

基于纹理分析的茶青在线分类^①

汤 哲^{②*} 江才华^{③*} 张 立^{**} 周建勇^{**}

(* 中南大学信息科学与工程学院 长沙 410083)

(** 长沙湘丰茶叶机械制造有限公司 长沙 410100)

摘 要 将基于纹理分析和支持向量机(SVM)的分类技术应用于茶青的自动分类,实现了茶青的快速、准确在线分类。通过数字图像处理技术提取茶青的特征参数,并通过 SVM 建立不同茶叶的鉴别模型,实现不同茶青的快速在线分类。通过茶青图像的灰度共生矩阵(GLCM)提取出能量、相关性、对比度、逆差距 4 个纹理特征参数特征值,并将提取出的纹理特征值作为 SVM 的输入参数进行训练与分类。实验表明,将基于 SVM 的纹理分类方法应用于茶青的分类,能够取得很好的效果,分类判断的准确率达到 90%。

关键词 茶青分类,灰度共生矩阵(GLCM),纹理分析,图像分类,支持向量机(SVM)

0 引 言

目前的茶青分级可以通过传统的鲜叶分级机对茶叶进行初级分类,但分类精度并不是很高,分类后一般还是将单芽头与一芽一叶的混合在一起,一芽一叶和一芽两叶的混合在一起,一芽两叶与一芽三叶、多叶的混合在一起。此外,传统的茶青分级机由于本身结构的限制,不能实现自动分级与茶叶自动化加工生产线连续生产的要求。根据茶叶加工品质的要求,不同类型的茶叶加工参数要求不一致,比如单芽头的茶叶杀青温度比一芽一叶的高,一芽一叶又比一芽两叶的高,同样对于分选风力大小、揉捻时间、摊凉时间、滚炒时间和温度等参数,不同类型的茶叶也都不一样,而茶叶自动化生产线加工要满足名优茶在外形、汤色和含水率等关键指标的要求,就必须准确知道被加工茶青中每一种茶叶的比例。但目前还没有对茶青进行自动分类与识别的有效方法,对茶叶加工的品种造成很大的影响,不利于茶叶附加值的提高。随着计算机技术的发展,人们已将智能技术和计算机视觉技术广泛应用于食品和农产品的加工、分类与检测过程中。目前也有一些利用

视觉技术进行茶叶成本自动分级的研究,但利用计算机视觉技术对茶青进行分类识别与分类加工的相关研究与应用还很少。文献[1]利用多光谱成像的机器视觉技术及组合特征(形状特征、纹理特征)进行了茶叶等级区分,但它只能区分单片茶叶,不能进行成批茶叶的分类,并且分类效率很低。文献[2]利用基于支持向量机(SVM)的红外线光谱技术进行了成品绿茶、红茶、乌龙茶的分类,分别使用不同的核函数(径向基、多项式)进行了分析,取得了很好的效果。文献[3]对多光谱图像与 SVM 技术进行了研究,研究表明,SVM 的核函数为径向基时,能更准确、有效、快速地进行茶叶的分类。虽然该研究取得了很好的效果,但针对的是茶叶的成品。文献[4]将人工神经网络技术应用于鲜茶叶的分类,虽然取得良好的效果,但是它所需的特征参数太多,影响了茶叶分类的效率。考虑到茶青的纹理特征鲜明,使得将计算机视觉技术应用于茶叶的分类成为可能,再加上 SVM 在处理高维样本空间问题时,能够有效地避免维数灾难,并且所需样本少,具有良好的容错性、灵活性与自适应性。本文主要研究了基于纹理分析和支持向量机的茶青在线分类,并给出了研究的结果与分析。

① 国家自然科学基金(61103035)和创新基金(12C26214304866),湖南省自然科学基金(12JJ5026),湖南省科技计划(2012GK2010)和长沙市科技计划(K1306035-11)资助。

② 男,1977年生,博士,副教授;研究方向:自动化、智能控制和人形机器人;E-mail: tz@csu.edu.cn

③ E-mail: 819479374@qq.com

(收稿日期:2013-12-25)

