

基于机器视觉的苹果分级中特征参量选择方法*

殷勇 陶凯 于慧春

(河南科技大学食品与生物工程学院, 洛阳 471003)

【摘要】 为提高基于数字图像的苹果分级的准确性,常提取多特征信息。然而,使用多特征信息分级时会存在信息冗余等问题。为此,运用主成分分析(PCA)来融合特征参量,并借助 Wilks Λ 统计量选择对分级有显著作用的主成分;然后依据各特征参量对所选择主成分的贡献率筛选特征参量。Fisher 判别分析(FDA)结果表明:使用所选择的特征参量进行苹果分级,分级效果明显优于特征选择前,分级正确率和交叉验证正确率分别提高了 2.0% 和 1.5%。

关键词: 苹果 分级 图像处理 特征选择 主成分分析 统计量

中图分类号: TS255; TP391.41 **文献标识码:** A **文章编号:** 1000-1298(2012)06-0118-04

Feature Selection Method for Apple Grading Based on Machine Vision

Yin Yong Tao Kai Yu Huichun

(College of Food and Bioengineering, Henan University of Science and Technology, Luoyang 471003, China)

Abstract

In order to improve the accuracy of apple grading in digital image processing system, the multi-feature information was extracted for describing the apple features. However, this method may result in information redundancy and so on. So, the principal component analysis (PCA) was used to carry out information fusion of the feature parameters, and with the aid of Wilks Λ statistic the principal components (PC) which could promote grading results were selected. Then some features used in grading were selected based on the contribution rate to selected PC. The results of Fisher discriminate analysis (FDA) showed that the grading effect corresponding to the selected features was better than that of all features, and the grading accuracy and the cross-validation accuracy rose by 2.0% and 1.5%, respectively.

Key words Apple, Grade, Image, Feature selection, Principal component analysis, Statistic

引言

苹果分级是苹果市场销售前的一个重要环节。近几年,数字图像处理和模式识别等技术的进展推动了机器视觉在苹果无损分级技术领域的发展。目前,国内外专家学者对基于数字图像处理的苹果分级方法进行了大量研究,但大多侧重于大小、形状、颜色、表面纹理或缺陷等单一特征分级^[1~3]。由于使用单一特征信息进行分级存在片面性,难以满足苹果分级过程中对多指标进行描述的要求,分级结

果不够可靠、稳定,而使用多特征信息进行分级判别^[4~5],分级结果更准确、更可靠。但是在使用多特征信息进行苹果分级的过程中,如不加选择地利用所有特征参量进行分级,则会造成信息冗余、运算时间长、计算精度下降、系统不稳定等一系列问题^[6~8]。因此,选出对分级贡献大的特征参量是解决苹果分级问题的一个关键。本文将主成分分析与 Wilks Λ 统计量 2 种方法相结合,进行苹果分级特征参量的选择,并通过主成分分析和 Fisher 判别分析进行分级实验,以检验该方法的有效性。

收稿日期: 2011-09-07 修回日期: 2011-10-08

* 河南省科技创新杰出青年资助项目(624420017)

作者简介: 殷勇,教授,博士,主要从事农产品、食品品质无损检测技术研究, E-mail: yinyong@mail.haust.edu.cn

1 实验材料与方法

1.1 图像处理系统

实验使用的图像处理系统由实验室自行研制，硬件部分由 OK-AC1310 型 CCD 彩色摄像头、LED 光源、转动机构、图像采集室和计算机等部分组成（图 1）。

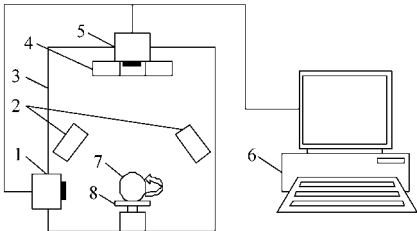


图 1 图像处理系统
Fig. 1 Image processing system

1. 侧部 CCD 摄像头 2. 侧部平板光源 3. 光源箱 4. 顶部环形光源 5. 顶部 CCD 摄像头 6. 计算机 7. 苹果 8. 旋转托盘

1.2 实验材料

实验样本为市售烟台富士苹果，共 200 个，按照鲜苹果国家标准（GB/T 10651—2008）^[9] 分级，其中优等品样本、一等品样本、二等品样本和等外品样本各 50 个。

