的实际分类类别。随后进行图像采集工作,在自制简易暗箱内放置固定光源,选用HUAWEI TNY-AL00摄像头(2 400万像素)作为图像采集设备,拍摄时将摄像头固定在水平云台上,使用微距模式(焦距4 mm)进行拍摄,共采集400副花生种子原始图像。采集到的花生种子图像为大小3 456×3 456(单位:像素)的JPG彩色图像,用MATLAB软件编程实现图像处理,采用SPSS软件进行数据回归处理。

1.2 图像预处理 采集到的图片分辨率过大,后续处理可能花费很多时间。在预处理中,首先选择截取花生目标部分,将图片分辨率降低到864×864(单位:像素)。由于识别目标大部分为粉、紫(紫黑)、红等颜色,需要通过转换公式将RGB颜色空间3个通道加权对图像进行灰度化处理,使得图像中前后景差异更为明显,同时消除图像中小块的阴影面积和各种噪音点,实现后续稳定的特征识别效果。采用转换公

基金项目 广东省科技计划项目(2017A040405056);广东省研究生教育创新计划项目(KA200192346)。
作者简介 马佳佳(1996—),女,广东汕头人,硕士研究生,研究方向:机器视觉。*通信作者,教授,硕士,从事人工智能研究。
收稿日期 2020-10-13

式(1)将预处理后的图像转为灰度图像[9]:

$$Gray=0.3r+0.59g+0.11b \quad (1)$$

式中,r、g、b 分别表示预处理后彩色图像的 R、G、B 3 个通道灰度值;Gray 为生成灰度图像对应灰度值。采集到的原始图像如图 1a 所示,灰度化处理后的图像如图 1b 所示。

接着对灰度图像进行滤波操作,分别选用 5×5 的中值滤波模板和高斯滤波模板消除图像中的随机噪声,得到质量相对增强的图像,滤波后图像如图 1c 所示。经过滤波操作后,图像中的噪点减少,同时能够较好保留对象边缘信息。采用 OTSU 大津阈值分割算法实现图像二值化,取 1 个最优阈值把图像分为前景与背景,2 部分的类间方差越大,说明差别越大,从而能有效分割图像,具有实时性好、计算量小的特点[10]。再根据误分割区域面积小于花生种子区域的特性,采用面积阈值剔除误分割区域,结合孔洞填充算法,完成花生种子对象与背景的分割,形成的二值图像如图 1d 所示。

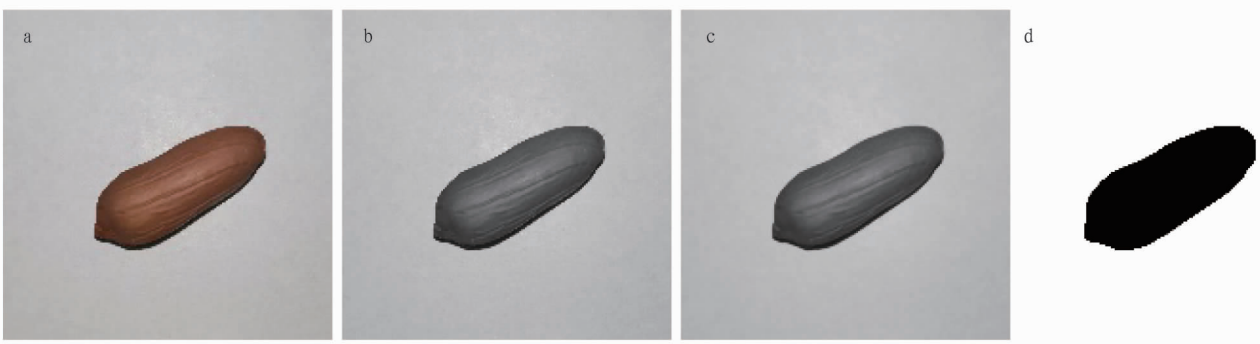


图 1 花生种子样本图像

Fig.1 Image of peanut seed sample

2 特征提取

2.1 尺寸特征提取

2.1.1 边缘检测。花生种子外观特征中的几何特征数据均与其图像边界有关[7],采用 Sobel 边缘检测算法对预处理后的花生种子图像进行边界提取,可以快速得到花生种子光滑、连续的边缘特征,结果如图 2 所示。Sobel 算子公式如式(2)、(3)所示:

$$Gx=[f(x+1,y-1)+2\times f(x+1,y)+f(x+1,y+1)]-[f(x-1,y-1]+2\times f(x-1,y)+f(x-1,y+1)] \quad (2)$$

$$Gy=[f(x-1,y-1)+2\times f(x,y-1)+f(x+1,y-1)]+[f(x-1,y+1]+2\times f(x,y+1)+f(x+1,y+1)] \quad (3)$$

式中,G_x 及 G_y 分别代表经横向及纵向边缘检测的图像灰度值,$f(a,b)$表示图像(a,b)点的灰度值。

图 2 边缘检测结果

Fig.2 Edge detection results

2.1.2 花生质量与图像中像素个数的关系。花生种子的质量和尺寸都是其重要的品质特征,要利用机器视觉的方法对花生种子样本的质量有一个粗略的估计,需要建立单颗花生种子质量与其采集图像中表现为对象像素个数,即样本尺寸大小的相关性模型,实现在检测过程中可以直接通过图像处理得到花生种子质量的相关数据以完成分类任务。

以此思路设计预试验:另随机选取 400 颗同一品种花生种子进行编号,用电子天平精确测量单颗质量并记录为数据 1;然后通过图像处理计算得每幅图像中花生种子对象的像素个数,并记录为数据 2;以数据 1、2 作为数据基础,用 SPSS 软件建立单颗花生种子质量(因变量)与图像中对象尺寸(自变量)之间的二次回归方程,记为

$$y=-0.046+4.614\times10^{-6}\times x+2.554\times10^{-11}\times x^2 \quad (4)$$

式中,y 为单颗花生种子的质量,单位为 g;x 为花生种子图像中花生对象所占的像素个数,单位为个。模型参数中 R 方值为 0.936,表示该模型能较好拟合数据,模型曲线拟合图如图 3 所示。

