

基于图像识别的小麦品种分类研究

何胜美^{1,2}, 李仲来², 何中虎^{1,3}

(¹ 中国农业科学院作物科学研究所/国家小麦改良中心, 北京 100081; ² 北京师范大学数学科学学院, 北京 100875;

³ CIMMYT 中国办事处, 北京 100081)

摘要: 基于数字图像分析, 利用小麦籽粒的 20 个形态特征和 12 个颜色特征对来自中国 4 个地点 7 个春小麦品种共 28 个样本进行分类和识别。对于不同品种和地区的样本, 分别利用逐步判别分析, 选取显著性较大的特征参量, 建立各地区和品种的贝叶斯分类器模型。结果表明, 对各地区品种识别的正确回判率和测试集的正确识别率均达到 100%。将各样本按品种合并, 再对合并后的样本进行品种识别, 除了新克旱 9 号的回判率为 98.3% 外, 其它品种的回判率均为 100%。测试集中, 龙麦 26 和青春 566 正确识别率分别为 97.5% 和 95.0%, 其它品种均为 100%。品种来源地识别也能达到较高的水平, 甘肃、宁夏、新疆和黑龙江的正确识别率分别为 88.6%、92.9%、72.9% 和 95.7%。说明利用籽粒图像对小麦品种进行识别高效可行。

关键词: 普通小麦; 品种; 图像处理; 模式识别

Classification of Wheat Cultivar by Digital Image Analysis

HE Sheng-mei^{1,2}, LI Zhong-lai², HE Zhong-hu^{1,3}

(¹ Crop Science Institute / National Wheat Improvement Center, Chinese Academy of Agricultural Sciences, Beijing 100081;

² School of Mathematical Sciences, Beijing Normal University, Beijing 100875;

³ CIMMYT-China Office, Chinese Academy of Agricultural Sciences, Beijing 100081)

Abstract: Digital image analysis was used to develop a pattern recognition algorithm to classify individual kernels of seven Chinese spring wheat cultivars grown at 4 locations. Totally, 20 morphological parameters and 12 color parameters were extracted. Three hundred kernels per sample were used as the training data set to develop identification model, and another 200 kernels were used as the test set. For the test set, the classification accuracy of wheat cultivars was 100% in each growing location. Except for Xinkehan 9 with 98.3%, the correct discrimination of the training set of collective samples is 100% for wheat cultivar. For the test set, the correct discrimination of Longmai 26 and Qingchun 566 were 97.5% and 95.0%, the others is 100%. For the origin of wheat grains, the classification of Gansu, Ningxia, Xinjiang and Heilongjiang were 88.6%, 92.9%, 72.9% and 95.7%, respectively. The results show that it is feasible to identify and classify wheat cultivar (grains) using digital image analysis.

Key words: Common wheat; Variety; Image process; Pattern recognition

计算机图像处理是一项综合性技术, 在遥感图像处理 and 生物医学图像分析方面已取得了卓越成效。近十多年来, 随着硬件成本下降和计算速度的提高, 计算机图像处理广泛介入农产品品质检测和分级, 成功地用于玉米、大麦、小麦、大豆、燕麦、黑麦草和水稻等种子检验^[1]。国外在小麦种子检测和识别方面开展了广泛的研究, 一方面成功地识别小麦品种和类别,

如大麦、小麦、燕麦和黑麦草的正确识别率均在 90% 以上^[2-6]; 另一方面有效识别不同类别的小麦种子, 对不同冬春性、质地和粒色的小麦类别判断效果好^[7,8], 但对同一类型不同品种的识别效果并不理想^[9]。在研究技术上, 提取了多种种子特征, 如形状特征^[2,3,6-11]、颜色特征^[4,6,11-13]和纹理特征^[5]等。Sapirstein 等系统研究了图像分析中抽样对分类精度

收稿日期: 2004-11-18

基金项目: “948” 重大农业国际合作项目 (2003Q01)

作者简介: 何胜美 (1979-), 男, 湖南新邵人, 硕士研究生, 主要从事生物数学和图像识别研究。何中虎为通讯作者, Tel: 010-68918547; E-mail: zhhe@public3.bta.net.cn