1 基于纹理分析的茶青在线分类系统的架构

茶青在线分类系统的基本原理是:茶青由输送带输送到摄像头下方,在人工光源照明下,图像传感器采集到合适的检测图像,由图像采集卡将其转化为数字信号保存在计算机内,再通过相关的算法对图像进行处理,从而得到所需要的各种目标图像纹理特征值,并在此基础上实现茶青的在线分类,最后用得到的茶青比例发送到茶叶生产线的数据采集与监控系统(SCADA)来控制茶叶的加工参数。茶青在线分类示意图如图 1 所示。

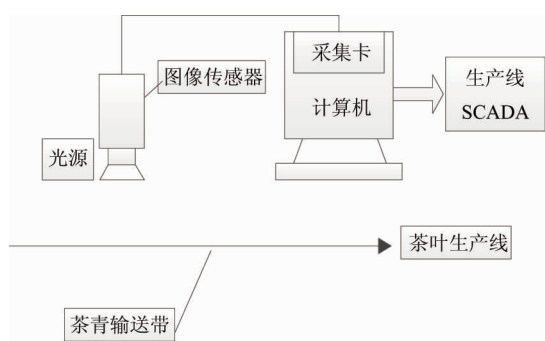


图 1 茶青在线分类示意图

基于纹理分析的茶青在线分类系统具体分成茶青采集、茶青图像预处理、茶青纹理特征值提取、茶青样本训练与分类等部分进行处理,基于纹理分析的茶青在线分类算法流程如图 2 所示。

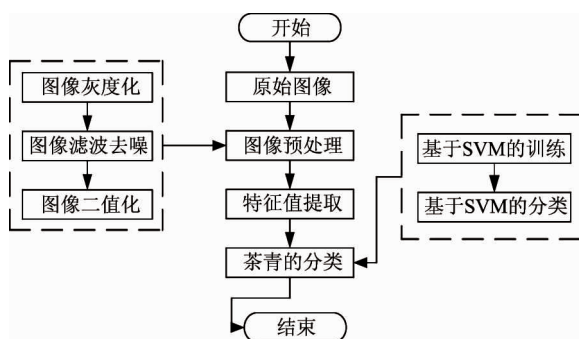


图 2 茶青分类的算法流程图

2 基于灰度共生矩阵的纹理特征提取

在纹理特征提取方式中,基于灰度共生矩阵(gray-level co-occurrence matrix, GLCM)是最有效的提取方法,具有较强的适应能力和鲁棒性,比其它方法更能准确地进行纹理提取。基于 GLCM 是一个

比较经典的统计分析方法,它是通过两个相邻位置像素点的联合概率密度函数建立的,以条件概率来提取纹理的相关特征,并统计处于同样位置关系的一对像元的灰度相关性,并用这一对像元的某种灰度的条件概率来表示纹理。文献[5]采用了基于 GLCM 与感知器分类技术进行了成品茶叶的分类。GLCM 提供了图像灰度间隔、方向和变化幅度的信息,从而提取用来定量描述纹理特征的统计属性,在本文中,用到的统计量有能量、相关性、对比度、逆差距^[6]等。

能量反映的是纹理均匀分布程度与粗细程度,如果 GLCM 越分散则能量值越小,GLCM 越集中则能量值越大。它的计算公式为

$$\sum_{i,j} p(i,j)^2 \quad (1)$$

相关性表示 GLCM 在行列上的相似程度,GLCM 越均匀相关值越大,它的计算公式为

$$\frac{\sum_{i,j} ijp(i,j) - u_x u_y}{\sigma_x \sigma_y} \quad (2)$$

对比度则表示图像的清晰度和纹理沟纹深浅度,纹理沟纹越深,值越大。它的计算公式为

$$\sum_{i,j} (i-j)^2 p(i,j) \quad (3)$$

逆差距反映的是图像纹理同质性,它主要是通过度量图像纹理局部变化的多少来体现。如果它的值越大,则说明图像纹理的不同区域间缺少变化、局部均匀,它的计算公式为

$$\sum_{i,j} \frac{1}{1 + (i-j)^2} p(i,j) \quad (4)$$

其中 $p(i,j)$ 表示矩阵第 i 行、第 j 列的像素值, u 为均值, σ 为方差。

本研究主要对茶青纹理能量、相关性、对比度、逆差距 4 个特征值进行了提取,并进行了分析与比较。图 3 是采集的茶青的样本,由于样本数量很多,所以图 3 显示了具有代表性的不同形状的茶青。