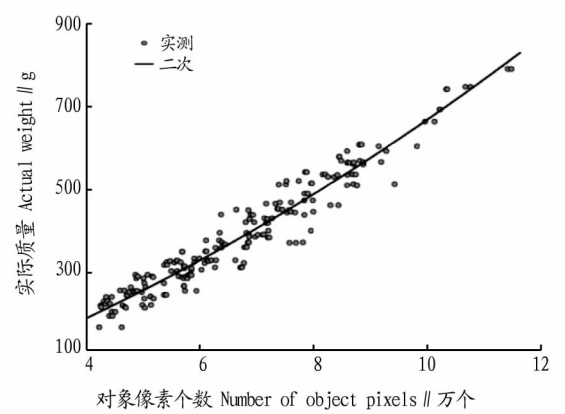


图 3 质量-像素回归模型曲线拟合

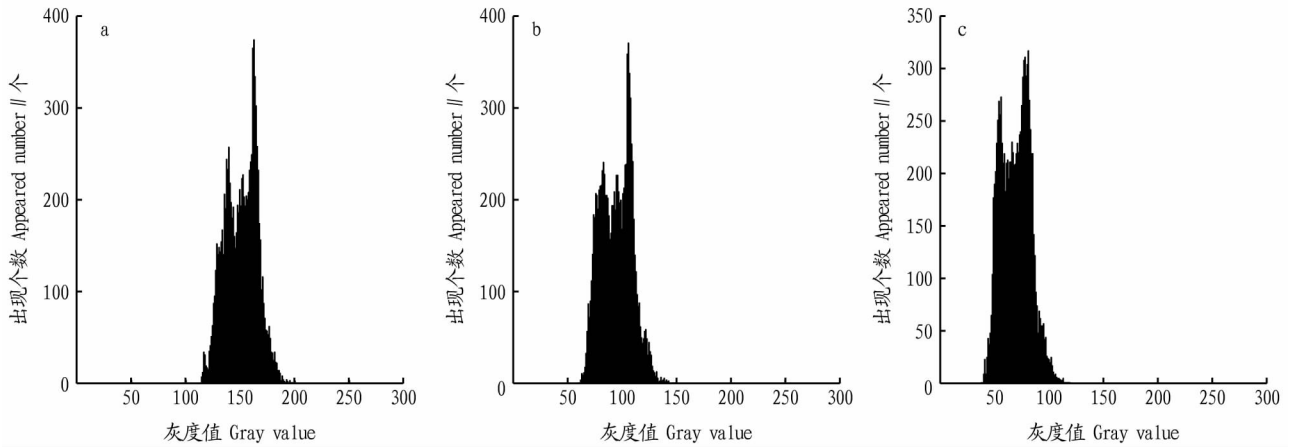
Fig.3 Curve fitting of quality-pixel regression model

在相关标准中已明确说明在花生的生产与销售中通常以种子百仁重表示花生种子大小,大体可分为大粒种、中粒

种、小粒种 3 个类别:百仁重在 80 g 以上为大粒种,50~80 g 为中粒种,50 g 以下为小粒种[11]。结合上述回归模型可以计算得出,花生种子图像中花生对象像素个数大于 112 879 记该花生种子对象为大粒种,在 81 669~112 879 内记为中粒种,小于 81 669 个像素单位记为小粒种。以此,基本实现从图像中提取得花生种子质量特征。

2.2　颜色特征提取　为了实现花生种子外观颜色分类,对降低了分辨率的原始图像(分辨率为 864×864,单位为像素)进行颜色特征提取,选取图像的中心部分像素,这里选择由(450,450)与(500,500)2 点围成的正方形像素区域进行颜色识别,识别结果为粉色(pink)、紫色(purple)、红色(red)和其他(others)。

在相同的图像采集条件下,同一颜色类别的花生种子图像 RGB 值较稳定,可对不同颜色的花生种子进行大数量的抽样,以确定每类颜色的具体 RGB 值取值范围,根据这一取值范围确定 3 个通道的灰度值阈值,示例图像的 RGB 3 通道灰度直方图结果如图 4 所示。处理待分类的花生种子图像时,提取其 RGB 3 通道的灰度值与阈值比较大小即可快速提取样本颜色特征,在保证准确率的同时通过自定义函数的方式进行颜色识别也便于后续对试验再做改进。



注:a.红色分量;b.绿色分量;c.蓝色分量

Note:a.Red component;b.Green component;c.Blue component

图 4　RGB 灰度直方图

Fig.4　RGB gray histogram

3　分类模型及性能测试

该研究所用的分类模型是支持向量机(Support Vector Machine,SVM)模型。该模型是一种基于统计学习理论的分类方法,通过监督学习方式对数据进行二元分类,其决策边界是对学习样本求解的最大边距超平面,具有相对优良的性能指标。由于待分样本集中的大部分样本不是支持向量,移去或者减少这些样本对分类结果没有很大影响, SVM 模型在小样本、非线性模式识别中具有一定的优势[12]。采用径向基函数核(radial basis function kernel, RBF kernel)作为 SVM 分类模型的核函数,其对应的映射函数将样本空间映射至无限维空间,是常用的 1 种核函数,其解析式如式(5)所示,其中核函数的惩罚参数 c 默认为 1,参数 gamma(g)默认为特征数的倒数。

$$k(X_1,X_2)=\exp\left(-\frac{\|X_1-X_2\|^2}{2\sigma^2}\right) \quad (5)$$

将从图像中提取到的花生种子质量特征用 a1、a2、a3 分别代表大粒种、中粒种、小粒种 3 类,颜色特征用 b1、b2、b3、b4 分别代表粉色、紫色、红色及其他颜色共 4 类,同时用每颗花生种子实际分类类别作为该样本的标签,以此进行模型训练,分类类别总数为 12。

