

# 利用高光谱图像技术评判茶叶的质量等级

陈全胜 赵杰文 蔡健荣 Vittayapadung Saritporn

(江苏大学食品与生物工程学院, 江苏 镇江 212013)

**摘要** 针对茶叶品质无损检测时内外品质难以同时兼顾的问题,利用高光谱图像技术检测茶叶质量。设计一套基于光谱仪的高光谱图像系统采集数据;通过主成分分析,从海量数据中优选出三个波长段的特征图像;从每个特征图像中分别提取平均灰度级、标准方差、平滑度、三阶矩、一致性和熵等 6 个基于统计矩的纹理特征参量,每个样本共有 18 个特征变量;再通过主成分分析对这 18 个特征变量进行压缩,提取 8 个主成分因子建立基于反向传播神经网络的茶叶等级判别模型。模型训练时的总体回判识别率为 97%;预测时总体识别率为 94%。结果表明,高光谱图像技术可以用于茶叶质量等级水平的评判。

**关键词** 图像处理;高光谱图像;主成分分析;反向传播神经网络;茶叶;评判

**中图分类号** TP391.41 **文献标识码** A

## Estimation of Tea Quality Level Using Hyperspectral Imaging Technology

Chen Quansheng Zhao Jiewen Cai Jianrong Vittayapadung Saritporn

(School of Food & Biological Engineering, Jiangsu University, Zhenjiang, Jiangsu 212013, China)

**Abstract** The tea quality level was estimated with hyperspectral imaging technology. A hyperspectral imaging system based on spectrometer was developed to perform acquisition of hyperspectral imaging data. The principal component analysis (PCA) was performed to select three optimal bands images. Then, six texture features (i. e., mean, standard deviation, smoothness, third moment, uniformity, and entropy) based on the statistical moment were extracted from each optimal band image, thus 18 variables for each tea sample in all. Finally, PCA was performed again to compress 18 features variables, and 8 principal components (PCs) were extracted as the input of back propagation (BP) neural net. Experimental results showed that discriminating rate was 97% in the training set and 94% in the prediction set. Overall results sufficiently demonstrate that the hyperspectral imaging technology can be used to estimate tea quality level.

**Key words** hyperspectral imaging; principal component analysis; back propagation neural net; tea; estimation

## 1 引 言

茶叶品质的指标是多方面的,即有外部品质指标(色泽和外形等),又有内部品质指标(滋味和香气等)。在茶叶综合品质检测方面,利用单一的检测手段一般不能全面地描述茶叶品质,如近红外光谱可以很好地表征茶叶的内部品质信息,但是,对茶叶外部品质特征,往往显得无能为力<sup>[1]</sup>;机器视觉技术能很好地表征茶叶的外部品质特征,但是无法获取反

映茶叶内部品质的有效信息<sup>[2]</sup>,信息反映的侧重点不同所带来的局限性必然影响到检测结果的精度和稳定性。所以,如何充分利用各种检测方法的长处,相互结合,优势互补,以提高检测的全面性、可靠性和灵敏度,是茶叶品质快速无损检测一个新的研究趋势。

目前,一种能集成光谱检测和图像检测优点的新技术——高光谱图像技术正好能满足茶叶内外综

**收稿日期:** 2007-06-19; **收到修改稿日期:** 2007-09-28

**基金项目:** 国家 863 计划(2006AA10Z263)和江苏省自然科学基金重点资助项目(BK2006707-1)资助课题。

**作者简介:** 陈全胜(1973—),男,安徽桐城人,博士,讲师,主要从事近红外光谱及高光谱图像技术在食品和农产品品质检测中应用等方面的研究。E-mail: chenjiang0518@yahoo.com.cn

**导师简介:** 赵杰文(1945—),男,江苏苏州人,教授,博士生导师,主要从事农产品加工、储藏、保鲜、品质检测、图像处理以及农业机器人视觉系统等方面的研究。E-mail: zhao\_jiewen@uj.s.edu.cn

合品质同时检测的需要。它是光谱分析技术和图像处理技术在最低层面上的融合技术,兼有这两种技术的优势,可以对研究对象的内外部品质特征进行可视化分析。近几年来有部分学者将该技术应用于肉类和果蔬类质量品质检测上<sup>[3~5]</sup>。该技术在茶叶品质检测中的应用,既可以克服光谱信息和图像信息在融合过程中存在的不足,又能兼顾到茶叶外部品质和内部品质的同时检测。