在本文实验中,计算机硬件环境大致如下:CPU 为 Intel(R) Core(TM) i5-2430M;内存为 2G;系统为 32 位操作系统;硬盘为 500G;软件环境主要是通过 Matlab2011 进行仿真。在本文中,GLCM 选择了 0° 、 45° 、 90° 、 135° 四个方向,距离分别为 1、2、3、4 像素点的共生矩阵,分别提取能量、对比度、相关性、逆差距四个特征量,特征值可分别见于图 4、图 5、图 6 和图 7,其中 Offset = [0 1; 0 2; 0 3; 0 4; -1 1; -2 2; -3 3; -4 4; -1 0; -2 0; -3 0; -4 0; -1 -1; -2 -2; -3 -3; -4 -4]。实验表明,根据

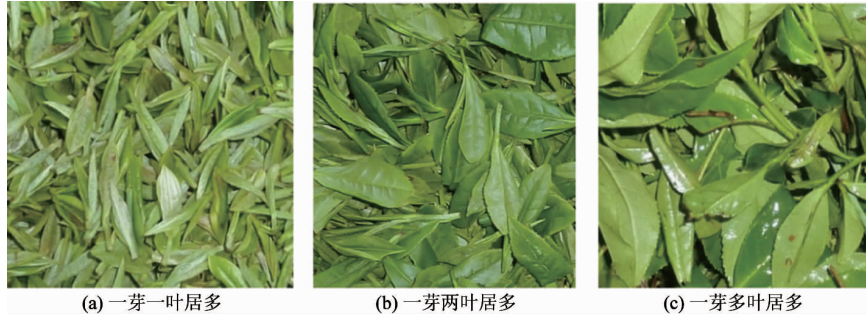


图3 茶青的样本

纹理相关性特征与逆差距特征能够很好地区分茶青的等级,当相关性特征值大于0.82及逆差距特征值在0.62至0.72之间时,茶青中一芽多叶所占的比例居多,这种类型的茶青比较粗糙,能够加工成次等茶;当相关性特征值在0.76至0.82之间及逆差距特征值在0.58至0.65之间时,茶青中一芽两叶所占的比例居多,能够加工成优等茶;当相关性特征值在0.72至0.76之间及逆差距特征值在0.52至0.6之间时,茶青中一芽一叶所占的比例居多,这种类型的茶青比较嫩,能够加工成名优茶。