1.3 实验方法

采集时将苹果置于水平转动托盘上，为了保证苹果的转动速度与摄像速度匹配，托盘每 24 s 旋转一圈，摄像头每 8 s 采集一次。苹果转动一周，侧部摄像头采集 3 次苹果侧面图像（图 2）。顶部摄像头采集苹果顶部图像之后，通过在 2 条传送带交界处设置一凹坑，苹果到凹坑处由触发器将其推出，进入相垂直的另一轨道，使苹果倒置，即可采集底部图像。这样，每个苹果共采集 5 幅图像，包含其全部的外部特征。实际上，可认为该实验方法模拟了生产线上苹果随传送带移动时的图像采集。

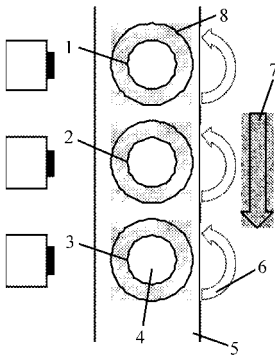


图 2 图像采集示意图

Fig. 2 Schematic diagram of image acquisition

1. 苹果 a 面 2. 苹果 b 面 3. 苹果 c 面 4. 苹果 5. 传送带 6. 托盘转动方向 7. 传送带运送方向 8. 转动托盘

1.4 图像预处理

为方便特征参量提取，需将 CCD 摄像头采集到的图像进行预处理。图像处理算法在 Matlab 2010b 平台上编译。

首先将图像从 RGB 颜色模型转换到 HSI 颜色模型中；然后，比较色调 H (hue)、饱和度 S (saturation)、强度 I (intensity) 分量的灰度直方图，发现在 S 分量图中苹果和背景最易区分，因此采用 S 分量提取苹果图像，并通过大津法 (Otsu) 去除背景；接着将无背景的图像进行二值化处理（图 3a），以提取苹果大小与形状类特征；将图 3a 与原图相乘得到分割后的图像（图 3b），以提取苹果的红色着色面积特征；最后将图 3b 转换成灰度图像（图 3c），以提取苹果的表面纹理和表面伤痕类特征。

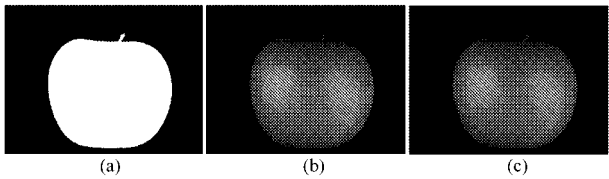


图 3 图像预处理

Fig. 3 Results of images preprocessing

(a) 二值图像 (b) 分割后图像 (c) 灰度图像

1.5 特征参量提取

根据国标，利用外部特征对苹果进行等级判定主要是依据苹果的大小与形状、红色着色面积、表面纹理特征^[10]和表面伤痕等 4 个方面的特征。采集自苹果水平放置时苹果侧面图像的特征包括：最小外接矩形水平倾角、面积、圆弧度、长宽比、矩形度、疤痕面积、红色面积，共 7 个；采集自苹果顶部与底部图像的特征包括：面积、圆弧度、长宽比、矩形度、疤痕面积、红色面积，共 6 个；采集自苹果全部图像的特征包括：平均亮度、平均对比度、平滑度、三阶矩、一致性、熵，共 6 个。

对苹果图像进行傅里叶变换，用极坐标函数 $S(r, \theta)$ 表示傅里叶频谱，其中 S 是频谱函数， r 和 θ 分别表示频率和方向。对于给定方向 θ ， $S(r, \theta)$ 可以看成一维函数 $S_\theta(r)$ ，同样对于给定频率 r ， $S_r(\theta)$ 也是一个一维函数，即

$$S(r) = \sum_{\theta=0}^{\pi} S_\theta(r) \quad S(\theta) = \sum_{r=1}^{R_0} S_r(\theta)$$

式中 R_0 ——以原点为圆心的圆半径

选取 $S(r)$ 均值、 $S(r)$ 方差、 $S(r)$ 最大值、 $S(\theta)$ 均值、 $S(\theta)$ 方差、 $S(\theta)$ 最大值等 6 个特征描述苹果图像的傅里叶频谱特性。

