

# 利用计算机视觉技术的烟叶质量分选系统研究

蔡健荣 方如明 张世庆 吴守一

(江苏理工大学)

张建平

(上海烟草(集团)公司技术中心)

**摘要:** 利用计算机视觉技术开发了烟叶质量分选系统。该系统对采集系统进行定标, 控制感光度。提取了 180 个特征参数并进行选择形成特征向量, 去除了标准样本中的奇异样本。利用神经网络对多个地区的烟叶进行学习和分类, 检测准确率均在 80% 以上, 半数地区检测准确率在 90% 以上。对烟叶分类该系统具有较高的实用价值。

**关键词:** 烟草; 计算机视觉; 神经网络; 质量检验

烟叶从采摘、收购到加工成卷烟需经过很多处理。一般当年收购的烟叶不能立即加工, 需存储一年以上进行调质处理。存储期间烟叶的组织发生了变化, 有些烟叶因存储条件问题发生压油、霉变现象。另外, 根据卷烟的配方要求, 需混合不同产地、不同等级的烟叶。由于各烟区每年收购的烟叶因气候和分级水平的差异, 而产生质量差异。为了保障烟民的身体健少受伤害, 稳定卷烟的口味和品质, 一般要进行人工挑叶, 剔除变质烟叶和不合格烟叶。高档卷烟的分选要求更为严格, 需逐片挑选。这些工作都由人工凭肉眼鉴别完成, 劳动强度极大。从人工挑叶的整个过程可见, 利用计算机视觉技术和适当的模式分类技术可以实现计算机自动检测。因此在已有研究基础上<sup>[1]</sup>, 我们开发了用于挑叶的计算机视觉分选系统。整个系统由几部分组成: 计算机视觉系统、标定系统、学习系统、分选系统。

## 1 视觉系统工作原理和系统组成

计算机视觉系统的主要技术是图像处理技术。它是用计算机实现部分人类视觉的功能, 把所测对象映射成数字图像, 并模拟人的判别准则去理解图像和识别图像, 进而对所摄图像进行分类或分级。烟叶质量主要从烟叶的颜色、身份、油分、组织结构、成熟度、大小等几方面加以分析。为了能使计算机识别烟叶质量, 首先必须把这些质量因素转变成计算机容易处理的数字量, 即特征提取。烟叶各质量因素大

部分和颜色有关。颜色的三大特性是: 色调、明度、彩度(饱和度)。要计算机模拟人的视觉, 必须选用符合人视觉习惯的颜色模型。孟塞尔颜色模型是根据大量视觉实验总结出来的模型, 最符合人的视觉特点<sup>[2]</sup>。它是用色调、明度、彩度三维空间表示颜色。本项目研究出了实现颜色从红、绿、蓝三原色系统向孟塞尔颜色系统的快速转换模型, 并将颜色的总体分布曲线, 经抽样产生特征参数。烟叶大小形状提取的特征主要有面积、长度、残伤及对烟叶轮廓线展开后的分布特征。烟叶的组织结构靠视觉较难实现, 但对图像表面进行纹理分析, 可实现对烟叶组织结构的间接描述。本系统采用了灰度共生矩阵法, 把烟叶纹理从细到粗分别提取, 形成从细纹到粗纹总体分布。因此, 在描述烟叶特征时, 用颜色、形状、纹理三类特征描述每片烟叶, 计算机自动提取 180 个特征参数。

计算机视觉系统的工作过程是: 首先标定采集参数, 然后采集图像, 根据图像提取特征参数; 对提取的特征进行分析和处理; 学习专家的分选思想, 对烟叶进行分类。实现烟叶的准确分类要包括对烟叶样本学习和检测两个过程, 如图 1 所示。

计算机视觉系统主要组成有: 计算机、图像监视器、摄像头、电动三可变镜头、光源(4 个灯管, 色温 6 500K)、漫反射光源箱、CA 6300 真彩色图像采集卡(采集速度 25 帧/s, 系统误差 1 个灰度级), 如图 2 所示。

