

文章编号: 0258-7025(2008)Supplement2-0345-06

基于可见光图像分析和支持向量机的鸭蛋新鲜度识别

刘 鹏^{1,2} 屠 康^{1,2} 潘磊庆¹ 刘 明¹

(¹ 南京农业大学食品科技学院, 江苏 南京 210095; ² 农业部农畜产品加工与质量控制重点开放实验室, 江苏 南京 210095)

摘要 研究鸭蛋内部的可见光图像信息(H 和 I)的变化来表征鸭蛋新鲜度的变化。采用可见光图像分析技术获取鸭蛋蛋心颜色参数,将哈夫单位作为新鲜度的验证指标,建立了基于支持向量机(SVM)的蛋心内部色泽参数的鸭蛋新鲜度预测模型。模型表明:当 SVM 类型为 epsilon-SVR,核函数为 RBF 模型特征参数 $C=2^7$ 、 $\sigma=2^3$ 时所建立的模型预测效果最好。其中模型预测效果参数 RMSEC=0.9520, EMSEP=0.4205。鸭蛋新鲜度预测值与测定值之间具有良好的线性关系,且预测值对实际值具有较大的搜索覆盖能力。该模型较好的解决了用普通的线性方法识别鸭蛋新鲜度存在的稳定性较差和置信度较小的问题。通过比较 SVM 和神经网络的识别结果可以发现: SVM 对鸭蛋蛋心颜色的识别性能优于神经网络的识别性能(SVM: 98.92% > ANN: 93.77%)。

关键词 图像处理; 识别模型; 支持向量机; 鸭蛋; 色泽参数; 新鲜度

中图分类号 O433.5 文献标识码 A doi: 10.3788/CJL200835s2.0345

Detection on Duck Egg's Freshness Based on Visible Light Image Analyzing and SVM

Liu Peng^{1,2} Tu Kang^{1,2} Pan Leiqing¹ Liu Ming¹

(¹ College of Food Science and Technology, Nanjing Agricultural University, Nanjing, Jiangsu 210095, China)
(² Key laboratory of Food Processing & Quality Control, Ministry of Agriculture, Nanjing, Jiangsu 210095, China)

Abstract This article applied the duck egg internal visible light image information (H and I) to characterize the freshness of duck egg. The visible light image analysis technology was applied to obtain the duck egg heart color parameters. The Hafu unit was used as the confirmation index of egg freshness through experiment, then established the forecast model based on SVM (support vector machines) which was developed using the egg heart color parameters to predicate the egg freshness. The model indicated that when we choose the SVM type as epsilon-SVR, the nuclear function as RBF, the model characteristic parameter $C=2^7$, $\sigma=2^3$, then the model forecast effect is the best. The model forecast effect parameters RMSEC reached 0.9520, and the EMSEP reached 0.4205. The predicted value of egg freshness and the practical value has good linear correlation and predicted value has a large covering and searching ability on the practical value. This model showed better stability and confidence interval compared with the ordinary linear method. Through the comparison of SVM and neural network recognition results we can get the conclusions that SVM has a better duck-egg freshness recognition performance than ANN by egg heart color's analysis (SVM reach to 98.92% > ANN reach to 93.77%).

Key words image processing; identification model; SVM; duck egg; color parameter; freshness

1 引言

鸭蛋保鲜及新鲜度分级,是蛋品加工及流通中的重要工序。目前,鸭蛋新鲜度检测用人工照蛋法及人工敲击辨听检测。由于人工检测生产效率低,

劳动强度大,对检测工技术要求高,检测稳定性较差,分级质量不能从根本上得到保证,这些严重制约着中国蛋品深加工技术的规模化发展。而随着人们生活水平的不断提高、市场竞争的日益激烈和科学

基金项目: 国家 863 计划(2007AA10Z213)、教育部“新世纪优秀人才支持计划”(NCET-05-0491)、江苏省自然科学基金(BK2006707-2)、江苏省攻关项目(BE 2007320)和校青年科技创新基金(Y200827)资助课题。