另外，一幅数字图像 $f(x, y)$ 的二维 $(p + q)$ 阶矩定义为

$$m_{pq} = \sum_x \sum_y x^p y^q f(x, y) \quad (p, q = 0, 1, \dots)$$

求和在跨越图像的所有空间坐标 x, y 的值上进行。相应的中心距为

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y)$$

其中

$$\bar{x} = \frac{m_{10}}{m_{00}} \quad \bar{y} = \frac{m_{01}}{m_{00}}$$

归一化 $(p + q)$ 阶中心距为

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma} \quad (p + q = 2, 3, \dots)$$

其中

$$\gamma = \frac{p + q}{2} + 1$$

对图像进行平移、旋转和比例缩放都不敏感的 7 个二维不变矩,分别为

$$\phi_1 = \eta_{20} + \eta_{02}$$

$$\phi_2 = (\eta_{20} + \eta_{02})^2 + 4\eta_{11}^2$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$

$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$$

$$\phi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})$$

$$\phi_7 = (3\eta_{21} + \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$$

选取 $\phi_1, \phi_2, \phi_3, \phi_4, \phi_5, \phi_6, \phi_7$ 共 7 个特征描述苹果的不变性特征。全部特征参量共有 32 个。

2 特征参量选择方法

设 A 表示 200 个样品各提取 32 个特征所形成的特征参量矩阵,对 A 进行主成分计算可得到 32 个主成分。由于每个主成分都是原 32 个特征参量的线性组合,因此每个主成分都融合了原 32 个特征参量信息。但是,并不是每个主成分对苹果分级都有效。因 Wilks Λ 统计量可以用来检验多个变量的判别效果和各个变量的判别能力^[11~12],所以可利用 Wilks Λ 统计量来检验各个主成分的判别能力,从而选择出对分级有显著性作用的主成分。Wilks Λ 统计量计算式为

$$\Lambda = \frac{|W|}{|T|} \tag{1}$$

式中 W ——组内离差矩阵
 T ——总离差矩阵

当选取 m 个指标进行研究时,设 $W_m = (w_{ij})_{m \times m}$, 记 $w_{ij}^0 = w_{ij}$, 则 $\Lambda_m = \frac{|W_m|}{|T_m|} = \frac{w_{11}}{t_{11}} \frac{w_{22}^{(1)}}{t_{22}^{(1)}} \cdots \frac{w_{mm}^{(m-1)}}{t_{mm}^{(m-1)}}$, 相应的 $|T_m| = t_{11} t_{22}^{(1)} \cdots t_{mm}^{(m-1)}$, 则

$$\Lambda_m = \frac{|W_m|}{|T_m|} = \frac{w_{11}}{t_{11}} \frac{w_{22}^{(1)}}{t_{22}^{(1)}} \cdots \frac{w_{mm}^{(m-1)}}{t_{mm}^{(m-1)}} \tag{2}$$

