

基于光电信号的玉米粒计数方法的研究

常礼^{1,2}, 马成学^{1,2}, 高理富²

(1. 中国科学技术大学自动化系, 安徽合肥 230027; 2. 中国科学院合肥智能机械研究所, 安徽合肥 230031)

摘要: 玉米的千粒重是玉米选育优质品种的主要依据. 为了准确、快速地测量出玉米的千粒重, 必须解决玉米粒的计数问题. 为此提出使用光电二极管获取玉米粒做自由落体时的图像信息, 并设计了计算玉米粒数的算法, 取得了较高的计数精度和计数速度. 该玉米粒计数方法和辨识算法对一般谷物的计数具有很好的参考价值.

关键词: 谷物计数; 千粒重; 玉米粒; 分块算法

中图分类号: TP216; TP311.11 **文献标识码:** A doi:10.3969/j.issn.0253-2778.2015.04.003

引用格式: CHANG Li, MA Chengxue, GAO Lifu. Corn-counting method based on photoelectric signal[J]. Journal of University of Science and Technology of China, 2015, 45(4): 275-279.

常礼, 马成学, 高理富. 基于光电信号的玉米粒计数方法的研究[J]. 中国科学技术大学学报, 2015, 45(4): 275-279.

Corn-counting method based on photoelectric signal

CHANG Li^{1,2}, MA Chengxue^{1,2}, GAO Lifu²

(1. Department of Automation, University of Science and Technology of China, Hefei 230027, China;

2. Robot Sensor and Human-Machine Interaction Laboratory, HeFei Institute of Intelligent Machines, HeFei 230031, China)

Abstract: 1000-grain weight is the single crucial criterion for breeding high quality corn. In order to measure the 1000-grain weight, the problem of counting the corn grains must first be solved. Photodiode was used to obtain the image information of corn grains when free falling, and an identification algorithm was designed for calculating the number of corn grains. The method and the algorithm are capable of high accuracy and a high speed, and have very good reference value to the general grain counting.

Key words: grain counting; 1000-grain weight; corn grain; block algorithm

0 引言

我国是世界上第二大玉米生产国, 种植面积仅次于美国, 年产量和消费量均在 1.2 亿吨以上. 所以提高玉米的产量和质量是我国农业发展的必然要求. 提高谷物产量和质量的主要方法是选育出优质品种^[1], 千粒重就是选育品种过程中一项重要的指标. 千粒重在现代农业科学研究中具有非常重要的

作用. 作物的千粒重与饱满度有直接的关系^[2], 千粒重越大亩产量越高. 玉米的千粒重与玉米本身的干物质相关, 不受人工干燥处理影响, 而且千粒重与容重之间存在正相关性^[3], 所以以千粒重替代容重作为衡量玉米品质的评价指标可以避免人工干燥对玉米等级评定的影响^[4]. 千粒重的获取过程中对种子的计数是不可缺少的内容. 玉米粒的计数方法主要有手工计数、基于 CCD 传感器的图像识别计数^[1,4-5]

收稿日期: 2014-10-09; 修回日期: 2015-01-20

基金项目: ITER 国内研究计划专项(2012GB102002)资助

作者简介: 常礼(通讯作者), 男, 1990年生, 硕士研究生. 研究方向: 模式识别、智能控制. E-mail: 13855133902@163.com

以及基于光电二极管的光电计数方法. 人工计数受人的影响因素很大, 而且效率低、成本高; 基于 CCD 的图像识别计数需要把玉米粒平铺到一个平面上, 再进行计算机图像处理, 速度慢、计算成本高, 而且谷物重叠现象不可避免, 给准确度造成很大的影响. 本文使用光电二极管采集玉米粒的图像信息, 用 C 语言编写了一种准确高效的玉米粒计数程序. 不仅提高了玉米千粒重计数的准确率, 而且大大降低了算法的复杂度, 程序已经成功地应用到 51 单片机上, 为玉米粒的技术提供了一种高效便捷的方法.

1 玉米粒信息的获取

为了能够识别玉米粒的数量, 必需对玉米粒的图像信息进行采集. 采集的方法一般有光电二极管采集和摄像头采集. 摄像头采集的优点是速度快, 如果玉米粒全部平铺在传送带上只需一次拍摄即可获得玉米粒的图像信息, 但是很难确保玉米粒相互重叠的现象, 给识别算法增加了难度. 光电采集需要将玉米粒分别垂直落下一粒一粒记录图像信息, 比较耗时间, 虽然在落下的同时也会有重叠的现象发生, 但是光电二极管记录的二值信号容量小, 可以大大减少识别算法的代价, 因此我们采用了光电二极管采集图像信息的方法, 用相互垂直的两排光电二极管来扫描玉米粒两个方向的图像信息, 如图 1 所示.