作者简介: 刘 鹏(1985—),男,博士研究生,主要从事农产品无损检测方面的研究。E-mail: llxx_2000@126.com

导师简介: 屠 康(1968—),男,博士,教授,主要从事农产品贮藏及物性方面的研究。E-mail: kangtu@njau.edu.cn

技术的飞速发展,用相应的技术手段代替人工及用无损检测代替破坏性检测来对农产品品质进行检测和分级是一种必然趋势。

根据蛋的光学特性可知,鲜鸭蛋对光的透射率会随着贮存时间的延长而下降。光的透射率变化,表明蛋的新鲜度等内部品质发生了变化,这种变化可以通过鸭蛋内部颜色信息的变化反映出来。将其应用于鸭蛋品质检测及分级,可大大提高检测效率和准确率。

支持向量机(SVM)是最近由统计学理论发展起来的一种小样本建模方法,在有限样本的情况下,支持向量机计算复杂性小,可获全局最优解和达到目前最好的泛化性能^[1]。该方法对于基于鸭蛋蛋心颜色的新鲜度识别这种非线性识别问题有较好预测性能。

因此,本研究将鸭蛋的哈夫单位值的变化和鸭

蛋颜色信息之间的变化联系起来,旨在探讨以建立基于 SVM 模型的方式对鸭蛋新鲜度进行无损识别,为构建自动检测系统提供参考。

2 实验方法与步骤

材料选购南京江宁青龙山农场的新鲜鸭蛋(三天以内,青白壳蛋各 130 枚,鸭蛋主要有白壳和青壳两种)260 枚,储藏于恒温箱中,其中温度 20 ℃,相对湿度 60%。

2.1 实验设计

实验设计方案如表 1 所示:其中后期数据处理和分析中无损对照组用以对测试组的数据进行校正,而测试组中选择哈夫单位符合要求的作为新鲜蛋样本,不符的作为非新鲜蛋样本。分别按 65:35 的比例构造训练和测试组。

表 1 实验设计概要

Table 1 Experiment design summary

Group	Testing group	Non-destructive detection group
Number of sample	240	20
Testing indexes	Visible light image of egg's content (visible light image acquisition device) Hafu unit (Hafu unit determinator) weight (electronic balance)	Visible light image of egg's content (visible light image acquisition device) Weight (electronic balance)
Testing times	10	10
Testing frequency /day	5	5

2.2 实验方法与步骤

实验方法与步骤如图 1 所示,具体分述如下。

鸭蛋新鲜度的有损标定即哈夫单位标定我们采用国标流行的美国国家禽蛋分级标准中的哈夫单位分级法,将蛋新鲜与否的哈夫单位阈值定为 60(符合 A 级标准)。首先进行材料预处理:所有试验用鸭蛋均去除表面的污斑,防止其在之后的信号采集和处理中产生干扰,同时这也是模拟工业生产中蛋

的预处理过程。去除裂纹蛋、变质蛋,将鸭蛋分为 10 批,每批 24 枚(其中青壳和白壳数量相同),先记录其无损指标,再记录其有损指标。实验步骤如下:

将预处理并分组过后的鸭蛋储存于温度 20 ℃,相对湿度 60%的恒温箱中备用。每次从样本组中随机挑选鸭蛋 24 枚,其中白壳、青壳各 12 枚,进行蛋壳表面颜色色差计测定,然后进入计算机视觉系统测定,获取图像处理参数,并记录。同时对其进行称重、