的影响,表明试验样本大小最好在 300 至 500 粒^[14]。但国内只有对麦苗图像识别的研究^[15,16],而对小麦种子图像识别的相关研究还未见报道。过去籽粒图像识别以单个小麦籽粒为单位,对籽粒选取的一致性要求较高,对实际品种的代表性偏低,而且操作比较麻烦。本研究应用数字图像处理技术,以整幅图像为单位识别来自中国 4 个地区 7 个春小麦品种共 28 个样本,同时进一步分析环境对小麦籽粒形态的影响。

1 材料与方法

1.1 材料

1.1.1 样本说明 所用样本包括 7 个小麦品种,分别是宁春 4 号、宁春 16 号、陇春 15 号、青春 566、新春 6 号、新克早 9 号和龙麦 26 号。将上述 7 个品种按产量比较试验种植于黑龙江哈尔滨、宁夏永宁、新疆乌鲁木齐和甘肃武威 4 个地点。籽粒样品中除去过小和干扁籽粒,每份样品为 500 粒,其中 300 粒用于建立判别分类模型,200 粒用于检验模型的效率。图像采样过程中,每幅图片共 20 个小麦籽粒,这样每个样品可测 500 个籽粒,得到容量为 25 的样本,其中 15 幅作为训练集,10 幅作为测试集。

1.1.2 图像获取和分析系统 图像采用 CCD (charge coupled device)彩色数码相机(型号: DSC F717, SONY)获取,相机装备一个 10×光学变焦(5 倍精确数码变焦)镜头(型号: Carl Zeiss Vario-Sonnar 10×)。制作一个底为边长为 40 cm 正方形、高 18 cm 的侧面一面开口的玻璃盒子。在玻璃盒子上底面中心凿一直径为 6.5 cm 的圆形小孔(刚好伸过镜头前部),相机固定在玻璃盒上,镜头通过小孔深入箱子中。通过玻璃盒侧面开口处在盒下底面置一可以自由抽动的木制载物台,作为籽粒背景平台,平台表面背景颜色为黑色。拍摄过程中,使用手动调节、延迟拍摄模式,以得到稳定的拍摄环境。将一根三基色环形荧光灯管(型号:松下 FL40ss.Ex-N/36, 36W, 色温 5000K)平放在样品平台周围,稍低于平台,用半径为 0.36 m 的半球形的灯罩扣在平台上方充当光源扩散体,目的是提供一致的扩散光照条件。将每 20 粒小麦籽粒背面朝上构成一 5×4 的矩阵排列于长 50 mm,宽 40 mm 的黑色背景上,用相机在 640×480 的分辨率下摄取背面图像,每像素点为 0.078 mm×0.833 mm 的矩形区域,面积为 0.0065 mm²。

图像处理和分析在台式计算机(Dell, Pentium 4)上完成。操作系统为 Windows2k,工程计算软件

MATLAB 6.5 的图像处理工具箱作为图像处理和分析主要平台。数据统计分析采用常用统计软件 SAS(SAS Institute Inc U.S.A. 1976)。

1.2 图像分析方法

利用上述图像获取系统得到的图像通过 USB 接口转入计算机处理。为了减少图像拍摄和传输过程中随机干扰带来的图像噪声的影响,本研究采取中值滤波平滑技术,在消除噪声的同时,较好地保持图像边缘^[3,17]。经过平滑处理的图像,利用阈值分割技术将小麦籽粒图像从背景中分割出来,灰度阈值选取基于全局阈值^[17]。首先选取一个初始估计阈值 $T=(max+min)/2$,其中 max 和 min 分别为图像灰度级的最大值和最小值。用 T 分割图像生成两组像素: G_1 由所有灰度值大于 T 的像素组成,而 G_2 由所有灰度值小于或等于 T 的像素组成。对区域 G_1 和 G_2 中所有像素分别计算平均灰度阈值 μ_1 和 μ_2 ,然后计算新的阈值: $T'=(\mu_1+\mu_2)/2$,比较计算出来的 T' 值与 T 值,如果它们之差的绝对值小于给定的参数 t_0 则停止迭代, T' 值作为最终图像分割阈值。否则,令 $T=T'$,重复计算 T' 。通过图像分割,将小麦籽粒图像单个从背景中分离出来,以便特征提取。