## 2 高光谱图像基本原理

高光谱图像是在特定波长范围内由一系列波长处的光学图像组成的三维图像块<sup>[6~8]</sup>。图 1 为高光谱图像三维数据块。 $x$  和  $y$  表示二维平面像素信息坐标轴, $\lambda$  是波长信息坐标轴。可以看出,高光谱图像既具有某个特定波长下的图像信息,又具有某个特定像素在不同波长下的光谱信息。其中,光谱信息能充分反映茶叶内部有效成分(包括多酚类、植物碱、氨基酸和蛋白质等含氢基团成分)的特征信息,这些特征信息又与茶叶的滋味和香味等品质因子密切相关;图像信息能充分反映茶叶的外部品质特征。因此,利用高光谱图像技术可以检测茶叶的内外综合品质。

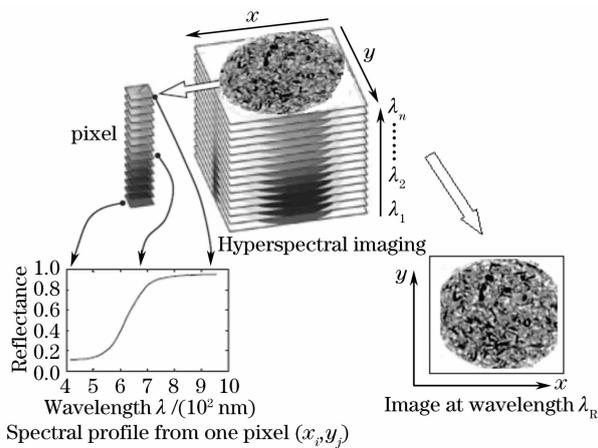


图 1 高光谱图像数据块

Fig. 1 Hyperspectral imaging datacube

高光谱图像是由一系列特定波长下的光学图像组成的三维数据块,特定波长的光可通过滤光片和光谱仪两种方式获得。因此,根据特定波长光获取的方式不同,可以分为基于滤光片的高光谱图像系统和基于光谱仪的高光谱图像系统。基于光谱仪的高光谱图像系统采集得到的数据精度高,可以用于寻找检测目标所需要的特定波长,但其数据量超大,数据处理时间长,仅适合于实验室研究阶段。反之,基于滤光片的高光谱图像系统采集得到的数据量

小,数据分析所需要的时间短,适合在线检测,但是数据过于简单,很难寻找到检测目标所需要的特征波长。一般情况下,首先在实验室阶段利用基于光谱仪的高光谱图像系统寻找检测目标所需要的最优波长,然后根据最优波长设计基于滤波片的高光谱图像系统,以实现在线检测。本实验属于实验室阶段的研究,研究的目的是验证高光谱图像检测茶叶品质等级的可行性,寻找检测茶叶品质等级的特征波长图像,因此采用基于光谱仪的高光谱系统。

## 3 材料与方法

### 3.1 实验材料

实验所用的材料为购自镇江市建林绿茶公司的炒青绿茶,分 4 个级别,茶叶的出厂期都为 2006 年 4 月。实验从每个等级中随机称取  $20 \pm 0.5$  g 茶叶作为一个样本,每个级别的茶叶分别选取 50 个样本,4 个级别总共 200 个样本。

### 3.2 实验装置

本实验的高光谱图像数据是基于光谱仪的高光谱图像系统采集得到的。图 2 为基于图像光谱仪的高光谱图像采集系统,由基于图像光谱仪的高光谱摄像机(ImSpector, V10E, 芬兰),一套 150 W 的光纤卤素灯(Fiber-Lite DC950 Illuminator, Dolan-Jenner Industries Inc, MA, 美国),一套输送装置(Zolix, SC300-1A, 北京)和计算机等部件组成。