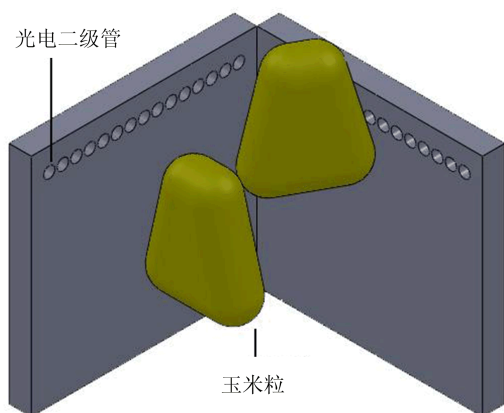


图 1 光电二极管阵列布局

Fig. 1 The layout of the photoelectric diode array

如果只有一粒玉米下落时, 采集的数据如图 2 所示. 其中“1”表示光电二极管被玉米粒遮挡; “0”表示没有被遮挡. 虽然我们尽量控制玉米粒一粒一粒的下落, 最好一粒与下一粒之间有些空隙(可以得到一行“0”, 方便分割数据). 但是玉米粒不可能都是一

样的大小, 如果遇到体积小一些的玉米粒则可能是两粒玉米交错着下落, 这时得到的图像信息如图 3 所示. 当然, 少数情况下也有 3 粒玉米粒交错下落的情况. 我们采集了 3 000 粒玉米下落时的图像信息, 为了方便分析, 把数据保存在一个文本文档里, 以便对数据的离线分析.

```
0001100000000000 0000000000000000
0011110000000000 0011111000000000
0011111000000000 0011111000000000
0001111100000000 0011111000000000
0001111100000000 0011111000000000
0001111100000000 0011111000000000
0000111100000000 0011111000000000
0000001000000000 0011111000000000
```

图 2 玉米粒的图像信息 1(1 粒)

Fig. 2 Image information of the corn(1 grain)

```
00000001111000000 0000000010000000
00000001111100000 00000001111100000
00111101111110000 00000001111110000
01111110111110000 00011001111110000
1111111001100000 00111110111110000
1111111000000000 0011111000000000
0111111000000000 0011111000000000
0000100000000000 0000110000000000
```

图 3 玉米粒的图像信息 2(2 粒)

Fig. 3 Image information of the corn(2 grains)

2 信息的存储

读取玉米粒的图像信息后, 必须把图像信息存储为我们方便处理的数据格式. 本文将信息存储在结构体数组中. 结构体的每个元素对应一块玉米粒的图像信息, 最多包含 3 粒玉米粒. 每个结构体元素包含数据序号(data_num)、图像行数(line_num)、总“1”数(all_1num)、图像的二维数组(photod_inf[30][36])、存储图像特征的数组(character[])、判定的玉米粒数量(corn_num)和实际玉米粒数量(real_corn_num)等结构体成员, 数据存储的流程如图 4 所示.

3 数据的特征分析

3.1 数据的块数

数据的块数是指在玉米粒的图像信息中, 由“1”组成的不相邻的数据块数, 采用 4 邻域法判定是否相邻. 对于如图 3 所示的图像信息就可以用 4 邻域法将数据分成两块. 分块可以采用深度优先搜索