1.2.1 形状特征提取 通过二值图像标号将单个小麦籽粒图像提取出来,利用籽粒区域和边界特性,提取籽粒形状特征。

面积:以小麦籽粒图像所包含的所有像素总数计算。

周长:小麦籽粒图像边界像素总和。

圆形度=(周长)²/(4π×面积)。

长轴长:小麦籽粒图像上距离最长的两个端点之间的欧氏距离。

短轴长:小麦籽粒图像上过长轴中点且垂直于长轴的直线间距离。

最小半径:籽粒边界上的像素到质心的最小距离。

最大半径:籽粒边界上的像素到质心的最大距离。

基于区域的不变矩:

Hu 提出了对于评议旋转和大小尺度变化不变矩,常用于形状识别,刻画事物的形状特征^[17]。首先定义一个规格化中心矩 η_{pq} ,表达式为:

$$\eta_{pq}=M_{pq}/M_{00}^p$$

其中:

$$M_{pq} = \sum_{i=1}^n \sum_{j=1}^m (i - i_G)^p (j - j_G)^q f(i, j), p, q = 1, 2, \dots,$$

$k, (i_G, j_G)$ 是籽粒的质心, $f(i, j)$ 是像素 (i, j) 的灰度值。

利用二阶和三阶中心矩可以导出以下不变矩:

$$M_1 = \eta_{20} + \eta_{02} \quad M_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$$

$$M_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$

$$M_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$

傅立叶描绘子

设 $d_k = [(i_k - i_G)^2 + (j_k - j_G)^2]^{1/2}$ 是籽粒边界像素到质心的距离, 其中 (j_k, i_k) 是边界上第 k 个像素, (i_G, j_G) 是籽粒的质心。傅立叶描绘子用以下公式计算:

$$FD_u = [R_u^2 + I_u^2]^{1/2} \quad u = 0, 1, 2, \dots, N-1$$

其中

$$R_u = \sum_{k=0}^{N-1} d \times \cos(2pk_u / N),$$

$$I_u = \sum_{k=0}^{N-1} d \times \sin(2pk_u / N),$$

N 为籽粒边界像素总数。

1.2.2 颜色特征提取 小麦籽粒颜色是一个重要特征, 通常采用 RGB 颜色系统, 本文将 RGB 颜色信息和 HIS 颜色配合使用识别小麦籽粒。由光度学知, HIS 系统直接用亮度 (intensity)、色调 (hue) 和饱和度 (saturation) 来描述颜色, 更符合人类视觉特征。在籽粒颜色特征提取中, 利用图像分割确定籽粒的位置, 然后综合原来的 RGB 图像对每颗籽粒分别求出红色、绿色、蓝色、色调、亮度和饱和度均值和方差, 作为 ss 颜色特征。

2 结果与分析

2.1 特征提取

总共选取了 32 个小麦籽粒特征参数, 其中 20 个形态特征分别为籽粒面积、周长、圆形度、长/短轴长、最大/最小半径、半径比、半径均值、离心率、等价半径、一到四阶不变矩和零到五阶边界傅立叶描绘子, 其中离心率是采用椭圆来拟合籽粒边界所对应的椭圆的离心率, 和圆形度一起, 作为小麦籽粒圆和扁程度的描述, 当籽粒为圆形时, 圆形度最大为 1, 籽粒形状越扁, 圆形度越小, 离心率越大。一到四阶不变矩对平移、旋转和大小尺度变化不变, 主要反映了区域中灰度相对于灰度中心是如何分布的度量。边界特征中, 一些低阶傅立叶描绘子能够反映物体的大体形状, 而高阶傅立叶描绘子是精确定义形状特征 (比如拐角和直线) 所需的。12 个颜色特征主要是红色、绿色、蓝色、色调、饱和度和亮度的均值和方差。本研究首次采用离心率和等价半径来描述小麦籽粒特征, 对各品种和来源地的方差分析表明, 离心率和等价半径在地区间和品种间存在显著差异, 同时识别结果表明这两个特征具有很好的区分度。