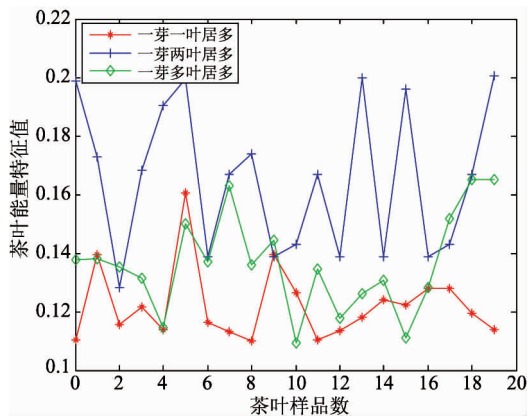


图4 不同茶青所占比例的能量特征值

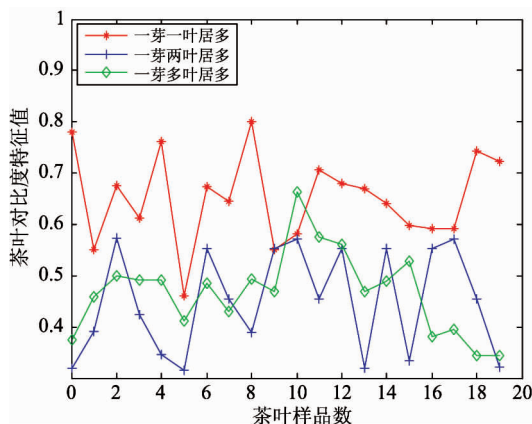


图5 不同茶青所占比例的对比度特征值

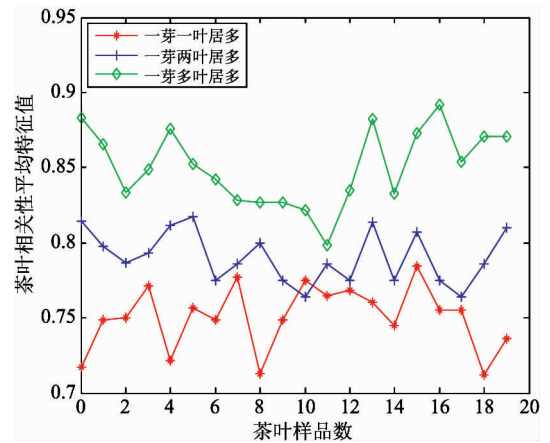


图6 不同茶青所占比例的相关性特征值

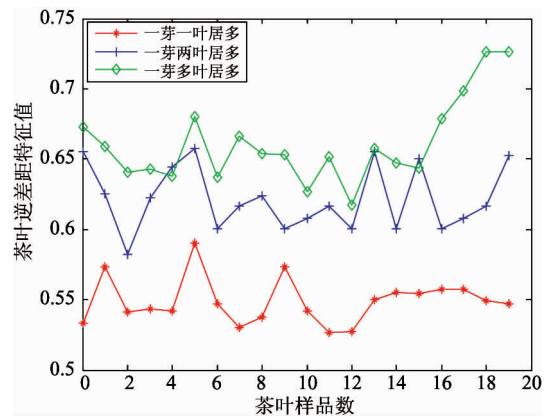


图7 不同茶青所占比例的逆差距特征值

3 基于支持向量机的茶青纹理分类

纹理图像的分类是指将图像中提取的纹理特征值,按照一定的决策规则归类到已知的某一纹理类别中。纹理分类器包括监督式纹理分类器、无监督式纹理分类器和基于支持向量机(SVM)的纹理分类。监督式纹理分类器的设计首先需要确定判别函数的结构和决策规则;然后根据一定数量的已知类别样本进行判别函数中参数的设定;最后对未知类别的纹理样本进行特征提取和选择,通过判别函数

的计算结果,并根据决策规则完成对未知类别的纹理分类。它的设计方法主要包括:以最小错误率贝叶斯决策为代表的概率判别规则、fisher 分类器的线性分类方法^[7]和以最近邻法的距离判别规则^[8]。无监督式纹理分类器设计需要选定某种距离度量作为纹理特征间的相似度量,同时需要确定准则函数用来评价聚类结果质量,然后给定纹理特征的某个初始分类,最后用迭代算法找出使准则函数取极值的最优纹理分类结果。它主要包括基于近邻法则的 K 均值聚类^[9]和 ISODATA 聚类。

SVM^[10]是一种基于结构风险最小化的统计学习方法,它的基本思想是将分类问题通过非线性变换将输入空间变换到一个高维的特征空间,其中非线性变换主要是通过核函数得到的,然后在新高维特征空间中重新构造线性判别函数来实现原空间中的非线性判别函数,求取最优线性分类面,从而实现非线性变换后的线性分类。常用的核函数有以下几种:

线性核函数:

$$K(x_i, x_j) = (x_i \cdot x_j) \quad (5)$$

多项式核函数:

$$K(x_i, x_j) = (x_i \cdot x_j + c)^d \quad (6)$$

径向基核函数:

$$K(x_i, x_j) = \exp(-g \|x_i - x_j\|^2) \quad (7)$$

Sigmoid 核函数:

$$K(x_i, x_j) = \tanh[k(x_i \cdot x_j) + c] \quad (8)$$

其中, x_i, x_j 分别是原空间中的向量, k 为一个标量, c 为位移参数。上面的核函数所对应的分类器一般被称为标准 SVM, 图 8 表示了 SVM 的原理与结构, 先把数据从输入空间映射到特征空间, 然后在特征空间构造线性判决函数。