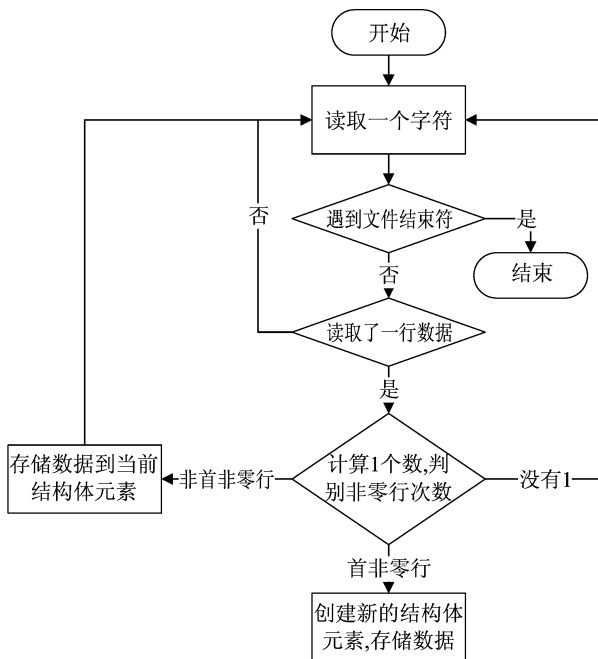


图 4 数据存储流程图

Fig. 4 The flow chart of data storage

(depth_first search, DFS)^[6]算法. 这种遍历方法结合了递归和堆栈两种数据结构的技巧, 一般要加入一个变量来判断该点是否已遍历完毕, 而且递归算法需要不断执行进栈和出栈操作, 需要很大的存储成本. 而我们最后要将算法应用到 51 单片机上, 所以不能使用 DFS 算法. 虽然可以设计 DFS 的非递归算法^[7], 但是对堆栈要求比较高, 对识别块数不够直观且不方便对每一个数据块做后期处理和分析.

基于以上原因, 本文设计了基于结构体链表的数据块数识别算法. 算法的流程图如图 5 所示, 总共使用了 3 类结构体链表 (Remain, Class [i], NneighborA). 首先将图像信息中的所有“1”的坐标存储在链表 Remain 中, Remain 就表示剩余没处理的“1”坐标组成的链表; Class [i] 表示每个数据块的“1”坐标组成的链表; NneighborA 表示 Class [i] 中所有“1”的邻居坐标. 整个算法可以看作先找一个点, 然后不断的寻找他的邻居点, 直到没有邻居点, 将已经找到的点算作一个数据块, 数据块数增加 1; 然后再找另外一个点, 寻找另一个数据块, 直到没有剩下的点. 使用这种数据分块算法可以很好地解决两粒玉米或三粒玉米同时下落交错但不接触的情况. 通过对 3 000 粒玉米粒的实验, 我们得出了如图 6 所示的统计图. 从图中可以看到, 块数和实际个数大部分是吻合的, 但是也有一部分块数为 1 的数据实际

是两个玉米粒. 这就说明块数大于 1 的数据绝对不可能为 1 粒玉米粒, 而块数为 1 的数据却有可能大于 1 粒玉米粒. 用分块算法得出的结果提供了判定的最小玉米粒数.