2.2 品种识别

对宁夏永宁、新疆乌鲁木齐、黑龙江哈尔滨以及甘肃武威四个地区的样本分别建立分类器模型, 进行品种识别, 效果非常显著。对宁夏永宁的样本, 采用 SAS 逐步判别分析选出前 7 个特征如表 1。

利用 SAS 判别分析建立各品种相应的线性判别模型, 陇春 15、龙麦 26、宁春 16、宁春 4、青春 566、新春 6 和新克早 9 对应的判别函数分别为:

表 1 利用逐步判别分析选取前 7 个贡献较大的特征

Table 1 Selection of seven features of individual wheat kernel using STEPDISC analysis

序号	变量	籽粒特征	平均典型相关系数	偏 R
No.	Variable	Kernel feature	Averaged squared canonical correlation	Partial R
1	X_1	一阶不变矩 M1 first invariant moment. M1	0.161	0.966
2	X_2	最大半径 Maximum radius	0.304	0.885
3	X_3	红色分量 Red	0.437	0.842
4	X_4	绿色分量 Green	0.551	0.792
5	X_5	亮度 Intensity	0.590	0.820
6	X_6	离心率 Eccentricity	0.675	0.634
7	X_7	籽粒面积 Area	0.714	0.485

表 4 总体样本测试集品种判别分析结果

Table 4 The discriminant results of the test set of collectivity samples on wheat variety

品种 Variety	陇春 15 Longchun 15	龙麦 26 Longmai 26	宁春 16 Ningchun16	宁春 4 Ningchun 4	青春 566 Qingchun 566	新春 6 Xinchun 6	新克早 9 Xinkehan 9
陇春 15 Longchun15	40(100%)	0	0	0	0	0	0
龙麦 26 Longmai 26	0	39(97.5%)	0	0	0	0	1
宁春 16 Ningchun16	0	0	40(100%)	0	0	0	0
宁春 4 Ningchun 4	0	0	0	40(100%)	0	0	0
青春 566 Qingchun 566	0	0	2	0	38(95.0%)	0	0
新春 6 Xinchun 6	0	0	0	0	0	40(100%)	0
新克早 9 Xinkehan 9	0	0	0	0	0	0	40(100%)

上述结果表明，图像识别能对小麦品种进行很好识别，同时也表明不同品种小麦籽粒的上述形状特征和颜色特征存在显著差异。

2.3 来源地识别

对于所用 7 个品种，分别建立分类模型，以识别每个样本的来源地，对训练集的正确回判率和测试集的正确识别率分别列于表 5 和表 6。

从判别结果来看，除新克早 9 号之外，其它 6 个样本都取得了很好的识别效果。新克早 9 号在回判过程中，15 个来自甘肃的样本中有 5 个（33.3%）被判为新疆，而 15 个来自新疆的样本中有 3 个（20.0%）

判为甘肃；在测试集识别中，有 9 个（90.0%）来自甘肃的样本误判成为新疆，7 个（70.0%）新疆样本错判为甘肃。出现这种情况的可能原因是新克早 9 号为光敏感型品种，在不同地区之间表现差异较大，而新疆和甘肃的纬度和生态环境较为接近，二者表现较为相似。依地区把各样本合并，对得到的合并样本来源地识别结果如表 7。总体识别效果理想，但在甘肃样本和新疆样本之间的误判率较高，有 13 粒（18.7%）的新疆样本误判为甘肃样本。因此，对甘肃和新疆两个点的样本提取新的特征进行单独判别分析是必要的。

表 5 各品种来源地识别模型训练集正确回判率

Table 5 The discrimination results of the training set about growing locations of each cultivar

来源 Location	陇春 15 Longchun 15	龙麦 26 Longmai 26	宁春 16 Ningchun16	宁春 4 Ningchun 4	青春 566 Qingchun 566	新春 6 Xinchun 6	新克早 9 Xinkehan 9
甘肃 Gansu	100	100	100	100	100	100	66.7
宁夏 Ningxia	100	100	100	100	100	100	100
新疆 Xinjiang	100	100	100	100	100	100	80
黑龙江 Heilongjiang	100	100	100	100	100	100	100