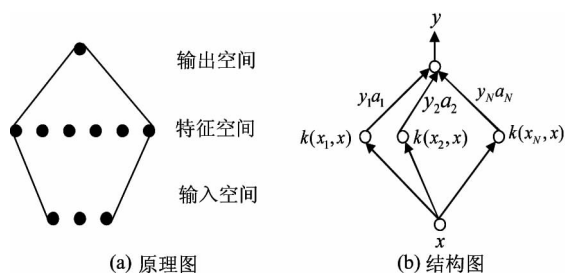


图 8 支持向量机原理与结构图

SVM 类似于人工神经网络, 需要一个训练过程, 将图像的纹理特征值作为 SVM 的输入参数进行训练获得支持向量, 然后将测试集输入到训练好的 SVM 中, 可以得到最后分类结果。通过上文茶青纹

理特征值的提取效果图可知, 纹理相关性特征值与逆差距特征值对茶青的分类有比较好的效果, 所以在基于 SVM 的纹理分类中, 主要是用到茶青的纹理相关性特征值与逆差距特征值进行有效的结合。

在本文实验中, SVM 的核函数选择了径向基核函数(RBF), 误差评测标准为 K 折交叉确认误差, 惩罚因子为 1000。目标函数为:

$$\min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l \zeta_i \quad (i = 1, 2, \dots, l) \quad (9)$$

其中 $\zeta_i \geq 0$ ($i = 1, 2, \dots, l$), l 为样本数, ζ_i 为松弛变量。

下面分别对不同形状的茶青进行训练与分类, 不同形状茶青的训练集如图 9 ~ 图 11 所示。由图可知, 茶青中形状不同的混合茶青都能够很容易进行区分。茶青训练完后, 需要对茶青样本进行分类, 不同形状茶青的分类集如图 12 ~ 图 14 所示。实验结果表明, 将 SVM 技术应用到茶青分类中能够取得比较准确的分类效果。

