

基于外部特征信息的番茄果实质量预测模型

何 微 牛智有 李晓金

华中农业大学工学院, 武汉 430070

摘要 为实现番茄果实质量的预测, 提出一种应用计算机视觉技术对番茄果实质量进行自动识别的方法。采用 Matlab 平台构建算法, 利用数学形态运算及图像局部性质运算等构成的识别算法, 对番茄样本图像进行预处理, 提取番茄图像投影面积、轮廓周长、最大内接圆直径和最小外接圆直径外部特征参数, 分别建立番茄果实质量与 4 个特征参数之间的一元线性、二阶多项式、幂指和多元线性预测模型。试验结果显示: 采用多元线性预测模型预测结果最佳, 其决定系数为 0.926 7, 标准差为 4.32; 利用检验样本对预测模型进行验证, 预测果实质量与实际果实质量的绝对误差均值为 3.260 g, 相对误差均值为 1.535%。研究结果表明, 基于计算机视觉技术外部特征信息的番茄果实质量预测方法是可行的。

关键词 番茄; 外部特征; 计算机视觉; 预测模型

中图分类号 TP 391.41; TS 255.35 **文献标识码** A **文章编号** 1000-2421(2013)06-0144-05

番茄果实的采收与分级主要依靠经验, 因缺乏现代无损分选技术, 极大地影响了番茄高端消费和出口贸易。鉴于此, 中国颁发了番茄质量等级标准 (GB8852-88)。该标准为番茄质量等级和规格判定提供了依据。目前, 番茄等级的确定主要采用人工评定的方法, 这种依靠人工的番茄等级评定方法存在明显缺陷: 一是缺乏客观的定量评价标准, 评定过程容易受到主观因素的干扰, 造成较大误差; 二是效率低, 难以满足工业化生产要求。因此, 采用现代化检测技术和方法判别番茄等级, 建立番茄等级的客观评价方法和基础理论, 以实现番茄质量的高效自动分级, 是当今番茄生产过程和流通领域中亟待解决的技术问题之一^[1]。笔者采用机器视觉技术, 研究番茄图像处理方法, 提取番茄有效的外部特征参数, 旨在分析果实质量与外部特征参数之间的相关性, 建立果实质量预测模型, 为番茄果实质量自动分级提供科学依据。

1 材料与方法

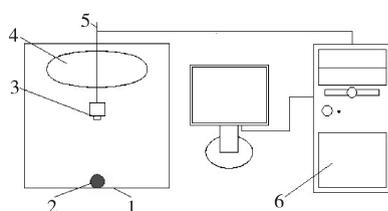
1.1 供试材料

供试样本为果蔬市场和农户种植收获后未经挑选的 134 个番茄果实, 主要品种为红宝石、华番 1 号

和精番 168。经过测定, 番茄果实质量为 58.1~341.0 g, 平均 135.5 g, 变异系数为 31.8%。根据试验要求对番茄果实进行人工分组和编号, 将供试样本分为 2 组, 其中 126 个样本用于建模, 其余 8 个样本用于对模型的检验。

1.2 图像采集系统

根据试验要求, 自制构建的番茄图像采集系统如图 1 所示。



1. 载物台 Stage; 2. 番茄 Tomato; 3. 电耦合装置 Charge couple device (CCD); 4. 环形光源 Circular light source; 5. 丝杆 Screw; 6. 计算机 Computer.

图 1 图像采集系统模型

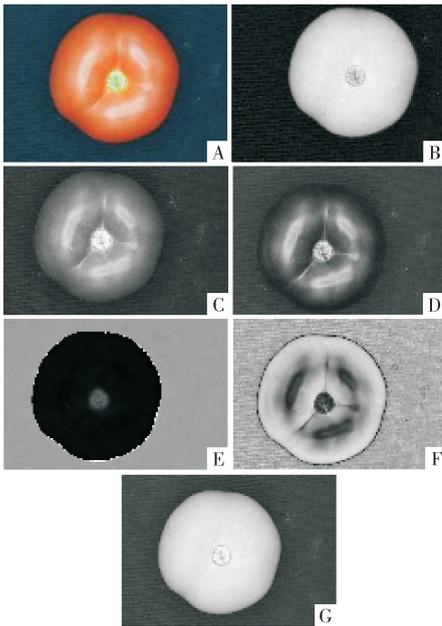
Fig. 1 Image acquisition system

图像采集系统主要包括硬件和软件 2 个部分。硬件部分主要有摄像头、光照箱、环形灯、图像采集卡和 PC 机。摄像头采用大恒公司生产的彩色 HV2000FC 型(带外触发), 配套使用 IEEE1394 接