## 2 数据采集环境设置和软件补偿

利用计算机视觉系统对烟叶进行检测分类, 首先必须确保系统误差尽可能小, 保证数据采集的稳定性和可靠性; 其次要使采集的信息可分性好。本研

收稿日期: 2000204212

蔡健荣, 博士生, 镇江市 江苏理工大学生物与环境工程学院, 212013

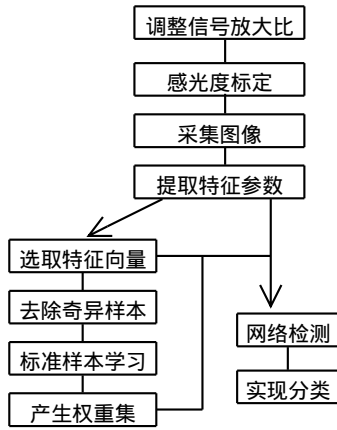
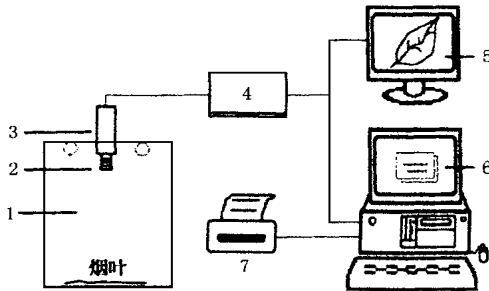


图1 分选系统工作流程图

Fig 1 Flow chart of classifying system



1. 光源箱 2 电动三可变镜头 3 摄像头  
 4 图像卡 5 图像监视器 6 计算机 7 输出设备  
 图 2 烟叶分选系统示意图

Fig 2 Schematic diagram of classifying system of tobacco leaves

究的目的是从经过调质处理后, 同级别烟叶中分选出合格、两用和不合格烟。从分类的角度看, 同一级别的烟叶属于同一类, 但相互间还存在区别, 因此要求系统要有极高的采集精度。本研究对硬件系统的参数进行了设置, 并对照明光强和光线的均匀性作了软件补偿。

### 2.1 调整采集卡输入信号的放大倍数

数据采集时, 一般要调整并稳定采集信号亮度、对比度、饱和度的放大比。这 3 个参数的放大倍数影响所采集图像的颜色逼真度。放大倍数太小, 采集的图像可分性不明显, 与本研究要求较好的可分性相矛盾; 放大倍数过大, 在放大信号的同时噪声也随之放大, 增加了系统误差, 降低采集信号的稳定性, 因此需选择合适的放大倍数。本研究针对检测对象的具体情况, 对同一样本在不同的信号放大倍数下, 重复进行特征值匹配试验, 得出信号放大倍数和采集精度之间的关系, 如图 3 所示。实验证明亮度放大倍

数在 31 以内可满足精度要求, 当放大倍数大于 31 则试验的重复性开始明显降低; 放大倍数小于 28 则样本的可分性开始降低, 因此放大倍数控制在 28~31 之间合适。对比度和饱和度放大倍数在 30 左右时, 采集图像的颜色也最逼真。

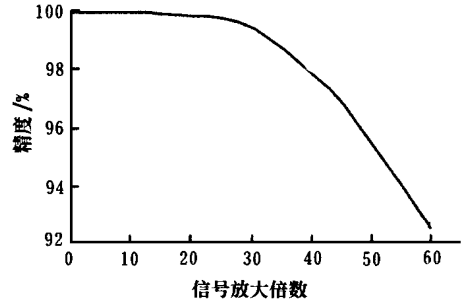


图 3 亮度不同放大倍数下测量精度

Fig 3 Measuring precision on different light amplification scales

### 2.2 CCD 光圈调整和光线不均匀的软件补偿

图像处理尤其是彩色图像处理对光线的强弱和偏色非常敏感, 因此必须控制这些影响因素。本系统选用的照明光源是色温和太阳光接近的灯管, 因此基本消除偏色。但光线亮度和光线的均匀性, 摄像机光圈和焦距等可变因素的控制也很重要。摄像机焦距控制主要靠观察者目测, 当焦距调节至一定量后, 如果焦距旋钮正转或反转都会使拍摄的动态图像变模糊时, 即认为调焦成功。而光线亮度和光线的均匀性靠目测很难实现, 系统软件提供了这些控制手段, 研究中设置了标定亮度为 230(最高亮度 255)。

**光圈调整:** 这是辅助调整摄像机的光圈的手段。制备一专用白板, 供每次检测前标定。检测前, 将白板置于光源箱内, 由计算机实时测量, 动态显示白板亮度值。实验者根据计算机提示, 手工调节摄像机镜头光圈, 使亮度平均值在 230 左右, 这是对亮度的粗调。