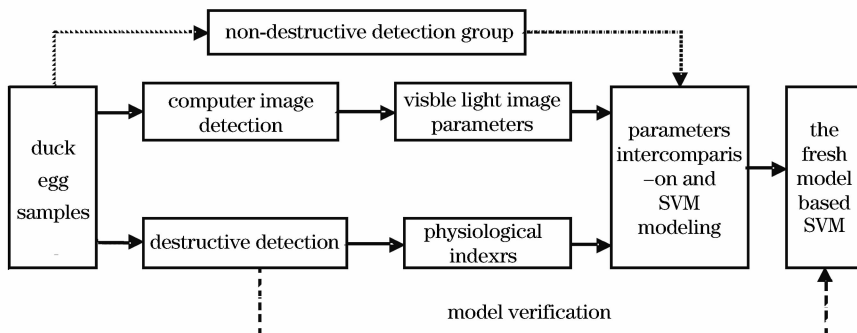


图 1 实验方法与步骤流程图

Fig. 1 Flow of experiment method and step

测气室高度及破壳测量哈夫单位值,记录相关生理指标。每五天定时试验,重复以上步骤,共 10 次试验。

2.3 可见光图像采集测定装置

该系统主要包括摄像头,图像采集卡,图像处理系统(自编),高亮节能光源,UPS 电源等。将鸭蛋放入预先经过背景消除处理的照蛋箱,并确保被测蛋与透光孔间无缝隙以避免光直接从缝隙中进入光室,打开透射光源后开启摄像头,然后将光室中鸭蛋透射光所成的像采集到计算机中;通过自编图象分析程序对图像进行处理并得出蛋心区域的色信息。装置如图 2 所示。

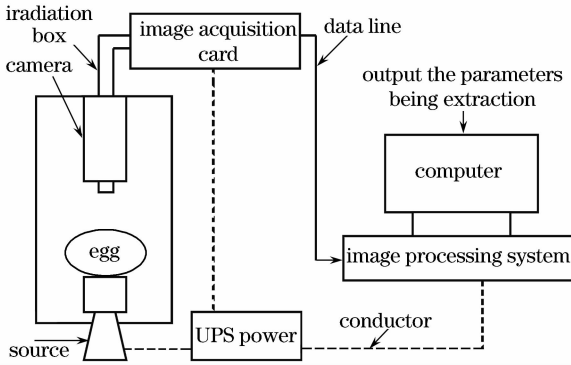


图 2 实验装置图

Fig. 2 Experiment equipment diagram

3 可见光图像预处理及特征参数提取

3.1 图像分割

拍摄到的鸭蛋蛋心剖面图像为彩色图像,图像经过分割,去除背景,保留完整的蛋心区域,以便进行鸭蛋蛋心的特征提取。为分析鸭蛋蛋心区域与背景区域的差别,运用自行开发的禽蛋图像分析系统提取计算图像的颜色分量直方图。图像的 R 分量直方图如图 3 右半部分所示,从图中可以看出 R 分量直方图分为明显的 2 块区域,分别对应蛋心颜色区域和背景区域。因此,在 R 分量上设置阈值区间

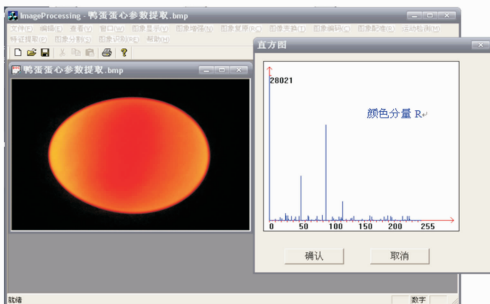


图 3 鸭蛋蛋心剖面图和 R 分量直方图

Fig. 3 Image of duck egg heart and the histogram of R component

R component

(75~100)即可以获得很好的分割效果。

3.2 颜色空间选取

在图像处理中,经常使用的颜色空间是 RGB 颜色空间和 HIS 颜色空间。RGB 颜色系统是在三基色原理下建立起来的颜色系统,是面向计算机、摄像机硬件的,RGB 成分与人对颜色的感觉并无直接的联系。人眼需要通过颜色的亮度、色度及饱和度来辨别颜色,而 HIS 颜色系统是根据人眼的这一特性来描述颜色的。HIS 模型的 3 个要素是色调 H 、饱和度 S 和亮度 I 。因此,本研究选取 HIS 颜色空间模型来提取鸭蛋蛋心的颜色特征值。由于通过图像分析软件提取透射光图像中的蛋心处颜色参数的 RGB 分布均值,因此需要对其按照(1)式进行非线性变换。

$$\begin{cases} I = (R + G + B) / 3, \\ S = 1 - \min(R, G, B) / I, \\ W = \arccos \left[\frac{2G - G - B}{2(R^2 + G^2 + B^2 - RG - GB - BR)^{1/2}} \right], \\ H = W, \quad (B \leq G) \\ H = 2\pi - W, \quad (B > G) \end{cases} \quad (1)$$