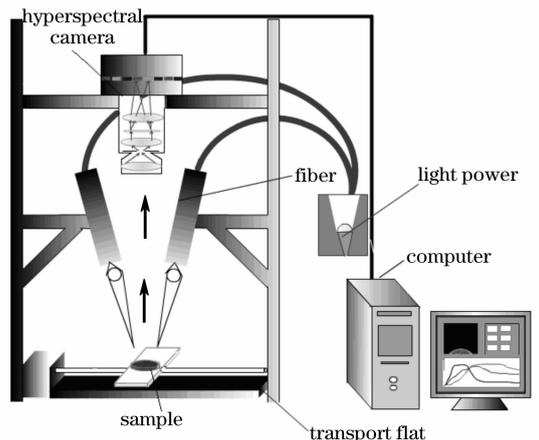


图 2 基于光谱仪的高光谱图像系统

Fig. 2 Hyperspectral image system based on spectrometer

### 3.3 高光谱图像采集

每次称取  $20 \pm 0.5$  g 的茶叶作为一个样本,将茶叶均匀地平铺在规格为  $\phi 10$  cm  $\times$  1 cm 的培养皿中,然后进行高光谱图像采集。在高光谱图像数据采集前,预先确定好高光谱摄像头的曝光时间以保证图像的清晰;确定好输送装置的速度以避免图像尺寸和空

间分辨率失真。曝光时间为 30 ms, 输送装置的速度为 1.25 mm/s。数据采集时, 线阵的探测器在光学焦面的垂直方向作横向扫描( $x$  轴方向), 获取条状空间中每个像素在各个波长处的光谱信息; 同时随着样本的纵向前进( $y$  轴方向), 线阵探测器扫出整个平面完成整幅茶叶样本图像数据的采集。采用的高光谱摄像头的图像分辨率为  $600 \times 1280$ , 光谱范围是  $408 \sim 1117$  nm, 光谱分辨率为 2.8 nm, 采样间隔为 0.67 nm, 采集得到 1024 个波长下的图像, 最终得到一个大小为  $600 \times 1280 \times 1024$  的高光谱图像数据块。

### 3.4 高光谱图像标定

由于光源的强度在各波段下分布不均匀以及摄像头中暗电流噪声的存在, 造成在光源强度分布较弱的波段下, 所获得的图像含有较大的噪声<sup>[9]</sup>。因此, 需要对所获得的高光谱图像进行黑白标定。在与样品采集相同的系统条件下, 扫描标准白色校正板得到全白的标定图像  $W$ , 关闭相机快门进行图像采集得到全黑的标定图像  $B$ , 完成高光谱图像的标定, 使采集得到的绝对图像  $I$  变成相对图像  $R$

$$R = \frac{I - B}{W - B} \quad (1)$$

所有高光谱图像数据的采集都是基于 SpectralCube (Spectral Imaging Ltd., Finland) 软件平台; 数据的处理是基于 ENVI V. 4. 3 (Research System, Inc., USA) 和 Matlab V. 7. 0 (Mathworks Co., USA) 软件平台。

## 4 结果与讨论

由高光谱原理可知, 高光谱图像上的每个像素都有光谱信息。图 3 为样本图像上的不同像素在  $408 \sim 1117$  nm 范围内的光谱曲线, 其中, 上面的 4

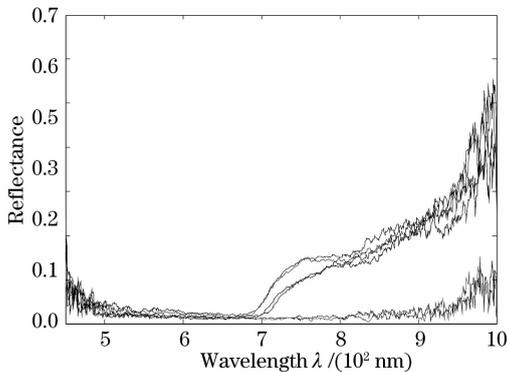


图 3 茶叶高光谱图像不同区域的光谱曲线

Fig. 3 Spectral profile from the different regions in the hyperspectral image of tea samples

条曲线表示茶叶的光谱曲线, 下面的两条表示背景的光谱曲线。从图 3 可看出, 在 700 nm 以下区域茶叶的光谱曲线与背景的光谱曲线接近; 而 850 nm 以上, 这些区域的光谱曲线存在明显噪音。因此, 在后期的数据处理过程中, 选取 700~850 nm 范围内的高光谱图像数据进行分析。