口的图像采集卡；光照箱是定制的长 500 mm、宽 600 mm、高 600 mm 的铝合金箱，顶端有 1 个可以调节摄像头高度的丝杆；环形灯采用 32 W 的日光灯。软件部分主要包括图像采集控制程序和图像处理程序。图像的采集、预处理和特征提取均采用 MatlabR2008a 软件编程实现。图像采集部分主要完成图像大小、采集频率以及图像存储工作，图像采集在封闭的灯箱内进行，箱底部用黑色吸光布作为背景，分别采集番茄的正面、反面和侧面图像。图像处理程序主要用于对图像的颜色转换、图像分割、轮廓提取以及特征参数的提取。

1.3 番茄果实图像预处理

1) 图像色彩空间转换。采用图像采集系统所采集的样本图片是 RGB 空间图像，它们的物理意义很清楚，但并不适应人的视觉特点，因此，要把图像转换到更与人类感知接近的 HSV 模型。图像的后续处理都是基于 *H* 分量的图像。RGB 和 HSV 模型各部分如图 2 所示。



A. 原始图像 Original image; B. R 分量图像 R component image; C. G 分量图像 G component image; D. B 分量图像 B component image; E. H 分量图像 H component image; F. S 分量图像 S component image; G. V 分量图像 V component image.

图 2 RGB 及 HSV 各分量显示

Fig. 2 Display of component

2) 图像分割。阈值分割是一种简单有效的图像分割方法，该法是用一个或者几个阈值将图像的灰度级分为几部分，将隶属于同一部分的像素视为相

同的区域。阈值分割包括双峰法、迭代法、最大类间方差法等^[2-4]。

本试验采用最大类间方差法 (OSTU 阈值分割算法)。该算法的优点是计算简单、稳定有效、适用性比较强，根据图像自动寻找最佳阈值。

最大类间方差法是在判决分析或最小二乘法原理的基础上推导出来的，最大类间方差分割图像的计算方法如下^[5-7]：

一幅数字图像 $f(x, y)$ 中的像素按灰度级用阈值 T 分为 C_0 和 C_1 类，即

$$C_0 = f_1(x, y) | f_{\min} \leq f(x, y) \leq T$$

$$C_1 = f_2(x, y) | f_{\max} \geq f(x, y) > T$$

其中， f_{\min} 、 f_{\max} 分别为图像 $f(x, y)$ 中灰度的最小值和最大值。

设 N_i 是灰度值为 i ($f_{\min} \leq i \leq f_{\max}$) 的像素数，则图像 $f(x, y)$ 总的像素为 $N = \sum_i N_i$ ，各灰度级出现的概率为 $P(i) = N_i / N$ ，则 C_0 类出现的总概率为 $P_0 = \sum_{i=f_{\min}}^T P(i)$ ，均值为 $u_0 = \sum_{i=f_{\min}}^T iP(i) / P_0$ ； C_1 类出现的总概率为 $P_1 = \sum_{i=T+1}^{f_{\max}} P(i)$ ，均值为 $u_1 = \sum_{i=T+1}^{f_{\max}} iP(i) / P_1$ ，则图像 $f(x, y)$ 的均值为

$$u = \sum_{i=f_{\min}}^{f_{\max}} iP(i) = \sum_{i=f_{\min}}^T iP(i) + \sum_{i=T+1}^{f_{\max}} iP(i) = P_0 u_0 + P_1 u_1$$

定义两类的类间方差为

$$\sigma^2(T) = P_0(u - u_0)^2 + P_1(u - u_1)^2$$

最大类间方差把两类方差作为阈值选择的判决依据，认为最好的阈值 T^* 应该是使类间方差 $\sigma^2(T)$ 取得最大值时的阈值，即

$$T^* = T^* | \sigma^2(T^*) \geq \sigma^2(T), \forall T \in [f_{\min}, f_{\max}]$$

通过采用以上最大类间方差法，获得样本二值图像的最佳阈值为 77，其最佳阈值分割图像如图 3 所示。

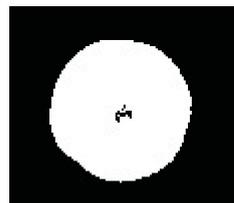


图 3 最佳阈值分割图像

Fig. 3 Segmentation image

3) 图像形态学处理。形态学图像处理的基本思想是用具有一定形态的结构元素去度量和提取图像中对应形状，以达到对图像分析和识别的目的^[8-9]。

为了得到番茄果实的外部轮廓,需要把番茄果实蒂处的像素腐蚀掉,以便后续的番茄果实外部特征参数的提取。采用形态学中的闭运算来填充目标内的细小空洞和裂缝、连接断开的临近目标,同时可以平滑图像,保持目标物尺寸大小不变。经闭运算处理后的番茄果实图像如图 4 所示。