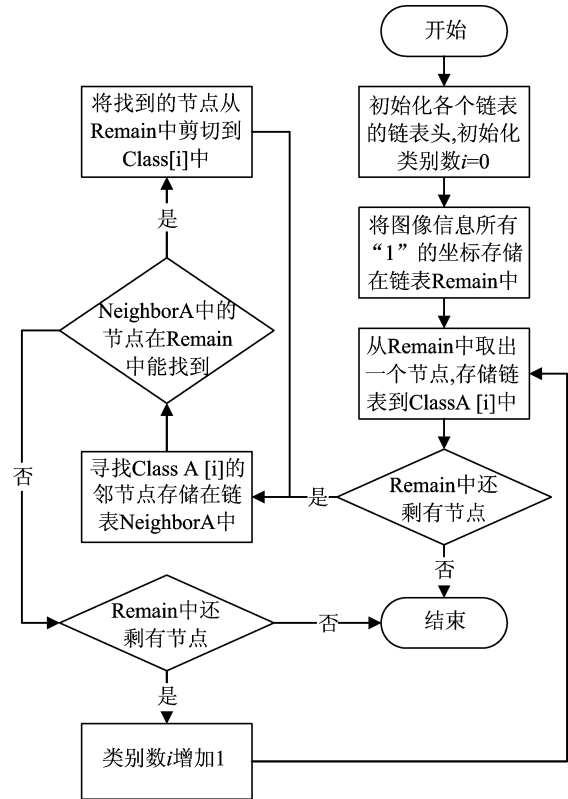


图 5 数据块数识别流程图

Fig. 5 The flow chart of the number of data blocks

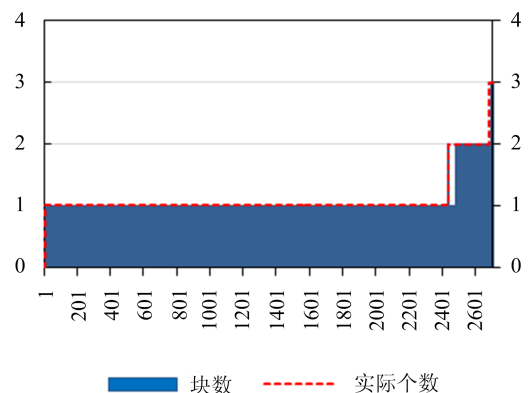


图 6 块数与实际玉米粒数的关系

Fig. 6 The relationship between the block's number and the real grains' number

3.2 数据的方差

数据的方差是指玉米粒图像信息中, 每行“1”的横坐标均值的方差. 计算方差的程序流程图如图 7 所示. 数据的方差反映了图像信息位置左右波动的

程度. 一般情况下, 数据中只有 1 粒玉米粒时, 方差值会非常小, 甚至为零. 通过对 3 000 粒玉米粒的实验, 我们得出了如图 8 所示的统计图. 从图 8 中可以看出, 在玉米粒实际个数为 1 的时候, 方差大部分是小于 2 的, 最高也不会超过 4; 当实际个数为 2 时, 方差有了明显的增大, 绝大部分在 4 以上, 只有极少数在 4 以下; 由于实际个数为 3 的情况非常少, 所以没有得出统计规律. 可以通过设定一个方差的阈值来判定一些数据包含的最小个数.

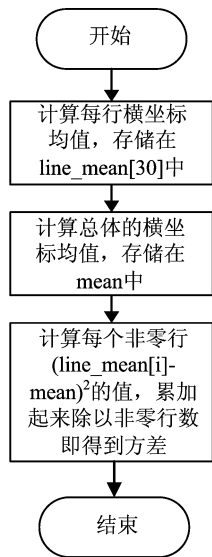


图 7 方差计算流程图

Fig. 7 The flow chart of the variance calculation

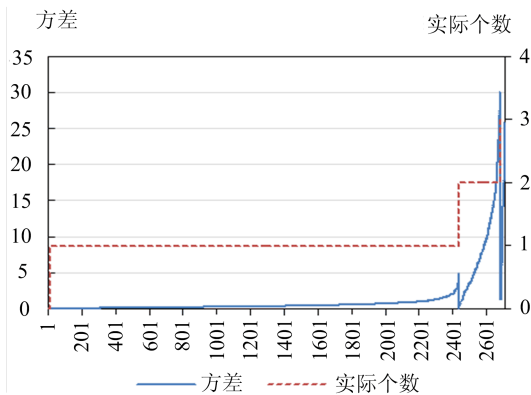


图 8 方差与实际玉米粒数的关系

Fig. 8 The relationship between the variance and real grains' number

3.3 数据总“1”数

数据总“1”数表示存储的玉米粒图像信息中, 所有“1”的个数. 总“1”数的获取非常简单, 可以在信息存储的时候完成计数, 只要添加 1 个计数变量就可以实现. 通过对 3 000 粒玉米的实验, 我们得出了如

图 9 所示的统计图. 从图 9 中可以看出, 当总“1”数小于某个阈值的时候, 玉米粒的数量是不可能超过 1 的, 也就是说, 要想含有 2 个以上的玉米粒, 数据中必须有足够多的“1”. 此外, 总“1”数对 3 粒玉米粒的判别也有重要意义, 从图中 9 可以看到, 实际为 3 粒玉米粒的数据都包含了大量的“1”, 所以可以设定一个总“1”数的下限来判定 3 粒玉米的情况.

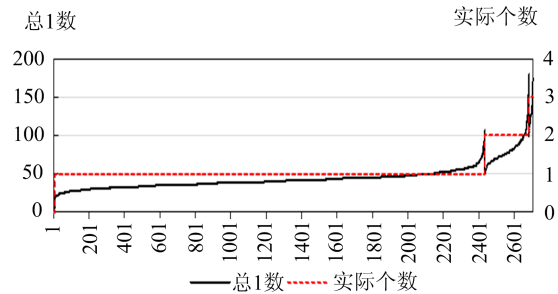


图 9 总“1”数与实际玉米粒个数的关系

Fig. 9 The relationship between the total number of “1” and the real grains' number

3.4 其他特征

除以上分析的特征以外, 从玉米粒的图像信息中还可以得到很多特征, 比如数据的行数, 平均每一行“1”的个数, 每行“1”数乘积之和等, 虽然从统计数据来看都没有很明显的规律, 但是在后期对算法的优化上还是很有帮助的, 尤其是一些特殊的样本信息.