采用交叉验证法进行验证试验,400 个样本按照比例3∶1随机划分为训练集和测试集,在随机选择的条件下,每次测试的结果不同,可达到验证目的。多次试验得到该模型对测试集的分类准确率达 86%,能较好完成花生种子批量初步分类,绘制试验结果图像如图 5 所示,坐标系中横坐标表示随机划分为测试集的 100 个样本,纵坐标表示花生种子分类类别代码;星号代表样本的实际类别即样本标签,圆圈代表模型预测得到的结果,2 点重合则表示模型预测准确,表 1 说明了样本分类类别代码含义。

4　结论

植物器官外观表现是个体细胞分裂、生长、分化与代谢相互作用的最终形态体现,花生种子外观特征与其内部品质之间密不可分,寻求其中的联系有利于花生种子选育良种。

该研究建立了一个基于机器视觉的花生种子外观品质检测与分类系统,通过图像处理提取单粒花生种子图像中 2 个代表性的特征参数,采用 SVM 分类模型完成 12 个类别的分类,试验分类准确率达 86%。该研究对代表性特征的选择和识别样本的简化使得分类系统更符合实际生产需要,对同品种花生种子的不同品质分类以及不同品种花生种子的直接分类有积极意义。但需要说明的是,该研究仅用单一花生品种对分类方法进行验证,若应用于其他品种花生种子的分类上,则需对算法中的必要参数及计算质量与图像像素个数的回归方程参数进行优化,才能保证较高的分级准确率。

(下转第 231 页)

产与需求无缝对接,助推共享农业发展[12]。

其二,使用大数据方便消费者监督农作物种植、收获和配送等全过程。企业在大数据平台上提前向消费者展示农作物在每个阶段的质量标准,将农作物的全部信息上传到大数据平台,消费者通过可视化方式了解农作物种植过程和状况是否达到标准。全方位地了解农作物信息,增强消费者对企业的信任,也便于消费者监督企业的生产经营行为。

其三,使用大数据打破了企业发展的地域限制。CSA 社会企业随着市场经济的发展,经营地址已不再局限于当地。如分享收获农场的五常大米供应地在黑龙江五常市。利用大数据可以打破地域的界限,将生产资源和物流资源优化配置,减少企业成本,增加会员便利。

其四,使用大数据促进资源合理流动。通过大数据支持企业发展,可让政府、互联网企业、培训和支持机构建立和健全信任关系,实现各种资源的合理流动。与此同时,社会企业还可以与观光农业、休闲农业行业共享数据,充分调动各方积极性,促进资源有序流动,提高效益,减少成本损耗,实现农村、农业、农民之间的良性互动,提升其竞争合力。

参考文献

[1] 周飞跃,勾竞懿,梅灵.国内外社区支持农业(CSA)体系的比较分析[J].农业经济问题,2018(7):78-87.

[2] SCHNELL S M.Food with a farmer's face:Community-supported agriculture in the United States[J].Geographical review,2007,97(4):550-564.

[3] BASU S,O'NEILL J,SAYER E,et al.Population health impact and cost-effectiveness of community-supported agriculture among low-income US adults:A microsimulation analysis[J].American journal of public health,2020,110(1):119-126.

[4] 石嫣,程存旺,雷鹏,等.生态型都市农业发展与城市中等收入群体兴起相关性分析:基于"小毛驴市民农园"社区支持农业(CSA)运作的参与式研究[J].贵州社会科学,2011(2):55-60.

[5] BLÄTTEL-MINK B,BODDENBERG M,GUNKEL L,et al.Beyond the market—New practices of supply in times of crisis:The example community-supported agriculture[J].International journal of consumer studies,2017,41(4):415-421.

[6] PAUL M.Community-supported agriculture in the United States:Social,ecological,and economic benefits to farming[J].Journal of agratian change,2019,19(1):162-180.