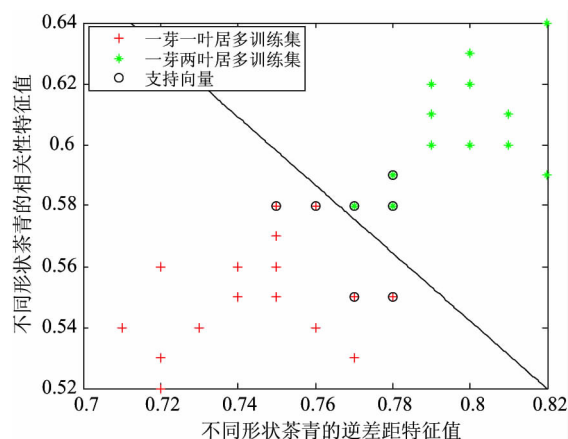


图 9 茶青(一芽一叶居多与一芽两叶居多)的样本训练集

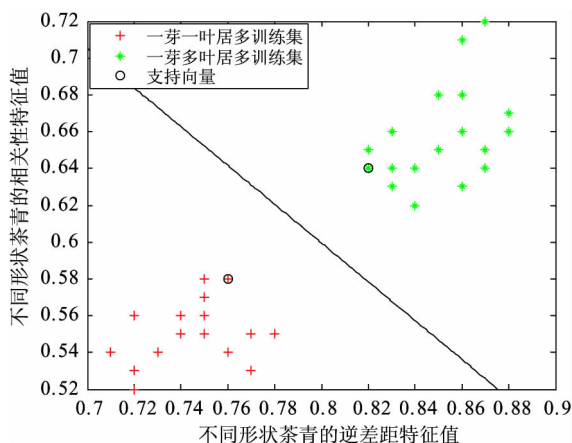


图 10 茶青(一芽一叶居多与一芽多叶居多)的样本训练集

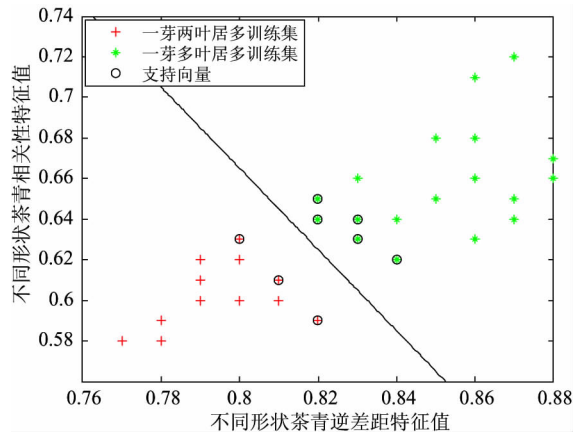


图 11 茶青(一芽两叶居多与一芽多叶居多)的样本训练集

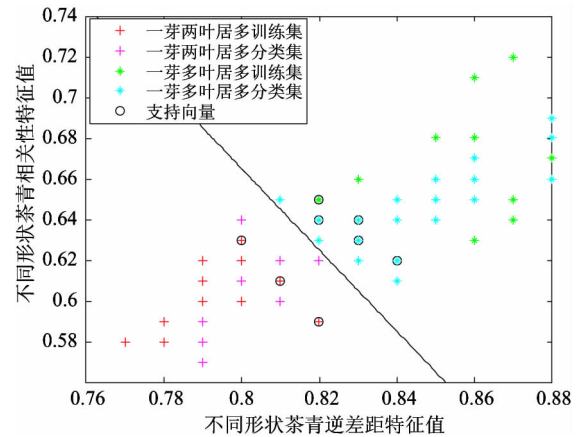


图 14 茶青(一芽两叶居多与一芽多叶居多)的分类结果

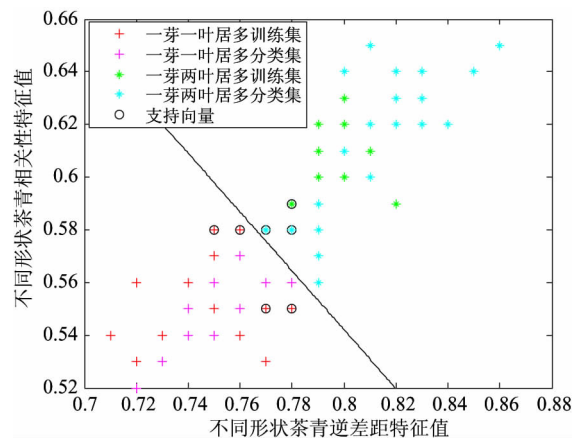


图 12 茶青(一芽一叶居多与一芽两叶居多)的分类结果

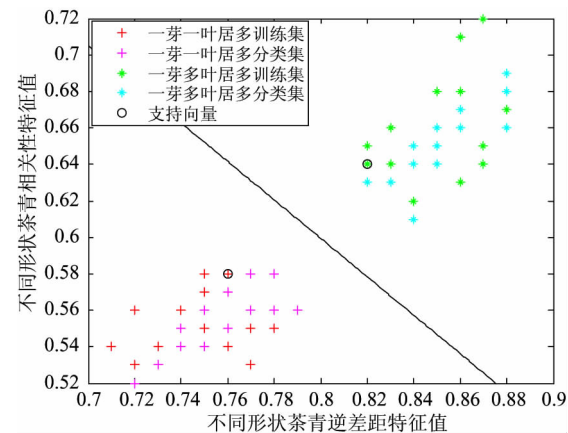


图 13 茶青(一芽一叶居多与一芽多叶居多)的分类结果

4 实验结果与分析

根据上文的分析思路与实验结果,使用 MATLAB 仿真实现了茶青纹理特征值提取及利用支持向量机进行茶青的训练和分类。本研究中,采集了不同形状的茶青各100多张,用来作为样本进行训