**亮度微调:** 光圈调整后感光强度基本达到要求, 但与标定值仍会有一定的差异, 用程序自动调整亮度放大倍数, 使感光量保持在标定值不变。如果亮度放大倍数已达到许用范围(28~31), 采集系统感光量仍不满足亮度要求, 系统会提示加大或缩小光圈, 需反复调整光圈和亮度放大倍数, 直到达到亮度的标定值。

**光源不均匀软件补偿:** 光源箱是由漫反射材料组成, 消除了部分光源不均匀性。有时部分灯管老化或其它原因, 摄像机可视区域仍会存在光线不均匀

现象。为了减轻对检测的影响,这里用标准白板做样板,换算出补偿系数。在每次检测前作标定,并在采集数据时作自动补偿。

### 3 学习系统的建立

烟叶质量的评价虽然有统一的标准,但产地和出产年份的不同又有具体的衡量准则。计算机学习也应因地制宜,并能够对烟叶的特征参数进行适当的调整。因此在学习时应做到不同时域和地域的烟叶分别建立特征库和学习库。

#### 3.1 特征参数和标准样本的选择

本系统中,从烟叶中提取了 180 个特征值作为特征参数集,它基本覆盖了各类烟叶特征,可视为烟叶的总体特征。特定时域的烟叶样本,其特征参数总是烟叶总体特征的一部分。在确保特征参数能代表烟叶特征的前提下,供学习的特征值越少,学习和检测速度越快,因此要对这 180 个特征进行筛选。选择方法是利用统计分析法产生最后特征值数。

首先,对烟叶提取的所有特征参数作统计分析,一般特征值分布均服从正态分布,且特定时域的烟叶样本的特征值集中分布在某一区间。设定一置信区间,将分布在置信区间内的特征参数作为学习时的特征向量。烟叶产地不同,经过程序计算会自动产生特征向量。一般烟叶样本典型有序,则生成较少的特征参数作为输入向量;如果烟叶样本规律性较差,则会生成较多的特征参数作为特征向量。经过对多个地区的实验,一般特征向量元素在 20~40 之间。

其次,对提供学习的标准样本进行聚类分析,剔除有矛盾数据的样本,以利于提高学习速度。具体采用 ISODATA 算法,根据经验估计出类的数目,在特征空间中将其彼此靠近的样本归属于同一类。两类间距离采用重心间距离公式

$$D(X_i, X_c) = D\left(\frac{1}{N} \sum_{i=1}^N X_i, \frac{1}{M} \sum_{j=1}^M Y_j\right) \quad (1)$$

式中  $X_i, X_c$ ——类别;  $X_i = (x_{i1}, x_{i2}, \dots, x_{in})$ ,  $Y_j = (y_{j1}, y_{j2}, \dots, y_{jn})$ ——表示两类烟叶样本的特征向量,  $n$  表示特征向量的元素个数,即特征空间维数;  $M, N$ ——各类中的样本数。

聚类分析是通过类的合并或类的分裂来达到合理分类。大量样本经过聚类都能很好地实现分类,但当某一类的样本数只有一个或相对很少时,说明该样本特征奇异,无法实现合理归类,这些样本视为奇异样本予以剔除。

#### 3.2 神经网络学习方法

经过聚类分析,可以得到分类,但由于聚类过程是无导师学习,经常将实际一类的样本细分成多类,给检测带来困难,因此需要有目标值的学习方法。

人工神经网络技术是一种模仿人脑活动的技术。它可以通过对样本数据的训练和学习,获得对客观事物的认识和理解。神经网络由诸多神经元组成,接受外界的刺激(即输入信息),经过分析综合传递到下一级神经元。在烟叶分析系统中用  $X = (x_1, x_2, \dots, x_n)$  表示烟叶的各特征值,  $W = (w_1, w_2, \dots, w_n)$  表示相应的连接强度即权重值,  $u$  表示某一神经元输入的总和,  $H$  为神经元本身具有的阈值,该神经元的输出  $Y$  可描述为

$$Y = f(u) = f\left(\sum_{j=1}^n w_j x_j - H\right) = f(W \cdot X - H) \quad (2)$$