[7] 石嫣.中国农业需要可持续发展——对《四千年农夫》的思考[J].中国合作经济,2012(7):62-63.

[8] 郭于华.安全食品的生产与社会信任的重建:CSA 社会企业的探索之路[J].学海,2020(3):11-20.

[9] 张卫东,林菁璐.我国社区支持农业的发展策略研究[J].中州学刊,2017(10):38-42.

[10] SPROUL T W,KROPP J D.A general equilibrium theory of contracts in community supported agriculture[J].American journal of agriculture economics,2015,97(5):1345-1359.

[11] 石嫣.发展互联网生态农业需要构建信任机制[J].中国农村科技,2014(7):15.

[12] 吴瑞兵."互联网+现代农业"助推精准扶贫的模式研究:基于社区支持农业视角的分析[J].价格理论与实践,2018(6):134-137.

(上接第 227 页)

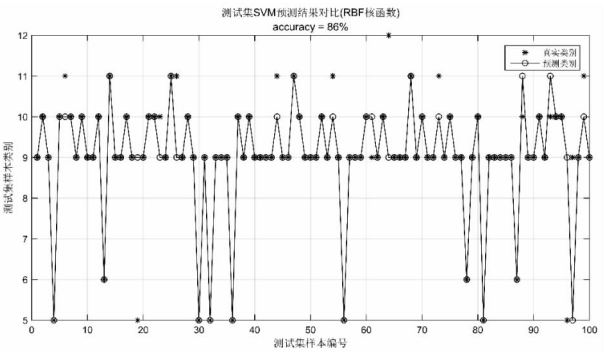


图 5　花生种子分类结果

Fig.5　Classification results of peanut seeds

表 1　花生种子分类类别代码含义

Table 1　Meaning of peanut seed classification code

颜色 Color	大 Big	中 Medium	小 Small
粉 Pink	1	5	9
紫 Purple	2	6	10
红 Red	3	7	11
其他 Others	4	8	12

参考文献

[1] 陈迪,吕杰.中国花生产业国际竞争力评价研究[J].财经问题研究,2013(1):35-39.

[2] 刘妍,周新奇,俞晓峰,等.无损检测技术在果蔬品质检测中的应用研究进展[J].浙江大学学报(农业与生命科学版),2020,46(1):27-37.

[3] 颜秉忠,王晓玲.基于计算机视觉技术大枣品质检测分级的研究[J].农机化研究,2018,40(8):232-235,268.

[4] 关卓怀,陈科尹,丁幼春,等.水稻收获作业视觉导航路径提取方法[J].农业机械学报,2020,51(1):19-28.

[5] 何东健,张庭锋,李邦国,等.除草机器人机器械视觉自动导航技术研究[J].安徽农业科学,2020,48(18):201-205.

[6] 彭明霞,夏俊芳,彭辉.融合 FPN 的 Faster R-CNN 复杂背景下棉田杂草高效识别方法[J].农业工程学报,2019,35(20):202-209.

[7] 颜秉忠.机器视觉技术在玉米苗期杂草识别中的应用[J].农机化研究,2018,40(3):212-216.

[8] 王红军,熊俊涛,黎邹邹,等.基于机器视觉图像特征参数的马铃薯质量和形状分级方法[J].农业工程学报,2016,32(8):272-277.

[9] 杜云,曹世佳,贾科进,等.基于图像处理的大蒜蒜种品质无损分级方法研究[J].河北工业科技,2018,35(5):317-321.

[10] 张婧茹,莫锦秋,曹家勇,等.一种鲁棒的室外金属标记物图像分割算法[J].机械设计与研究,2016,32(3):31-37.

[11] 国家粮食储备局无锡科学研究院.花生:GB/T 1532—2008[S].北京:中国标准出版社,2009.

[12] 秦丰,刘东霞,孙炳达,等.基于深度学习和支持向量机的 4 种苜蓿叶部病害图像识别[J].中国农业大学学报,2017,22(7):123-133.