表 6 各品种来源地识别模型测试集正确识别率

Table 6 The discrimination results of the test set on growing locations for each cultivar

来源 Location	陇春 15 Longchun 15	龙麦 26 Longmai 26	宁春 16 Ningchun16	宁春 4 Ningchun 4	青春 566 Qingchun 566	新春 6 Xinchun 6	新克早 9 Xinkehan 9
甘肃 Gansu	100	100	100	90	100	100	0
宁夏 Ningxia	100	100	100	100	100	100	100
新疆 Xinjiang	100	90	80	100	100	90	20
黑龙江 Heilongjiang	100	100	100	80	100	70	90

表 7 总体测试集来源地的判别分析结果

Table 7 The discrimination results of growing locations for each cultivar with all samples

来源 Location	甘肃 Gansu	宁夏 Ningxia	新疆 Xinjiang	黑龙江 Heilongjiang	总计 Total
甘肃 Gansu	62 (88.6%)	0	4	4	70
宁夏 Ningxia	2	65 (92.9%)	3	0	70
新疆 Xinjiang	13	1	51 (72.9%)	5	70
黑龙江 Heilongjiang	3	0	0	67 (95.7%)	70

3 讨论

本研究表明,应用籽粒的形状特征和颜色特征能对小麦的品种和来源地进行有效识别,品种正确识别率均在 95%以上,来源地的正确识别率平均为 87.5%,说明应用计算机视觉进行小麦识别是可行的。本研究中将每幅图片 20 粒小麦各特征的平均值作为一个观测值,更能真实反映小麦籽粒的特征和方便实际操作,进一步提高了识别效果。例如对陇春 15,以单个籽粒特征为单位来识别来源,甘肃、宁夏、新疆和黑龙江的正确识别率分别为 87.5%、95.0%、93.5%和 92.0% (表 8),但以图片为单位的正确识别率均能达到

100% (表 6),其它品种除新克早 9 号以外,正确识别率都有一定的提高。本文中提出用离心率和等价半径两个新的特征来刻画小麦籽粒形态,丰富了小麦籽粒图像识别的研究。对同一品种不同栽培地点的识别结果表明,籽粒外观形态和颜色对外部生长环境非常敏感,进一步肯定和补充了国外的研究工作,计算机图像识别具有客观性和可重复性,为小麦品种识别提供了客观可行的方法。同时,研究中主要采用了比较具有代表性的中国主要麦区春小麦品种,但其方法也可以应用到粒状农产品识别领域,比如冬小麦,大米,玉米等,国内外已经涌现了不少应用的实际例子 [1,18]。

表 8 以单个籽粒为单位对陇春 15 号测试集来源地识别结果

Table 8 The discrimination results of growing location on Longchun 15 with traditional method

来源 Location	甘肃 Gansu	宁夏 Ningxia	新疆 Xinjiang	黑龙江 Heilongjiang	总计 Total
甘肃 Gansu	175(87.5%)	15	0	10	200
宁夏 Ningxia	2	190(95.0%)	7	1	200
新疆 Xinjiang	1	12	187(93.5%)	0	200
黑龙江 Heilongjiang	16	0	0	184(92.0%)	200

4 结论

小麦籽粒的形状和颜色是小麦的重要特征,应用小麦籽粒的形状特征和颜色特征能对小麦的品种和来源地能有效识别,为小麦种质纯度检验提供了客观方法。相比肉眼识别而言,计算机识别结果更客观、准确和一致。

References

- [1] 刘燕德, 应义斌, 成芳. 计算机视觉技术在种子纯度检验中的应用. 农业机械学报, 2003, 34(5): 161-163.
Liu Y D, Ying Y B, Cheng F. Research of machine vision in purity inspection of seed. *Transactions of The Chinese Society of Agricultural Machinery*, 2003, 34(5): 161-163. (in Chinese)
- [2] Sapirstein H D, Neuman M, Wright E H, Shweddyk E, Bushuk W. An