$f(u)$  为神经元的输入输出关系的函数,本文用 Sigmoid 函数

$$f(x) = (1 + e^{-x+c})^{-1} \quad (3)$$

式中  $c$ ——常数。神经网络各神经元之间相互连接在一起,组成网络拓扑结构。在烟叶检测系统中采用了误差后向传递网络,是一种有教师的学习算法。

本系统网络拓扑结构,有 3 层,即输入层、中间层和输出层。输入层的节点数为特征空间的维数;中间层可人为调整,本系统取 5~10 之间;输出层的节点数根据目标值个数而定。烟叶的特征值通过输入层进入网络,经过中间层的顺序模式变换,由输出层产生结果。

误差后向传递网络学习算法属于 D 学习律,设某层有  $N$  个处理单元,训练集包含有  $M$  个样本,则根据 (2), (3) 式可计算出第  $p$  个样本 ( $p = 1, 2, \dots, M$ ), 单元  $j$  ( $j = 1, 2, \dots, N$ ) 的输入总和  $u_{pj}$  及输出  $Y_{pj}$ , 如果任意设定网络初始权值,那么对每个样本  $p$ , 网络的输出  $Y_{pj}$  与期望的输出  $d_{pj}$  一般总有误差,其误差为各单元误差之和。

$$E_p = \frac{1}{2} \sum_{j=1}^N (d_{pj} - Y_{pj})^2 \quad (4)$$

$E_p$  称为误差函数。对于 D 学习规则,权重的改变实际上是利用了梯度最速下降法,即使权值沿着误差函数的负梯度方向改变。在学习过程中,学习因子  $G$  的取值越小越好,  $G$  值越大,每次权值修改剧烈,可能会导致学习过程发生振荡。因此,为了使  $G$  取值足

够大, 又不致产生振荡, 在权重修改公式中通常加上一动态项, 即

$$w_{ji}(t+1) = w_{ji}(t) - \alpha E_p(t) \delta v_{ji}(w_{ji}(t)) + A(w_{ji} - w_{ji}(t-1)) \quad (5)$$

式中  $A$ ——动量因子, 它决定了上一次学习的权重变化对本次权重更新的影响程度。

### 4 检测结果与分析

本研究的目的是按照企业标准, 从经过调质处

理后同级别的烟叶中分选出适合高档卷烟配方的合格烟叶。因此, 要将一批烟叶分别归类为合格、两用和不合格三类。其中两用烟叶是根据烟叶货源而确定用或不用的烟叶。为此分别采集辽宁、河南、云南、贵州等地的 C1L (中部柠檬黄一级) 烟和 C1F (中部桔黄一级) 烟进行学习 and 检测。在对标准样本进行学习时, 网络训练误差设置为 0.05。经网络训练后对标准样本的回判结果如表 1 所示, 对未参与学习的大量样本进行检测, 检测结果如表 2 所示。

表 1 标准烟叶样本学习后回判结果

Tab 1 Regrading results of standard tobacco samples after learning

地区	辽 宁			河 南			云 南			贵 州		
吻合率/%	94.0			97.2			96.4			97.0		
类别	合格	两用	不合格	合格	两用	不合格	合格	两用	不合格	合格	两用	不合格
计 算 机	合格	13	-	15	-	-	10	-	-	12	-	-
	两用	1	7	1	11	-	-	9	1	1	11	-
	不合格	-	1	-	-	10	-	-	8	-	-	10
样本总数	14	8	11	16	11	10	10	9	9	13	11	10

表 2 人工评判结果和计算机检测结果对比

Tab 2 Comparison of human being grading results and computer vision grading results

地区	辽 宁			河 南			云 南			贵 州		
吻合率/%	86.4			81.7			92.2			93.1		
人工评判	合格	两用	不合格	合格	两用	不合格	合格	两用	不合格	合格	两用	不合格
计 算 机	合格	33	-	33	4	-	40	2	-	42	-	-
	两用	5	25	6	23	3	3	17	2	5	23	3
	不合格	-	5	-	5	20	-	1	38	-	1	56
样本总数	38	30	13	38	32	23	43	20	40	47	24	59