3.3 特征参数提取

由于蛋心处颜色参数 H 或 I ,用以度量鸭蛋的新鲜程度,该方法已由王巧华等验证^[2],其中青壳蛋新鲜度与 H 值相关性最高,因此我们计算青壳蛋的蛋心 H 值均值 \bar{H} 和均方差 σ_H 。而白壳蛋新鲜度与 I 值相关性最高,因此我们计算白壳鸭蛋的蛋心 I 值均值 \bar{I} 和均方差 σ_I 。其中均值表示蛋心处的整体颜色,而均方差表示的是鸭蛋蛋心区域各像素颜色的离散程度。其中各指标计算方法如(2)式~(5)式所示。

$$\bar{H} = \frac{1}{N} \sum_{i=1}^N H_i, \quad (2)$$

$$\sigma_H = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (H_i - \bar{H})^2}, \quad (3)$$

$$\bar{I} = \frac{1}{N} \sum_{i=1}^N I_i, \quad (4)$$

$$\sigma_I = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (I_i - \bar{I})^2}, \quad (5)$$

式中 N 为鸭蛋蛋心像素总数, H_i 为鸭蛋蛋心处第 i 个像素的色调值, I_i 为鸭蛋蛋心处第 i 个像素的亮度值。

自编软件实现了 RGB 系统向 HIS 系统的转换功能。操作界面如下图 4 所示:其中左边为原始图

像,右侧为用 R 的阈值区间进行背景纯化分割后的图像,用该图像进行色彩参数的提取,再利用

Spss16.0 计算颜色参数的平均值和均方差。

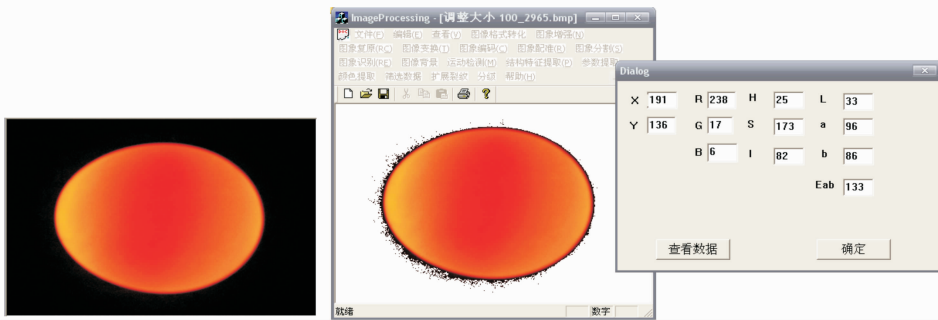


图 4 图像预处理及颜色特征参数提取

Fig. 4 Image pretreatment and extraction of color characteristics parameter

4 鸭蛋新鲜度的识别

青壳蛋的色调均值 \bar{H} 和色调均方差 σ_H , 白壳蛋的亮度均值 \bar{I} 和亮度均方差 σ_I 的分布如图 5 所示,可以看出,以 $\bar{H}, \sigma_H, \bar{I}, \sigma_I$ 作为特征值采用普通的线性方法来识别蛋的新鲜度是比较困难的(无法用线性方法进行划分。因此,必须用非线性模式识别方法来判断蛋心颜色是否为正常,从而判断出蛋是否新鲜。支持向量机能很好的解决非线性识别的问题,所构造的模型具有很好的预测性能^[3~8]。因此,本研究采用支持向量机来识别蛋心颜色。

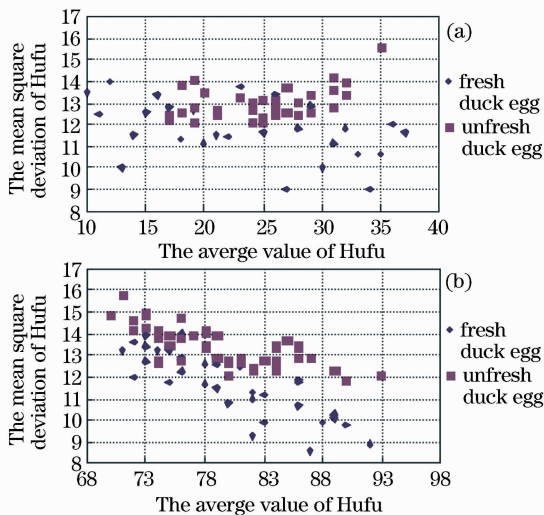


图 5 新鲜蛋和非新鲜蛋的颜色参数分布。(a)青壳鸭蛋, (b)白壳鸭蛋

Fig. 5 Color parameters distribution of the fresh egg and un-fresh egg. (a) Cyan shell duck egg, (b) white shell duck egg