Λ_m 可以用来度量 m 个指标 $X_1, X_2, \dots, X_m$ 对 k 个总体的判别效果, Λ_m 越小,判别效果越好。

在选择过程中,每个主成分相当于一个变量。设在第 1 步引进变量 X_1 , 则其值是 $\Lambda_1 = \frac{w_{11}}{t_{11}}$, 为方便描述,恰有 $\Lambda_1 = \min \Lambda_i (1 \leq i \leq m)$ 。第 2 步引进变量 X_r , 则其值 $\Lambda_{1r} = \frac{w_{11} w_{rr}^{(1)}}{t_{11} t_{rr}^{(1)}}$, 记 $\Lambda_{r+1} = \frac{w_{rr}^{(1)}}{t_{rr}^{(1)}}$, 则有 $\Lambda_{1r} = \Lambda_1 \Lambda_{r+1}$, 其中 Λ_{r+1} 表示在第 2 步引入 X_r 后引起的 Λ 的变化。可以看出, Λ_1 可以显示变量 X_1 的重要性, Λ_{1r} 可以显示变量 X_1, X_r 的重要性, Λ_{r+1} 表示在变量 X_1 给定时,再引进变量 X_r 的判别能力。一般地, Λ_{r+l} 表示给定 l 个变量时再引进新变量 X_r 引起的 Λ 的变化量,它可以检验引进 X_r 所增加的判别能力。这样依次在第 1 至第 l 步中引进了 l 个重要变量 $X_1, X_2, \dots, X_l$ 。则第 $l + 1$ 步时,在未选进的变量中选出一个变量,使得 $\Lambda_{1,2,\dots,l,r} = \Lambda_{1,2,\dots,l} \Lambda_{r+1,2,\dots,l} = \Lambda_{1,2,\dots,l} \frac{w_{rr}^{(l)}}{t_{rr}^{(l)}}$ ($r = l + 1, l + 2, \dots, p$) 达到最小。现仍假设恰有 X_{r+1} 使 $\Lambda_{1,2,\dots,l,l+1} = \min_{l+1 \leq r \leq p} \Lambda_{1,2,\dots,l,r}$ 。为保证每步选入的变量是重要的,对该步所选入变量的判别能力做显著性检验,计算 F 统计量的值 $F_0 = \frac{1 - \Lambda_{r+1,2,\dots,l}}{\Lambda_{r+1,2,\dots,l}} \frac{n - k - l}{k - l}$, 给定显著性水平

$\alpha = 0.05$, 若 $p = P(F \geq F_0) < \alpha$, 则认为 X_{r+1} 显著, 可以引进。如不显著,则停止引进,并考虑较早选中的变量的重要性是否发生较大变化,剔除的原则等同于引入的原则。按照上述方法,使用 SPSS 进行分析,按照先后顺序,入选的主成分分别为: $PC_2, PC_1, PC_3, PC_4, PC_{11}, PC_5, PC_6, PC_{20}, PC_{16}, PC_{10}, PC_8, PC_{17}, PC_{24}, PC_{25}$, 其中 PC_i 表示第 i 个主成分。

在生成各主成分的过程中,各原始特征变量对应的线性组合系数表示了该变量在各主成分生成过程中贡献的大小,选取贡献最大的特征参量作为代表该主成分的特征参量。表 1 是经上述方法入选的特征参量。其中 22 和 26 号特征参量入选了 2 次,所以实际选入了 1、2、5、6、7、18、22、23、25、26、27、32 号共 12 个特征参量。

表 1 入选的特征参量

Tab.1 Selected component vectors

	PC ₂	PC ₁	PC ₃	PC ₄	PC ₁₁	PC ₅	PC ₆	PC ₂₀	PC ₁₆	PC ₁₀	PC ₈	PC ₁₇	PC ₂₄	PC ₂₅
向量系数极值	-0.40	0.29	0.32	0.39	-0.50	0.56	-0.59	-0.42	-0.50	0.53	0.48	-0.56	0.54	-0.66
特征参量	7	23	18	22	26	1	6	22	32	5	26	27	25	2

3 实验结果与分析

用特征选择前、后的特征参量进行 Fisher 判别分析(FDA)。在 CPU 为 AMD7750、内存 2 GB 的计算机上,特征选择前计算耗时 0.209 244 s,特征选择后计算耗时 0.120 008 s,耗时缩短了 42.6%。表明使用选择特征参量进行判别分析,运算效率得到了提高。图 4、5 分别是原 32 个特征参量和 12 个选择特征参量的 FDA 分类结果。从图 4 可以看出,等外品被完全区分开,二等品和一等品之间存在个别混杂,一等品和优等品混杂严重。苹果的等级混杂主要是发生在优等品和一等品之间。主要是由于这 2 种等级的苹果在外部形态上比较接近,苹果的表面伤痕、大小和形状较为接近,只能依靠表面红色面积以及表面纹理类特征进行分级。经 SPSS 计算,样