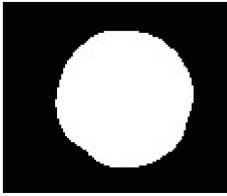


图 4 经闭运算处理后的图像

Fig. 4 Image after closing process

4) 轮廓提取。提取番茄果实的轮廓前,将小目标移除。对于这一批番茄果实图像移除二值图像中少于 50 个像素的连通元素,生成一个新的二值图像。利用图像分析和统计跟踪边界并用红色线条在二值图像区标记出轮廓线,其实例如图 5 所示。

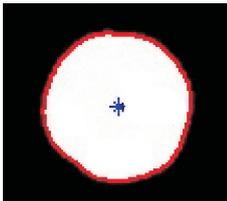


图 5 轮廓图

Fig. 5 Outline image

1.4 番茄果实外部特征参数提取

从样本轮廓图上提取番茄果实的外部图像特征信息,分析图像的局部性质,提取番茄果实的图像特征信息参数,统计标记区域内的像素点个数、标记边界的边界长度。外部特征信息参数主要包括图像投影面积 S 、轮廓周长 L 、最大内接圆直径 d_1 和最小外接圆直径 d_2 [10-13]。

由于提取的特征参数是基于图片本身信息得到的,为了得到番茄的实际参数,用已知参数的标准滤纸作为参照,在相同条件下进行图像处理,计算比例系数,通过统计值乘以比例系数便可得到番茄果实的实际特征参数值(表 1)。

由测量参数值和像素统计值的比值得到标准滤纸的标定结果,计算得到番茄果实直径系数为 0.009 7 cm,周长系数为 0.008 8 cm,面积系数为 $9.43 \times 10^{-5} \text{ cm}^2$ 。

表 1 标定结果

Table 1 Calibration results

项目 Items	直径/cm Diameter	周长/cm Perimeter	面积/cm ² Area
测量参数值 Measured values	12.0	39.2	122.7
像素统计值 Statistical values	1 287.3	4 440.9	1 301 525

1.5 数据处理

数据处理主要通过 Matlab 和 EXCEL 软件对数据进行统计分析。利用检验样本建立投影面积的面积 S 、轮廓周长 L 、最大内接圆直径 d_1 和最小外接圆直径 d_2 与番茄果实质量之间的关系及预测模型,以决定系数、标准差为评价指标,分析番茄果实质量与特征参数之间的相关性及预测模型的预测精度,并用检验样本对预测模型进行实证。

2 结果与分析

2.1 预测模型的建立

在实际生产过程中,由于在生产线上采用机器视觉对番茄果实进行实际图像采集时,不能保证摄像头对番茄果实拍照总是同一个面,所以在试验中对每个番茄果实样本进行包括正面、反面、侧面等不同部位 6 个角度的图像进行采集,提取每张图像的外部特征信息,在分析数据时选取同一组数据进行分析。采用这样的图像采集与数据分析是符合生产实际的。

通过以上图像采集于外部特征参数的提取与计算方法,选择番茄果实投影面积 S 、轮廓周长 L 、最大内接圆直径 d_1 和最小外接圆直径 d_2 ,以这 4 个参数作为建模的特征参数。

以番茄果实质量为目标函数,以 4 个特征参数为设计变量,分别建立一元线性、二阶多项式、幂指数及多元线性预测模型 [11-13]: 一元线性模型 $Y = A + BX$; 二阶多项式模型 $Y = A + BX_i + CX_i^2$; 幂指数模型 $Y = AX_i^B$; 多元线性模型 $Y = A_0 + \sum_{i=1}^4 A_i X_i$ 。式中 Y 为果实质量(g); $A, B, C, A_0, A_i (i=1, 2, 3, 4)$ 均为系数; $X_i (i=1, 2, 3, 4)$ 为特征参数。

分别采用以上 4 种模型对果实质量与番茄果实质量特征参数之间进行拟合,建立了基于外部特征参数的番茄果实质量预测模型,利用决定系数 R^2 、标准误差 SE 为评价指标,其预测统计结果如表 2 所示。