练,生成分类训练网络。为了检验分类的效果,采集了900多张图片用于分类,将图片分成三组,每组300张,其中不同形状的茶青各100张。茶青分类的准确率计算公式为

$$s = m/n \quad (10)$$

其中, s 为茶青准确率, m 为准确分类数目, n 为样本总数目。分类的准确率结果如表1所示。

表 1 茶青分类的准确率

组别	一芽一叶居多 (%)	一芽两叶居多 (%)	一芽多叶居多 (%)
A 组	88	86	98
B 组	86	89	94
C 组	90	92	96
平均	88	89	96

5 结 论

本文通过对茶青纹理特征的提取,用支持向量机的方法对茶青的类型进行了有效的分类,在茶叶行业内首次用计算机视觉的方法实现了对茶青的自动分类,为提高茶叶自动化加工的品质提供了重要保障。实验表明,在利用支持向量机进行茶青的训练与分类时,将纹理相关性特征值与逆差距特征值作为支持向量机的输入参数,能够很好地区分茶青的等级,分类的准确率达到90%。

参考文献

- [1] 李晓丽,何勇. 基于多光谱图像及组合特征分析的茶叶等级区分. 农业机械学报, 2009, 40(S1): 113-118
- [2] Chen Q, Zhao J, Fang C H, et al. Feasibility study on identification of green, black and Oolong teas using near-

- infrared reflectance spectroscopy based on support vector machine (SVM). *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2007, 66(3): 568-574
- [3] Wu D, Yang H, Chen X, et al. Application of image texture for the sorting of tea categories using multi-spectral imaging technique and support vector machine. *Journal of food engineering*, 2008, 88(4): 474-483
- [4] 陈怡群, 常春, 肖宏儒等. 人工神经网络技术在鲜茶叶分选中的应用. 农业网络信息, 2010, (007): 37-40
- [5] Gill G S, Kumar A, Agarwal R. Nondestructive grading of black tea based on physical parameters by texture analysis. *Biosystems Engineering*, 2013, 116(2): 198-204
- [6] Du C J, Sun D W. Recent developments in the applications of image processing techniques for food quality evaluation. *Trends in Food Science & Technology*, 2004, 15(5): 230-249
- [7] Zhang X, Yan W, Zhao X, et al. Nonlinear biological batch process monitoring and fault identification based on kernel fisher discriminant analysis. *Process biochemistry*, 2007, 42(8): 1200-1210
- [8] Hwang W J, Wen K W. Fast KNN classification algorithm based on partial distance search. *Electronics letters*, 1998, 34(21): 2062-2063
- [9] de O Bastos L, Liatsis P, Conci A. Automatic texture segmentation based on k-means clustering and efficient calculation of co-occurrence features. In: Proceedings of the 15th IEEE International Conference on Systems, Signals and Image Processing, 2008. 141-144
- [10] Wei C T, Wang N. Remote sensing image classification based on texture features. *Guilin Ligong Daxue Xuebao*, 2013, 33(1): 80-85

Research on online classification of fresh tea leaves based on texture analysis

Tang Zhe^{*}, Jiang Caihua^{*}, Zhang Li^{**}, Zhou Jianyong^{**}

(^{*} School of Information Science and Engineering, Central South University, Changsha 410083)

(^{**} Changsha Xiangfeng Tea Machinery Manufacturing Co., Ltd, Changsha 410100)

Abstract

A classification method based on texture analysis and support vector machine (SVM) was applied to automatic identification of fresh tea leaves to realize the fast, accurate online classification of fresh tea leaves. The texture characteristic parameters of tea leaves were obtained through the digital image processing technology, and different models for identification of fresh tea leaves were derived through SVM to realize the fast online classification of fresh tea leaves. The texture parameters including energy, correlation, contrast and homogeneity were obtained through the gray level co-occurrence matrix (GLCM), and they were used for SVM training and classification. The experimental results showed that the SVM application for tea classification achieved very good results, and the accuracy rate for classification can reach as high as 90%.

Key words: classification of fresh tea leaves, gray-level co-occurrence matrix (GLCM), texture analysis, image classification, support vector machine (SVM)