### 4.1 特征波长的选取

实验选取 700~850 nm 范围内的高光谱图像数据进行分析, 分析前, 将原始高光谱图像数据在 ENVI 上进行重采样, 得到一个  $600 \times 600 \times 217$  的三维数据块。它是由 217 张波长在 700~850 nm 范围内且大小为  $600 \times 600$  的图像组成, 它包含的数据量比二维图像和一维光谱的数据量都要大得多。数据量过大影响后期数据处理的速度; 波段相邻的两幅图像之间又具有较强的相关性, 造成高光谱图像数据中存在大量的冗余信息<sup>[10]</sup>。因此, 有必要对高光谱数据进行降维处理以寻找最能表征茶叶品质指标的特征图像, 这样即可以提高后期数据处理的速度, 又可以去除数据中的冗余信息。

主成分分析是沿着协方差最大的方向由高维数据空间向低维数据空间投影。主成分分析得到的各个主成分向量之间互相独立, 即可以实现降维, 又能消除原始数据中的冗余信息。实验通过主成分分析来优选特征波长, 根据方差贡献率的大小提取前面几个主成分图像, 从中找到最能表征原始信息的主成分图像。每个主成分图像都是由原始数据中的各个波长下的图像经过线性组合而成的, 通过比较线性组合中的权重系数, 最大权重系数所对应波长下的图像为最佳特征图像<sup>[11, 12]</sup>。

图 4 为经过主成分分析后得到的前 4 个主成分

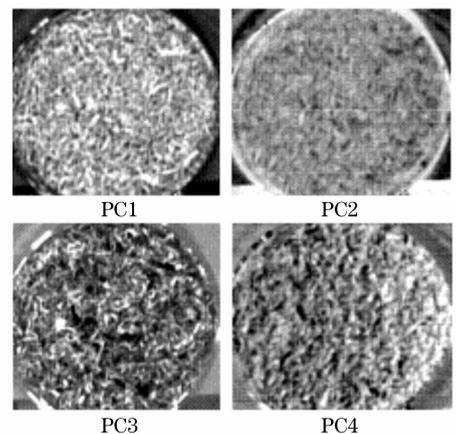


图 4 主成分分析得到的前 4 个主成分图像

Fig. 4 Front top four principal component score images with principal component analysis

图像 PC1、PC2、PC3 和 PC4。可以看出第一主成分图像 PC1 最能表征茶叶图像的原始信息,因此,根据 PC1 寻找特征波长图像。PC1 是由 217 个波段下的图像经过线性组合而成的:

$$PC1 = \alpha_1 \cdot Im_1 + \dots + \alpha_{91} \cdot Im_{91} + \dots + \alpha_{136} \cdot Im_{136} + \dots + \alpha_{200} \cdot Im_{200} + \dots + \alpha_{217} \cdot Im_{217}. \quad (2)$$