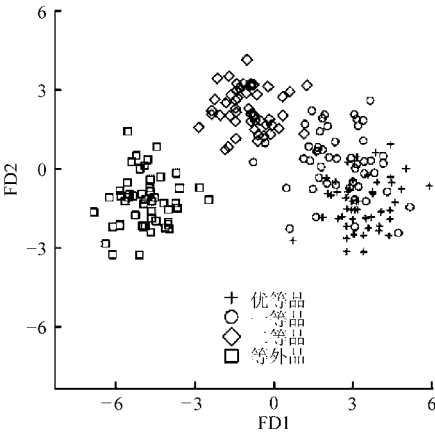


图 4 特征选择前 4 个等级苹果的 FDA 分级结果

Fig.4 FDA results of the first four grades apples before feature selection

本分级正确率为 94.5%,交叉验证正确率为 90.5%。在图 5 中,等外品被完全区分出来,二等品和一等品之间存在个别混杂,一等品和优等品间的混杂情况大大改善,经 SPSS 计算,样本分级正确率为 96.5%,交叉验证正确率为 92%,比特征选择前分别提高了 2.0% 和 1.5%。

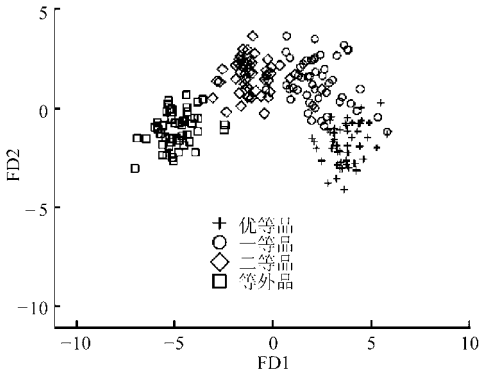


图 5 特征选择后 4 个等级苹果的 FDA 分级结果

Fig.5 FDA results of the last four grades apples after feature selection

特征选择前、后的 FDA 结果表明,该特征选择方法能够减少信息冗余对分级结果造成的不利影响,减少了运算时间,提高了分级正确率。

4 结束语

在主成分分析和 Wilks Λ 统计量计算相结合的基础上,提出了一种多特征选择方法。分级实验结果表明,这种选择方法在苹果分级中可选择出有效的特征参量,提高了分级正确率,减少了运算时间。虽然该方法是在苹果分级研究中提出的,但不失一般性,对其他农产品的分级也有参考价值。

参 考 文 献

1 冯斌,汪懋华. 基于计算机视觉的水果大小检测方法[J]. 农业机械学报,2003,34(1):73~75.
Feng Bin, Wang Maohua. Detecting method of fruit size based on computer vision[J]. Transactions of the Chinese Society for Agricultural Machinery,2003,34(1):73~75. (in Chinese)

2 Whitelock D P, Brusewitz G H, Stone M L. Apple shape and rolling orientation[J]. Applied Engineering in Agriculture, 2006, 22(1):87~94.

3 Li Qingzhong, Wang Maohua, Gu Weikang. Computer vision based system for apple surface defect detection[J]. Computers and Electronics in Agriculture, 2002, 36(12): 215~223.

4 龙满生,何东健,宁纪峰. 基于遗传神经网络的苹果综合分级系统[J]. 西北农林科技大学报: 自然科学版,2001, 29(6):108~111.
Long Mansheng, He Dongjian, Ning Jifeng. An integrated apple grading system based on genetic neural network[J]. Journal of Northwest A&F University: Natural Science Edition,2001,29(6):108~111. (in Chinese)