4.1 模型筛选

建立基于蛋心颜色的鸭蛋新鲜度 SVM 识别模型是本文研究的重点。应用 SVM 建立识别模型首

先是核函数的选择,不同的核函数对所建立模型的性能影响很大。在没有先验知识指导的情况下,采用径向基核函数往往能够得到令人满意的结果^[9]。这是因为线性内核只有在样本数据线性可分的情况下才能得到较好的分类效果,而径向基核函数可以将非线性样本数据映射到高维特征空间,因此可以处理具有非线性关系的样本数据,而且当径向基核函数取特定参数值时与线性内核相同。

而 Sigmoid 内核在实际中很少应用,因此本研究中用径向基函数(RBF)作为 SVM 的核函数建立识别模型。

为了验证以上推断。我们采用台湾大学林智仁副教授等开发设计的 SVM 模式识别与回归的软件包 LIBSVM,对蛋的新鲜度用颜色参数进行 SVM 回归建模和预测。分别选取 epsilon-SVR、nu-SVR 两种 SVM 类型,并采用 4 种不同核函数:Linear 核函数、Poly 核函数、RBF 核函数与 Sigmoid 核函数进行 SVM 回归建模,校正模型的各有关参数如表 2 所示。以所建 SVM 回归模型对预测集样品的鸭蛋新鲜度进行了预测,结果如表 2 所示。表中 RMCEC 表示校正模型对预测集样品预测值与测定值的测定系数,EMSEP 表示校正模型对预测集样品的预测标准差,它们表征模型的预测能力(或泛化能力)。

从表 2 可以看出,无论是 epsilon-SVR 还是 nu-SVR 类型 SVM,在 Linear、Poly、RBF 与 Sigmoid 核函数中,支持向量机核函数为 RBF 时,所建立的模型预测效果要好些。对于相同的核函数,epsilon-SVR 比 nu-SVR 型 SVM 的预测效果好。总之,当 SVM 类型为 epsilon-SVR,核函数为 RBF 所建立的模型预测效果最好其中 RMSEC=0.9520,EMSEP=0.4205。

因此通过筛选,选择径向基函数作为 SVM 的

核函数建立识别模型。由于径向基函数性能直接受到参数大小的影响。误差惩罚参数 C 实现在错分样本的比例和算法复杂度之间的折衷,即在确定的特征子空间中调节学习机器置信范围和经验风险的比例以使学习机器的推广能力最好。它的选取一般是由具体的问题而定,并取决于数据中噪声的数量,在确定的特征子空间中 C 的取值小表示对经验误差的惩罚小,学习机器的复杂度小而经验风险值较大;如果 C 取 ∞ ,则意味着训练样本必须要准确地分类。径向基核参数 σ 直接影响 SVM 的性能优劣。特征子空间的维数决定了能在此空间构造的线

性分类面的最大维数,也就决定了线性分类面能达到的最小经验误差。同时,每一个特征子空间对应唯一的推广能力最好的分类超平面,如果特征子空间维数很高,则得到的最优分类面就可能比较复杂,经验风险大但置信范围小;反之,如果特征子空间维数低,则得到的最优分类面简单,经验风险小,但置信范围大。这两种情况下得到的 SVM 都不会有好的推广能力,只有首先选择合适的 σ ;将数据投影到合适的特征空间,才可能得到推广能力良好的 SVM 分类器^[10,11]。