经过对高档烟配方所用不同地区烟叶的多次检测分析表明, 系统检测性能稳定。与人工评判结果吻合率均在 80% 以上, 半数地区总吻合率在 90% 以上。从实验中同时也发现了一些问题。如果一些不同级别的烟叶样本参与训练, 会使网络无法收敛。经过去除奇异样本后, 网络迅速收敛, 这是因为检测样本不典型所致。系统对颜色较好, 但身份较轻的烟叶检测结果还不够理想。

经过大量的实验, 提出了以下建议: 1) 用于训练的烟叶样本必须具有代表性和规律性, 检测时才能有较好的结果。2) 代表每一类型烟叶的训练样本必须全面且应有一定的数量, 否则会因找不到匹配样本使检测产生波动, 无法获得满意的效果。3) 不同级别的烟叶不能混入训练样本进行学习, 否则会

产生离奇的结果。4) 网络训练误差应根据训练样本数量及样本的典型性、规律性控制在一合适的范围。训练误差不能太小, 否则会导致学习样本吻合率很好而检测样本吻合率很低的后果; 训练误差也不能太大, 否则学不到规律, 导致训练样本和检测样本吻合率都很低。本系统学习误差应在 0.02~ 0.08 之间为宜。5) 应定期检查光线的强度, 以免因采集误差产生系统噪声。

### 5 结 语

计算机视觉技术在烟草领域有着广阔的应用前景。不仅在烟叶的质量分类, 在烟丝的质量监控, 烟叶 40 级质量分级等诸多领域都有潜在的发展前景。

## [参 考 文 献]

- [1] 张建平. 计算机视觉在烟草行业中的研究及应用展望  
烟草科技, 1998, 2: 22~ 23
- [2] 汤顺青主编. 色度学. 北京: 北京理工大学出版社, 1990
- [3] 李孝安, 张晓愤编著. 神经网络与神经计算机导论. 西安: 西北工业大学出版社, 1995

## Application of Computer Vision Technique to Research on Classifying System of Tobacco Leaves

Cai Jianrong Fang Runing Zhang Shiqing Wu Shouyi

(Jiangsu University of Science and Technology, Zhenjiang 212013)

Zhang Jianping

(Shanghai Tobacco (Group) Corp. Technical Center)

**Abstract:** Examining the external quality of tobacco leaves is now mostly relied on the sense of human being. A computer vision system of tobacco leaves grading was introduced in this paper. This system has some subsystems. The demarcation system can control and set the phototonus of CCD camera. The learning system is used to extract and select eigenvalue, then to constitute characteristic vector. It can also remove the freak samples from the standard samples. The neural network is used to learn and grade the tobacco leaves in several areas. And the results of grading indicated that the examination precision of computer vision is all above 80% coincided with grading of human being, and about a half of products from different areas are more than 90%. So this system is highly practical in tobacco leaves garding.

**Key words:** tobacco; computer vision; neural network; quality examination

### 信息技术与农业科技革命

科技部副部长韩德乾在谈信息技术在农业中的应用前景时说,随着农业新科技革命的到来,信息技术在农业中的作用越来越重要。让高科技从贵族化色彩走向平民化,是当前要着力解决的问题之一。

他认为,发展信息技术,按照现在的搞法还不够,必须要有革命性的措施,其中包括:切实建好农业农村信息体系,以严谨态度做好数据库、局域网、互联网工作;进一步提高地理信息系统、全球定位系统、遥感系统技术在农业和农村的应用水平;加强软件产业发展和农业专家智能系统的研究、制作和开发;发展信息传感和自动控制技术,进一步发展精确农业;加强DNA芯片研究;研究开发特殊恶劣情况下不同作用的农业机器人等。

韩德乾副部长强调说,现在不少农民对信息

技术存在神秘感,觉得高科技高不可攀。怎么让高科技摆脱贵族化色彩,解决平民化问题,值得认真研究。例如,把高水平的计算机专家和农业专家结合起来,研究、制作、开发农业专家智能系统,根据粮食作物、经济作物或畜牧业的品种、特性及其所需的最佳环境条件等制作成软件,再根据当地的种植、养殖品种和环境条件,进行软件的第二次开发。农民只要提供自己经营的几个必要数据(如面积、土壤类型、农家肥投入量、资金投入量等),计算机通过智能化处理后,就能打出“明白纸”。“明白纸”经当地农技人员审查后,给农民使用。农民按照这样的“明白纸”发展精确农业,对增产、节支、提高产品质量具有重要意义。

——摘自《科技日报》2000203220