- 2 李里特,关东胜,张泽俊,等.一种制备强酸化水的装置:中国,96244992.X[P].1998-03-25.
- 3 叶清源.电解水装置的一种电解槽:中国,200420028723.4[P].2005-07-20.
- 4 耿庆仁.无离子隔膜电解技术制备酸性氧化电位水工艺:中国,200410021043.4[P].2004-12-29.
- 5 沈晓盛,刘长军,蔡友琼,等.电解海水的抑菌活性及对食品加工表面材料的消毒效果[J].微生物学通报,2008,35(11):1833~1839.
Shen Xiaosheng, Liu Changjun, Cai Youqiong, et al. Antibacterial effects of electrolyzed oxidizing seawater against food-borne pathogens in bacteria suspensions and on food processing surfaces[J]. Microbiology, 2008, 35(11):1833~1839. (in Chinese)
- 6 Cao W, Zhu Z W, Shi Z X, et al. Efficiency of slightly acidic electrolyzed water for inactivation of *Salmonella enteritidis* and its contaminated shell eggs[J]. International Journal of Food Microbiology, 2009, 130(2):88~93.
- 7 铃木铁也.电解水の食品分野での利用[J].食品と开发,1998,33(3):10~12.
- 8 伊藤喜久.电解水・オゾン水装置の開発动向[J].食品と开发,1997,32(3):20.
- 9 Abadias M, Usall J, Oliverira M, et al. Efficacy of neutral electrolyzed water (NEW) for reducing microbial contamination on minimally-processed vegetables[J]. International Journal of Food Microbiology, 2008, 123(1~2):151~158.
- 10 郝建雄,王愈,李里特.酸性电解水处理对草莓采后生理的影响[J].温室园艺,2006(5):36~38.
- 11 Ayebah B, Hung Y C. Electrolyzed water and its corrosiveness on various surface materials commonly found in food processing facilities[J]. Journal of Food Process Engineering, 2005, 28(3):247~264.
- 12 刘海杰,朱叶,乔支红,等.弱电解水对豆腐品质的影响[J].中国农业大学学报,2007,12(6):67~70.
Liu Haijie, Zhu Ye, Qiao Zhihong, et al. Effect of weakly electrolyzed water on quality of tofu[J]. Journal of China Agricultural University, 2007, 12(6):67~70. (in Chinese)
- 13 Hisae K, Mamoru Y. Disinfection of seawater from fishing ports by an electrolytic apparatus and its application to fisheries sanitation[J]. Nippon Suisan Gakkaishi, 2003, 69(6):955~959.
- 14 Huang Y R, Hug Y C, Hang D F. Application of electrolyzed water in the food industry[J]. Food Control, 2008, 19(4):329~345.
- 15 肖卫华,李里特,王慧敏,等.电生功能水防治黄瓜白粉病试验初报[J].植物保护,2003,29(4):50~51.
- 16 郝建雄,李里特,马占鸿,等.电生功能水防治小麦条锈病试验研究[J].中国植保导刊,2006,26(6):21~23.
Hao Jianxiong, Li Lite, Ma Zhanhong, et al. Study on electrolyzed functional water against wheat stripe rust[J]. China Plant Protection, 2006, 26(6):21~23. (in Chinese)
- 17 李里特.电生功能水在农业上的应用研究及展望[J].中国农业信息,2006(1):15~16.
- 18 李法德,李里特,刘海杰,等.浸没式电解混合装置:中国,201010556255.8[P].2011-04-13.
- 19 张铁垣.化验员手册[M].2版.北京:中国电力出版社,1996:133~350.

(上接第121页)

- 5 李先锋,朱伟兴,花小朋,等.融合形状和颜色特征的苹果等级检测[J].计算机工程与应用,2010,46(35):202~204.
Li Xianfeng, Zhu Weixing, Hua Xiaopeng, et al. Apple grading detection based on fusion of shape and color features[J]. Computer Engineering and Applications, 2010, 46(35):202~204. (in Chinese)
- 6 Leemans V, Magein V, Destain M F. Defects segmentation on 'Golden Delicious' apples by using colour machine vision[J]. Computers and Electronics in Agriculture, 1998, 20(1):117~130.
- 7 Zhu Bin, Jiang Lu, Luo Yaguang, et al. Gabor feature-based apple quality inspection using kernel principal component analysis[J]. Journal of Food Engineering, 2007, 81(4):741~749.
- 8 Zou Xiaobo, Zhao Jiwen, Li Yanxiao, et al. In-line detection of apple defects using three color cameras system[J]. Computers and Electronics in Agriculture, 2010, 70(1):129~134.
- 9 GB/T 10651—2008 鲜苹果[S]. 2008.
- 10 Rafael C Gonzalez, Richard E Woods. 数字图像处理[M].2版.阮秋琦,阮宇智,译.北京:电子工业出版社,2007.
- 11 高惠璇.应用多元统计分析[M].北京:北京大学出版社,2005.
- 12 孙香丽,殷勇.乳制品电子鼻分类中传感器阵列的一种优化方法[J].传感技术学报,2008,21(7):1124~1127.
Sun Xiangli, Yin Yong. Optimization method of sensor array for dairy products identification by electronic nose[J]. Chinese Journal of Sensors and Actuators, 2008, 21(7):1124~1127. (in Chinese)