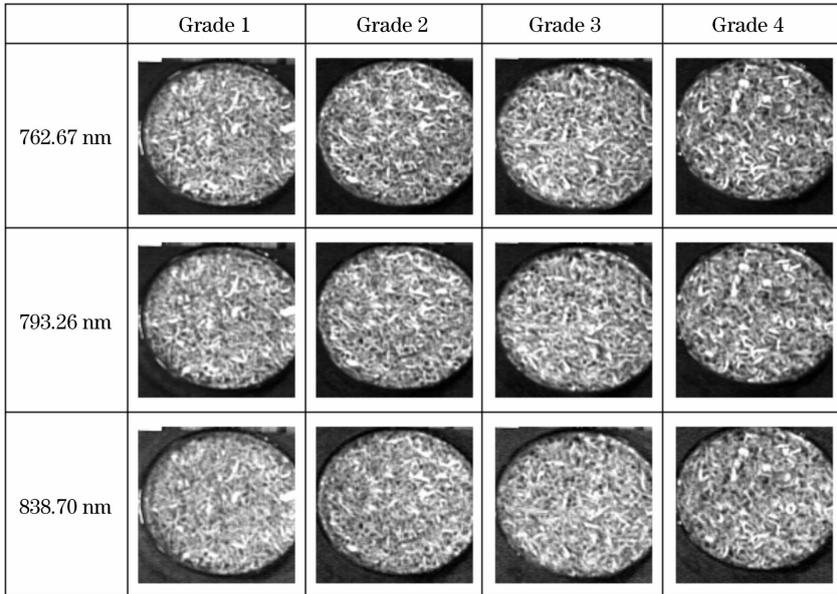


图 5 四个等级的茶叶在三个特征波长下的图像

Fig. 5 Four grades tea samples images at three feature bands

## 4.2 特征提取

本文采用了基于灰度统计矩方法来计算平铺茶叶图像的纹理,提取三个特征波长下的纹理特征作为特征变量。对每个特征波长图像,找到图像的中心位置点,然后以该点为中心截取一个  $300 \times 300$  的中心区域,然后按照(3)式~(8)式分别提取平均灰度级  $m$ 、标准方差  $\delta$ 、平滑度  $R$ 、三阶矩  $\mu_3$ 、一致性  $U$  和熵( $e$ )等 6 个基于统计矩的纹理特征参量

$$m = \sum_{i=0}^{L-1} z_i p(z_i), \quad (3)$$

$$\delta = \sqrt{\sum_{i=0}^{L-1} (z_i - m)^2 p(z_i)}, \quad (4)$$

$$R = 1 - 1/(1 + \delta^2), \quad (5)$$

$$\mu_3 = \sum_{i=0}^{L-1} (z_i - m)^3 p(z_i), \quad (6)$$

$$U = \sum_{i=0}^{L-1} p^2(z_i), \quad (7)$$

$$e = - \sum_{i=0}^{L-1} p(z_i) \ln p(z_i). \quad (8)$$

由高光谱的原理可知,在特定的波长下,图像上

在该线性组合中,权重系数越大,所对应波长下的图像对 PC1 图像的贡献越大,通过比较这 217 个权重系数,前 3 个较大的权重系数分别是  $\alpha_{136}$ 、 $\alpha_{200}$  和  $\alpha_{91}$ ,它们所对应的波长分别为 793.26 nm, 838.70 nm 和 762.67 nm。因此,根据主成分分析优选出这三个波长下的图像作为特征图像,图 5 为 4 个级别的茶叶在这 3 个特征波长下的图像。

点的灰度值与该点在该波长下的光谱值一一对应,也就是说,图像上每个像素在特定波长下的光谱值可以通过特定波长下图像的灰度值来表现。茶叶中含有大量与品质相关的有效成分(如氨基酸、多酚类和植物碱等),这些有效成分大多存在含氢基团(C—H、O—H 和 N—H 等),能在某些特定波长下产生倍频与合频吸收。茶叶品质质量不同,其内部有效成分含量与比例可能存在差异,这种差异在特定波长下表现为光的吸收强度不同,即表现为不同的光谱值,又由于图像上每个像素在特定波长下的光谱值可以通过特定波长下图像的灰度值来表现。因而不同质量等级的茶叶,其内部有效成分的差异可以通过某特定波长下的图像灰度等级来体现。在某个特定波长下,图像的平均灰度级  $m$  可以用来表示茶叶样本上很多点的平均光谱值,因而,特征图像上的平均灰度级  $m$  能在一定程度上表征茶叶内部的品质信息。茶叶的等级质量也与茶叶的外部品质有一定的关系,可以通过其它 5 个纹理特征来描述茶叶的外部品质。例如一级茶叶的外部形状相对比较均匀,在特征图像中,灰度分布的标准方差  $\delta$ 、平

滑度  $R$  和熵  $e$  相对较小,一致性  $U$  相对较大,三阶矩  $\mu_3$  趋于 0。

#### 4.3 模型建立与结果

在模型建立之前,首先从所用的 200 个样本中提取 100 个样本进行训练建立识别模型,另 100 个样本用来验证该模型的可靠性。本文采用反向传播神经网络作为模式识别的方法来建立识别模型,该反向传播神经网络模型的输入层单元数为主成分特征变量个数,它的输出层单元数为 1,即茶叶的等级;选择非线性的 Sigmoid 型函数作为模型的作用函数;训练过程中的学习因子和动量因子都置为 0.1,模型中间层单元数置为 8,此时模型的性能最优。