- instrumental system for cereal grain classification using digital image analysis. *Journal of Cereal Science*, 1987, 6: 3-14.
- [3] Majumdar S, Jayas D S. Classification of cereal grains using machine vision: I. Morphology models. *Transaction of the ASAE*, 2000, 43: 1 669-1 675.
- [4] Majumdar S, Jayas D S. Classification of cereal grains using machine vision: II. Color models. *Transaction of the ASAE*, 2000, 43: 1 677-1 680.
- [5] Majumdar S, Jayas D S. Classification of cereal grains using machine vision: III. Texture models. *Transaction of the ASAE*, 2000, 43: 1 681-1 687.
- [6] Majumdar S, Jayas D S. Classification of cereal grains using machine vision: IV. Combined morphology, color, and texture models. *Transaction of the ASAE*, 2000, 43: 1 689-1 694.
- [7] Symons J S, Fulcher R G. Determination of wheat kernel

- morphological variation by digital image analysis: I. Variation in eastern Canadian milling quality wheats. *Journal of Cereal Science*, 1988, 8: 211-218.
- [8] Symons J S, Fulcher R G. Determination of wheat kernel morphological variation by digital image analysis: II. variation in cultivars of soft white winter wheats. *Journal of Cereal Science*, 1988, 8: 219-229.
- [9] Neuman M R, Sapirstein H D, Shwedyk E, Bushuk W. Discrimination of wheat class and variety by digital image analysis of whole grain samples. *Journal of Cereal Science*, 1987, 6: 125-132.
- [10] Zayas, Lai F S, Pomeranz Y. Discrimination between wheat classes and varieties by image analysis. *Cereal Chemistry*, 1986, 63: 52-56.
- [11] Utku H, Koksels H, Kayhan S. Classification of wheat grains by digital image analysis using statistical filters. In: Braun H J, Altay F, Kronstad W E, Beniwal S P S, McNab A, ed. *Wheat: Prospects for Global Improvement*. Netherlands: Kluwer Academic Publishers, 1998: 237-244.
- [12] Neuman M R, Sapirstein H D, Shwedyk E, Bushuk W. Wheat grain color analysis by digital image processing: I. Methodology. *Journal of Cereal Science*, 1989, 10: 175-182.
- [13] Neuman M R, Sapirstein H D, Shwedyk E, Bushuk W. Wheat grain color analysis by digital image processing II: wheat class discrimination. *Journal of Cereal Science*, 1988, 10: 183-188.
- [14] Sapirstein H D, Kohler J M. Effects of sampling and wheat grade on precision and accuracy of kernel features determined by digital image analysis. *Cereal Chemistry*, 1999, 76: 110-115.
- [15] 李少昆, 索兴梅, 白中英, 祁之力, 刘晓鸿, 高世菊, 赵双宁. 基于 BP 神经网络的小麦群体图像特征识别. *中国农业科学*, 2002, 35(6): 616-620.
- Li S K, Suo X M, Bai Z Y, Qi Z L, Liu X H, Gao S J, Zhao S N. The machine recognition for population feature of wheat images based on BP neural network. *Scientia Agricultura Sinica*, 2002, 35(6): 616-620. (in Chinese)
- [16] Gonzalez R C, Woods R E. *Digital Image Processing* (2nd ed.). Beijing: Publishing House of Electronics Industry Press, 2003.
- [17] 边肇祺. 模式识别. 北京: 清华大学出版社, 2000.
- Bian Z Q. *Pattern Recognition*. Beijing: Tsinghua University Press, 2000. (in Chinese)
- [18] 王桂琴, 杨子彪, 郑丽敏, 朱虹, 廖树华, 单成钢, 吴富宁. 计算机视觉在农产品检测中的应用. *中国农业科技导报*, 2003, 5(3): 52-56.
- Wang G Q, Yang Z B, Zheng L M, Zhu H, Liao S H, Shan C G, Wu F N. Application of computer vision technology testing agricultural products. *Review of China Agricultural Science and Technology*, 2003, 5(3): 52-56. (in Chinese)

(责任编辑 王芳)