4 玉米粒数量的判别

4.1 判别算法的设计

在获取了玉米粒的一些特征后, 需要分析每个特征对玉米粒数的影响, 并综合各特征对数据中玉米粒的数量作出判定, 也就是设计判别算法. 本文设计的判别算法的流程图如图 10 所示. 算法主要包括去噪声、判定肯定为 1 粒的数据、判定肯定为 3 粒的数据和根据总“1”个数通过不同的方差阈值来判定玉米粒的个数, 再加上其他一些特征的修正判别可以达到很好的判别效果. 对 3 000 粒样本数据判定的结果为 2 999 粒, 漏判和误判的总和在 20 个左右.

4.2 实际应用结果

将本文设计的玉米粒辨识算法烧录到 51 单片机中, 对预先数好的 3 000 粒玉米粒进行计数, 实验结果如表格 1 所示. 由实验结果得出, 计数的误差率可以控制在 5% 以下. 吉林大学的王刚^[4]使用

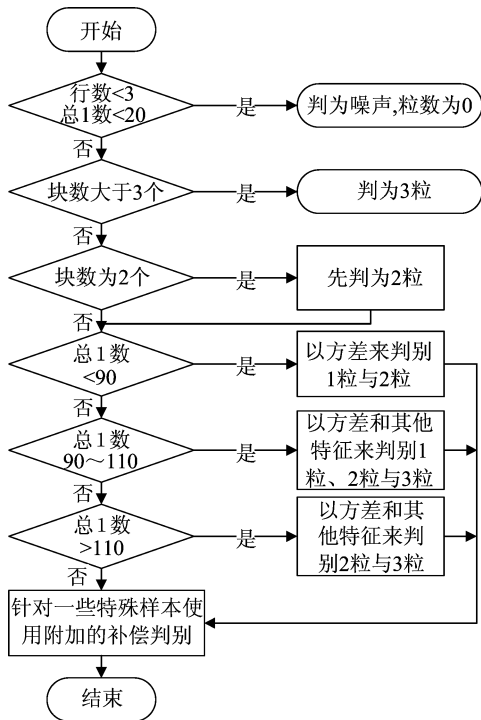


图 10 判别算法程序流程图

Fig. 10 The program flow chart of the discriminant algorithm
USB 摄像头装置采集平铺的玉米粒图像信息,测得的计数误差在 1.05%~1.56%之间,相比之下本文的方法提高了计数准确率。

表 1 算法实际应用结果

Tab. 1 The actual application results of the algorithm

实验序号	实际数量	判定数量	误差率
1	1 000	995	5%
2	1 000	997	3%
3	1 000	1 001	1%

本文使用的计数方法每 1 000 粒玉米计数所使用的时间在 1min 左右,可以实时显示计数过程中玉

米粒的数量. 而王刚的方法获得的速度在 12.28 粒/秒~22.33 粒/秒之间,没有明显区别,但是本文的方法不需要铺设玉米粒,计数更直接,可以显著提高计数的总体效率。

5 结论

通过实验分析和实际应用结果可以得出:在使用光电二极管获取玉米粒图像信息的基础上运用本文设计的玉米粒粒数识别算法可以明显地提高玉米粒计数的效率和准确率,这将大大减少人的劳动强度,为提高玉米的优质育种效率降低了成本。

误差率 5% 仍然有待于改进,玉米粒图像信息的特征仍然需要进一步分析. 随着样本数量的增多,采用统计学规律可以进一步改善判别算法。

参考文献 (References)

[1] 贾鹏, 李永奎, 赵萍. 基于 Matlab 图像处理的谷物颗粒计数方法研究[J]. 农机化研究, 2009, (1): 152-153, 156.

[2] 高学曾, 王忠孝, 许金芳, 等. 玉米穗粒数和千粒重与产量的关系[J]. 山东农业科学, 1989, (2): 4-7.

[3] 王琳. 玉米容重相关性分析[J]. 粮油仓储科技通讯, 2009, (3): 37-38.

[4] 王刚. 基于机器视觉的玉米千粒重快速检测仪的研制[D]. 吉林大学, 2012.

[5] 段丽英. 颗粒图象识别及计数在粮种品质测定中的应用[D]. 河北科技大学, 2010.

[6] 刘萍, 冯桂莲. 图的深度优先搜索遍历算法分析及其应用[J]. 青海师范大学学报(自然科学版), 2007, (3): 41-44.

[7] 刘中华, 张颖超. 深度优先搜索的非递归算法[J]. 科技信息, 2010, (25): 160-161.