针对每个样本,这 18 个特征变量是从 3 个特征波长下的图像中提取的,它们之间存在一定的相关性。所以在模型建立之前,有必要对此进行主成分分析,提取主成分的分向量构成模式识别的输入。模型在训练过程中,主成分变量的多少会影响到模型的精度稳定性,主成分因子数过少,信息损失过大,将会影响到模型的精度;但是主成分因子数过大,又会引入过多的冗余信息,影响到模型的稳健性,同时使数据处理的时间加长。因此,选择合适的主成分因子数对模型的建立很重要。

图 6 为在不同的主成分因子数下反向传播神经网络模型训练和预测的结果,可以看出,开始时,随着主成分因子数的增加,训练集和预测集中的识别率总体上呈现上升趋势,当主成分因子数达到 8 以后,模型的识别率都已经变化不大,甚至有下降的趋势。

由于前 8 个主成分所对应的累计方差贡献率达到 99.75%,能解释原始数据的 99.75% 信息,且它们之间相互独立,消除冗余信息,提高了模型的稳定性,因此,18 个特征变量经过主成分分析后,提取的前 8 个主成分因子作为反向传播模型的输入变量。经过 513 次迭代,最终得到 1 个  $8 \times 8 \times 1$  的反向传播神经网络模型,训练耗时为 200 ms。

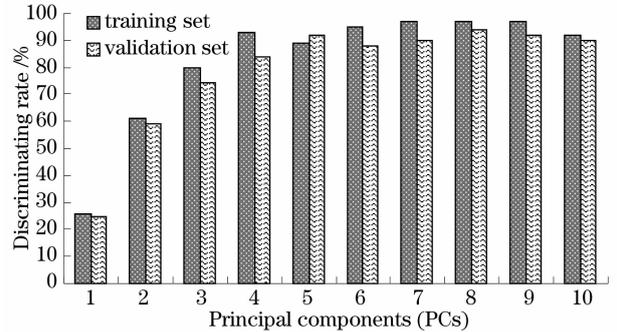


图 6 不同主成分因子数下的训练和预测时的识别结果

Fig. 6 Discriminating results in training set and prediction set with different PCs

表 1 和表 2 分别为各个级别样本的回判和预测的结果。从表 1 中可以看到,模型在训练时,将 1 个 1 级样本误判为 2 级,将 1 个 2 级样本误判为 3 级,将 1 个 4 级样本误判为 3 级;总体回判率为 97%。从表 2 中可以看到,模型在预测时,将 3 个 2 级样本分别误判为 1 级、3 级和 4 级,1 个 3 级样本误判为 2 级,2 个 4 级样本分别误判为 2 级和 3 级;总体识别率达 94%。表 1 和表 2 的结果表明模型的识别率和稳定性都达到一个很高的水平。

表 1 训练集中四个级别茶叶的判别结果

Table 1 Discriminating results of four grades of tea in training set

Grade	Sample number	Recognition result				Total recognition rate
		Grade 1	Grade 2	Grade 3	Grade 4	
Grade 1	25	24	1	0	0	97%
Grade 2	25	0	24	1	0	
Grade 3	25	0	0	25	0	
Grade 4	25	0	0	1	24	

表 2 预测集中四个级别茶叶的预测结果

Table 2 Discriminating results of four grades of tea in prediction set

Grade	Sample number	Recognition result				Total recognition rate
		Grade 1	Grade 2	Grade 3	Grade 4	
Grade 1	25	25	0	0	0	94%
Grade 2	25	1	22	1	1	
Grade 3	25	0	1	24	0	
Grade 4	25	0	1	1	23	

由表 1 和表 2 可以看出,发生误判的样本大部分集中在两个相邻的级别之间,这是由于实验材料中茶叶样本的级别是由专业人员通过感官评判得到的,其评判的结果受主观因素的影响,特别在两个相邻的等级之间,它们的品质指标往往十分相近。这些都会使模型的识别结果产生一定的误差,而这种误差又大部分集中在两个相邻级别的样本中。

## 5 结 论

利用一种新的无损检测技术——高光谱图像技术检测茶叶质量,以 4 个等级的炒青绿茶为实验对象。首先,搭建一套基于图像光谱仪的高光谱图像采集系统用于茶叶的高光谱图像数据采集;通过对茶叶高光谱数据进行主成分分析,从海量数据中优选出三个特征波长下的图像;然后,从每个特征波长下的图像中,分别提取 6 个基于灰度统计矩的纹理特征变量,共 18 个特征变量;最后,再对这 18 个变量进行主成分分析,提取 8 个主成分因子数建立反向传播模型最佳。模型训练时的总体回判率为 97%;预测时总体识别率为 94%。研究结果表明,高光谱图像技术可以用于茶叶质量等级的评判。