表 2 epsilon-SVR, nu-SVR 型支持向量机回归结果对比表

Table 2 Comparison of regression results between epsilon-SVR and nu-SVR

Type of SVM	epsilon-SVR			nu-SVR		
	No. of support vector	RMSEC	EMSEP	No. of support vector	RMSEC	EMSEP
Linear	76	0.9409	0.4813	110	0.9239	0.5109
Poly	77	0.9287	0.5247	110	0.9135	0.5231
Sigmoid	79	0.9416	0.4987	110	0.9231	0.5104
RBF	91	0.9520	0.4205	110	0.9348	0.4981

4.2 模型构建及验证

在明确选择径向基函数作为 SVM 的核函数建立识别模型的基础上,接下来我们通过试验的方法确定支持向量机模型中的参数 C 和 σ 。

目前,对参数的优化尚无比较成熟的方法,一般需要通过多次实验确定。通过对 240 个鸭蛋进行颜色特征值的提取,建立支持向量机。利用 H 均值、 σ_H 及 I 均值、 σ_I 来识别青壳蛋和白壳蛋的新鲜度,在 240 个样本中取 24 个新鲜蛋、132 个非新鲜蛋(符合实验设计中的 65:35)比例作为训练集,其余样本作为预测集检验模型的效果。依据训练样本回判准确率、预测样本预测准确率确定了参数 C 和 σ ,试验数据见表 3。

支持向量机模型对训练集样本的分类准确率不是最高,但是模型对预测集的分类准确率却是最高,此时模型的推广能力最强。为了比较 SVM 的识别能力,本研究用常规的人工神经网络进行鸭蛋新鲜度的识别。采用 MATLAB2007b 中的 NN 工具箱的 NFT 工具包,网络采用 4-8-1 结构(其中输入量 4 个;对青壳蛋为 H 均值、 σ_H 而白壳蛋为 I 均值和 σ_I ,隐层为单隐层结构,节点为 8 个,输出量 1 个;为鸭蛋新鲜度),学习速度取 0.2,动量因子取 0.2,进行训练。当系统最小平均学习误差为 0.001 时停止训练。训练集的回判准确率为 98.92%。将预测集输入神经网络,预测集的预测准确率为 93.77%。比较 SVM 和神经网络的识别结果可以看出,SVM 对鸭蛋蛋心颜色的识别性能优于神经网络的识别性能。

由表 3 可以看出,在参数 $C=2^7$ 、 $\sigma=2^3$ 时,虽然

表 3 预测结果总结

Table 3 Testing results summary

Evaluation parameters	Value be taken and results								
C	2^{-3}	2^{-1}	2^0	2^1	2^3	2^5	2^7	2^9	2^{11}
σ	2^{-7}	2^{-5}	2^{-3}	2^{-1}	2^0	2^1	2^3	2^5	2^7
Back evaluation accuracy /%	97.32	97.32	95.47	95.47	93.81	93.81	92.25	91.25	95.47
Forecast accuracy /%	78.34	89.23	92.51	92.51	92.31	93.67	94.31	92.36	93.12

5 结 论

本文应用支持向量机这一新颖的智能计算方法解决了鸭蛋蛋心颜色识别鸭蛋新鲜度这一复杂的非线性问题。主要得出以下结论:

1) 以蛋心处颜色参数 H 均值、均方差 σ_H 及 I 均值、均方差 σ_I 作为特征值,采用普通的线性方法来识别蛋的新鲜度存在稳定性较差,置信度较小的问题,该体系可以看作非线性问题,本研究采用支持

向量机较好的解决了该问题。

2) 通过经验分析和 SVM 模式识别与回归的软件包对蛋的新鲜度用颜色参数进行 SVM 回归建模和预测,我们发现当 SVM 类型为 epsilon-SVR,核函数为 RBF 所建立的模型预测效果较好。模型性能评估参数 RMSEC = 0.9520, EMSEP = 0.4205。说明蛋新鲜度预测值与测定值之间具有良好的线性关系且预测值对实际值具有较大的搜索覆盖能力。表明基于 SVM 和可见光图像分析技术预测鸭蛋新鲜度是可行的。

3) 在参数 $C=2^7$ 、 $\sigma=2^3$ 时,虽然支持向量机模型对训练集样本的分类准确率不是最高,但是模型对预测集的分类准确率却是最高,此时模型的推广能力最强。而比较 SVM 和神经网络的识别结果可以看出,SVM 对鸭蛋蛋心颜色的识别性能优于神经网络的识别性能(SVM: 98.92% > ANN: 93.77%)。