表 2 预测模型系数和预测统计结果

Table 2 Coefficient of prediction model and forecasting results

拟合曲线方程 Fitting curve equation	系数及评价指标 Coefficients and indicators	投影面积(S)/cm ² Projected area	轮廓周长(L)/cm Perimeter	最大内接圆直径(d ₁)/cm Maximum inscribed circle diameter	最小外接圆直径(d ₂)/cm Minimum inscribed circle diameter
Y=A+BX	A	6.057 1	12.095 5	4.272 1	4.271 9
	B	0.257 1	0.066 8	0.022 3	0.019 9
	R ²	0.916 5	0.920 1	0.922 2	0.891 7
	SE	5.28	4.95	4.75	7.41
Y=A+BX _i +CX _i ²	A	9.416 1	11.83	4.044 7	3.663 6
	B	0.211 2	0.082 1	0.025 4	0.028 2
	C	0.000 1	-0.000 5	-0.000 6	0.000 5
	R ²	0.918 0	0.922 4	0.923 1	0.899 4
Y=AX _i ^B	SE	5.18	5.01	4.89	6.94
	A	0.681 8	2.743 3	0.909 8	0.953 3
	B	0.834 8	0.427 3	0.462 9	0.408 2
	R ²	0.917 7	0.920 8	0.918 9	0.897 6
	SE	5.42	5.36	5.51	6.75

建立多元线性预测模型,得到 $Y=0.161 6S+32.208 8L-25.796 5d_1-48.021d_2-145.857 0$, 其中S为投影面积, L为轮廓周长, d₁为最大内接圆直径, d₂为最小外接圆直径, 相关系数为0.926 7, 标准差为4.32 g。

从表2中可以看出,采用一元线性模型,在4个特征参数中,番茄果实质量与最大内接圆直径有很好的相关性,其决定系数R²为0.922 2,标准差(SE)为4.74 g;采用二阶多项式模型,同样是番茄果实质量与最大内接圆直径有很好的相关性,其决定系数R²为0.923 1,标准差为4.89 g;采用幂指数模型,番茄果实质量与轮廓周长有很好的相关性,其决定系数R²为0.920 8,标准差为5.36 g。采用多元线性模型,预测模型的预测值与实际值之间的决定系数R²为0.926 7,标准差为4.32 g。

通过综合比较与分析可知,采用4个特征参数建立的多元线性模型相对于单个特征参数建立的其他3个模型而言,具有较高的决定系数和较小的标准差,所以选择以投影面积、轮廓周长、最大内接圆直径和最小外接圆直径为设计变量,建立的多元线性模型为最优果实质量预测模型。

2.2 预测模型的验证

利用构建的图像采集系统,采集8个番茄检验样本的图像,并对图像进行预处理,通过像素转换计算,获得8个检验样本的特征参数值和实际果实质量(表3)。将表3中8个检验样本的数据带入多元线性拟合模型,经过计算与验证,实际果实质量与预测果实质量之间的绝对误差均值为1.06 g,单个相对误差最大值为1.91%,最小值为0.02%,相对误差均值为0.59%(表4)。

表 3 检验样本实际特征参数值

Table 3 Characteristic parameters

样本号 Sample	S/cm ²	L/cm	d ₁ /cm	d ₂ /cm	果实质量/g Fruit mass
1	31.563 8	19.882 4	6.374	6.306 2	111.7
2	46.315 7	24.669 7	7.961 4	7.518 6	181.1
3	49.315 7	24.669 7	8.317 9	7.554 3	177.4
4	51.881 6	25.625 2	8.275 2	7.991 4	193.9
5	54.727 4	26.815 4	8.397 5	8.300 8	228.9
6	53.972 4	26.233 0	8.519 9	8.074 4	201.2
7	59.509 0	27.529 9	8.765 0	8.663 7	214.0
8	82.313 1	32.859 5	10.122 8	9.509 4	341.0

表 4 果实质量预测误差统计

Table 4 Prediction table of mass

样本号 Sample	实际质量/g Measuring mass	预测质量/g Predicted mass	绝对误差/g Absolute error	相对误差/% Relative error
1	111.7	111.9	0.209	0.187
2	181.1	176.8	4.315	2.383
3	177.4	178.0	0.640	0.361
4	193.9	193.2	0.737	0.380
5	228.9	218.7	10.194	4.453
6	201.2	205.2	4.011	1.993
7	214.0	218.4	4.443	2.076
8	341.0	339.5	1.531	0.449
均值 Average value			3.260	1.535