本实验是属于实验室研究阶段,基于光谱仪的高光谱图像系统采集得到的数据量大,处理时间长,并不适合在线检测。但是,通过该实验可以得到的最能表达茶叶品质等级的三个特征波长,然后根据相应的特征波长,设计基于滤波片式的高光谱图像系统以实现茶叶生产流通过程中的在线分级,避免滤波片波长范围选择的盲目性。

## 参 考 文 献

1 Chen Quansheng, Zhao Jiewen, Zhang Haidong *et al.*. Identification of Authenticity of tea with near infrared spectroscopy based on support vector machine[J]. *Acta Optica*

*Sinica*, 2006, **26**(6): 933~937  
 陈全胜,赵杰文,张海东等. 基于支持向量机的近红外光谱鉴别茶叶的真伪[J]. *光学学报*, 2006, **26**(6): 933~937  
 2 Chen Quansheng, Zhao Jiewen, Zhang Haidong *et al.*. Identification of tea color by using computer vision [J]. *J. Jiangsu University (Natural Science Edition)*, 2005, **26**(6): 1~4  
 陈全胜,赵杰文,张海东等. 利用计算机视觉识别茶叶的色泽类型[J]. *江苏大学学报(自然科学版)*, 2005, **26**(6): 1~4  
 3 M. S. Kim, Y. R. Chen, P. M. Mehl. Hyperspectral reflectance and fluorescence imaging system for food quality and safety[J]. *Trans. ASAE*, 2001, **44**(3): 721~729  
 4 B. Park, K. C. Lawrence, W. R. Windham *et al.*. Hyperspectral imaging for detecting fecal and ingesta contaminants on poultry carcasses [J]. *Trans. ASAE*, 2002, **45**(6): 2017~2026  
 5 Patrick M. Mehl, Yud-Ren Chen, Moon S. Kim. Development of hyperspectral imaging technique for the detection of apple surface defects and contaminations [J]. *J. Food Engineering*, 2004, **61**(1): 67~81  
 6 Yud-Ren Chen, Kuanglin Chao, Moon S. Kim. Machine vision technology for agricultural applications [J]. *Computers and Electronics in Agriculture*, 2002, **36**(2): 173~191  
 7 Peixin Hu, Qimin Lu, Rong Shu *et al.*. An airborne pushbroom hyperspectral imager with wide field of view [J]. *Chin. Opt. Lett.*, 2005, **3**(12): 689~691  
 8 Liu Muhua, Zhao Jiewen, Zheng Jianhong *et al.*. Review of hyperspectral imaging in quality and safety inspections of agricultural and poultry products [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2005, **35**(9): 139~143  
 刘木华,赵杰文,郑建鸿等. 农畜产品品质无损检测中高光谱图像技术的应用进展 [J]. *农业机械学报*, 2005, **36**(9): 139~143  
 9 Gerrit Polder, Gerie W. A. M., Van Der Heijden *et al.*. Calibration and characterization of imaging spectrographs [J]. *J. Near Infrared Spectroscopy*, 2003, **11**(3): 193~210  
 10 Jingwen Yan, Jiazhen Chen. SFCVQ and EZW coding method based on Karhunen-Loeve transformation and integer wavelet transformation [J]. *Chin. Opt. Lett.*, 2007, **5**(3): 153~155  
 11 Xun Lina, Fang Yonghua, Li Xin. A small target detection approach based on endmember extraction in hyperspectral image [J]. *Acta Optica Sinica*, 2007, **27**(7): 1178~1182  
 寻丽娜,方勇华,李新. 高光谱图像中基于端元提取的小目标检测算法 [J]. *光学学报*, 2007, **27**(7): 1178~1182  
 12 M. S. Kim, A. M. Lefcourt, Y. R. Chen *et al.*. Multispectral detection of fecal contamination on apples based on hyperspectral imagery: Part II. Application of hyperspectral fluorescence imaging [J]. *Trans. ASAE*, 2002, **45**(6): 2039~2047