参 考 文 献

- Li X L. Multi-scale statistical process monitoring in machining [J]. *IEEE Transactions on Industrial Electronics*, 2005, **52**(3): 924~927
- Wen Youxian, Wang Qiaohua, Zong Wangyuan. Research on yolk color grading model[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2001, **17**(6): 139~141
文友先,王巧华,宗望远等. 鸭蛋蛋心颜色等级模型研究[J]. *农业工程学报*, 2001, **17**(6): 139~141
- Huang Y, Liang S Y. Modeling of cutting forces under hard turning conditions considering tool wear effect [J]. *J. Manufacturing Science and Engineering, Transactions of the ASME*, 2005, **127**(2): 262~270
- George Paschos. Fast color texture recognition using chromaticity moments[J]. *Pattern Recognition Letters*, 2000, **21**: 837~841
- Burges C J C. A tutorial on support vector machines for pattern recognition[J]. *Data Mining and Knowledge Discovery*, 1998, **2**(2): 121~169
- Gómez A H, He Y, Pereira A G. Non-destructive measurement of acidity, soluble solids and firmness of Satsuma mandarin using Vis/NIR-spectroscopy techniques [J]. *J. Food Engineering*, 2006, **77**(2): 313~319
- Wang H L, Sh H, Ch M. On-line tool breakage monitoring in turning[J]. *J. Materials Processing Technology*, 2003, **139**(1~3): 237~242
- Zhao Jiewen, Zhang Haidong, Liu Muhua. Preprocessing methods of near-infrared spectra for simplifying prediction model of sugar content of apples[J]. *Acta Optica Sinica*, 2006, **26**(1): 136~140
赵杰文,张海东,刘木华. 简化苹果糖度预测模型的近红外光谱预处理方法[J]. *光学学报*, 2006, **26**(1): 136~140
- Zheng Xiaoxia, Qian Feng. Support vector machine with Gauss kernel classified method and model selection [J]. *Computer Engineering and Applications*, 2006, **42**(1): 77~79
郑小霞,钱 锋. 高斯核支持向量机分类和模型参数选择研究[J]. *计算机工程与应用*, 2006, **42**(1): 77~79
- Fu Xiaping, Ying Yibin, Liu Yande *et al.*. Detection of pear firmness using near infrared diffuse reflectance spectroscopy[J]. *Spectroscopy and Spectral Analysis*, 2006, **26**(6): 1038~1041
傅霞萍,应义斌,刘燕德等. 水果坚实度的近红外光谱检测分析试验研究[J]. *光谱学与光谱分析*, 2006, **26**(6): 1038~1041
- Tian Youwen, Zhang Changshui, Li Chenghua. Study on plant disease recognition using support vector machine and chromaticity moments [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2004, **35**(3): 95~98
田有文,张长水,李成华. 支持向量机和色度矩的植物病害识别研究[J]. *农业机械学报*, 2004, **35**(3): 95~98
- Ji Haiyan, Yan Yanlu. Application of principal components-artificial neural networks in near-infrared spectroscopy quantitative analysis [J]. *J. Instrumental Analysis*, 1999, **18**(3): 12~14
吉海彦,严衍录. 主成分-人工神经网络在近红外光谱定量分析中的应用[J]. *分析测试学报*, 1999, **18**(3): 12~14
- Wu Di, Huang Lingxia, He Yong. Visible-near infrared reflection spectroscopy for crop-weed discrimination [J]. *Acta Optica Sinica*, 2008, **28**(8): 1619~1622
吴 迪,黄凌霞,何 勇. 作物和杂草叶片的可见-近红外反射光谱特性[J]. *光学学报*, 2008, **28**(8): 1619~1622
- Shi Youming, Liu Gang, Liu Jianhong. Identification of auricularia auricula from different regions by Fourier transform infrared spectroscopy[J]. *Acta Optica Sinica*, 2007, **27**(1): 129~132
时有明,刘 刚,刘剑虹等. 不同产地黑木耳的傅里叶变换红外光谱鉴别[J]. *光学学报*, 2007, **27**(1): 129~132