3 讨论

试验结果表明,采用多元线性拟合模型对番茄果实质量预测能取得比较理想和精确的结果,但预测值与实际值仍存在较小偏差。通过对误差较大的样本进行分析,可能有几个方面的原因:一是由于番茄空果的出现导致预测结果比实际测量值大;二是由于番茄果实的水分会随着时间流失,使其质量也会随之减轻,导致误差较大;三是由于番茄品种间的物性(如果实密度)差异,也能导致较大误差。为了提高预测模型的精度,可以选择品种多、物性差异

较大的番茄样本进行多次试验,以建立更精确的预测模型。

本研究通过运用数字图像处理技术和图像提取算法获取了番茄果实图像投影面积、轮廓周长、最大内接圆直径和最小外接圆直径,并以投影面积、轮廓周长、最大内接圆直径和最小外接圆直径为外部特征参数,分别建立了一元线性、二阶多项式、幂指数及多元线性果实质量预测模型。通过比较分析,结果表明以多元线性预测模型效果最佳,其决定系数为 0.926 7、标准差为 4.32 g。运用多元线性预测模型对检验样本进行验证,预测果实质量与实际果实质量的绝对误差均值为 3.260 g,相对误差均值为 1.535%,取得比较理想的预测精度。

运用机器视觉技术、数字图像处理技术和图像提取算法,提取番茄果实图像外部特征参数,可以较准确地对鲜食番茄进行果实质量预测分级。同时,本研究方法可以适用于其他类球体果蔬或物品的质量预测。在今后的研究中可以考虑综合番茄果实不同侧面的特征参数建立模型,以达到更准确的效果。

参 考 文 献

- [1] 曹其新,刘成良,殷跃红,等.基于彩色图像处理的西红柿品质特征的提取研究[J].机器人,2001,23(7):652-656.
[2] 张志强,牛智有,赵思明,等.基于计算机视觉技术的淡水鱼质

- 量分级[J].农业工程学报,2011,27(2):350-354.
[3] 王江枫,罗锡文,洪添胜,等.计算机视觉技术在芒果重量及果面算坏检测中的应用[J].农业工程学报,1998,12(4):186-189.
[4] 刘刚.MATLAB 数字图像处理[M].北京:机械工业出版社,2010.
[5] 张倩,占君,陈珊.详解 MATLAB 图像函数及其应用[M].北京:电子工业出版社,2011.
[6] 张亚静,SHIBUSAWA S,李民赞.基于机器视觉的番茄内部品质预测[J].农业工程学报,2010,26(2):366-370.
[7] 郭峰,曹其新,谢国俊,等.基于 OHTA 颜色空间的瓜果轮廓提取方法[J].农业机械学报,2005,36(11):113-116.
[8] XU L M,ZHAO Y C. Automated strawberry grading system based on image processing [J]. Computers and Electronics in Agriculture,2010,72:32-39.
[9] JOSEPH D,EIFER T,GABRIEL C, et al. Prediction of raw produce surface area from weight measurement [J]. Journal of Food Engineering,2006,74:552-556.
[10] 展慧,李小昱,周竹,等.基于近红外光谱和机器视觉融合技术的板栗缺陷检测[J].农业工程学报,2011,27(2):345-349.
[11] WHITE D J,SVELLINGEN C. Automated measurement of species and length of fish by computer vision [J]. Fisheries Research,2006,80(4):203-210.
[12] BLASCO J,CUBERO S. Development of a machine for the automatic sorting of pomegranate arils based on computer vision [J]. Journal of Food Engineering,2009,90(7):27-34.
[13] 司永胜,乔军,刘刚.基于计算机视觉的苹果识别和形状特征提取[J].农业机械学报,2009,40(8):161-165.

Prediction model of tomato-mass based on external characteristic information

HE Wei NIU Zhi-you LI Xiao-jin

College of Engineering, Huazhong Agricultural University, Wuhan 430070, China

Abstract In order to classify tomatoes nondestructively, this paper put forward method, an image technology to predict tomato weight automatically. The algorithm was based on Matlab platform, using mathematical morphological operation and image local property operation to establish identification algorithm for the tomato. Linear model, second order polynomial model, power model were established through analyzing the relations between tomato external parameters and weight. Weight prediction model was established through regression analysis of the external characteristics. The experiment results showed that the weight was highly related to its area, perimeter, maximum inscribed circle diameter and minimum circumscribed circle diameter. Test results showed that the multivariate linear prediction model was the best. Its determination coefficient (R^2) was 0.926 7, the standard deviation (SE) was 4.32. Its mean relative error was 1.535%, and the mean absolute error was 3.260 g. The experiment results show that the program can quickly predict tomato weight through tomato external features.

Key words tomato; external characteristic; computer vision; prediction model