

DOI:10.3969/j.issn.1000-1298.2010.08.037

# 基于颜色分形的不同产地烟叶聚类分析\*

魏春阳<sup>1</sup> 张云鹤<sup>2</sup> 宋瑜冰<sup>1</sup> 李 锋<sup>1</sup> 薛超群<sup>1</sup> 蔡宪杰<sup>1</sup>

(1. 中国烟草总公司郑州烟草研究院, 郑州 450001; 2. 国家农业信息化工程技术研究中心, 北京 100089)

**【摘要】** 对较高精度的烟叶数字图像进行背景去除和分割预处理, 计算烟叶颜色各分量(红色、绿色、蓝色、亮度)的计盒维数, 绘制各颜色分量的计盒维数变化曲线, 进行烟叶颜色系统聚类分析。结果发现不同产地烟叶的颜色各分量计盒维数变化曲线存在较明显差异; 采用颜色分形维数作为聚类参数得到了最好的分类结果。这表明颜色分形维数可以作为按颜色分类的特征指标, 能更准确地描述图像颜色的分形分布特征。

**关键词:** 烟叶 系统聚类分析 颜色 分形 计盒维数

**中图分类号:** TP391.41; TS44<sup>+</sup>2 **文献标识码:** A **文章编号:** 1000-1298(2010)08-0178-06

## Hierarchical Cluster Analysis of Tobacco Leaves from Different Areas Based on Fractal Color

Wei Chunyang<sup>1</sup> Zhang Yunhe<sup>2</sup> Song Yubing<sup>1</sup> Li Feng<sup>1</sup> Xue Chaoqun<sup>1</sup> Cai Xianjie<sup>1</sup>

(1. Zhengzhou Tobacco Research Institute of CNTC, Zhengzhou 450001, China

2. National Engineering Research Center for Information Technology in Agriculture, Beijing 100089, China)

### Abstract

High definition digital images of tobacco leaves were obtained, which were made on removing the background and segmentation processing. All the components of tobacco leaf color, such as red, green, blue and luminance, were calculated, as new indexes quantitatively described the color distribution state of tobacco leaf surface. The box-counting dimension variation curve of every color component was plotted. A hierarchical cluster analysis was performed on tobacco leaf color with fractal dimension as parameter. The results showed that there was obvious difference in the box-counting dimension variation curve of every color component of tobacco leaves from different growing areas. Cluster analysis with fractal dimension as parameter was very well consistent with tobacco grade experts. These results indicated color fractal dimension, as characteristic index of color classification, could more accurately describe the fractal distributed features of image color.

**Key words** Tobacco leaf, Hierarchical cluster analysis, Color, Fractal, Box-counting dimension

### 引言

在实际烤烟生产中, 同类颜色烟叶在不同产地存在较大差异。同为桔黄颜色的烤烟, 黄淮地区多为浅色桔黄, 西南地区多为金黄, 而东南地区则多为深色桔黄。现有研究多是采用某一相同产地不同颜色或等级的烟叶, 运用计算机图像处理技术建立烤

烟的计算机视觉自动分级系统<sup>[1-5]</sup>, 而烟叶颜色特征仅作为构建识别模型时的一类辅助特征参数; 在烟叶颜色信息提取方面普遍采用颜色直方图计算颜色累计值<sup>[6-8]</sup>, 缺乏考虑烟叶颜色的表面分布特性, 难以揭示相同颜色组烟叶在不同产地间的颜色差异。分形理论为研究一些复杂的自然现象提供了新的方法和手段<sup>[9]</sup>。已有部分研究者将分形理论应

收稿日期: 2009-09-02 修回日期: 2009-09-25

\* 郑州烟草研究院院长基金资助项目(122008CA0400)

作者简介: 魏春阳, 工程师, 主要从事烟叶外观质量评价研究, E-mail: ztrisam@126.com

通讯作者: 蔡宪杰, 高级农艺师, 主要从事烟草栽培及烟叶质量评价研究, E-mail: caixj@ztri.com

用在颜色研究方面<sup>[10-12]</sup>。

本文基于“颜色分形”原理把颜色在烟叶表面的随机分布按统计意义上看成一个分形结构<sup>[13]</sup>,通过对获得的较高精度烟叶数字图像计算烟叶颜色各分量的计盒维数,作为定量描述烟叶颜色表面分布状态的新指标,并将分形维数作为参数进行颜色相似性分类,以揭示相同颜色组烟叶在不同产地间的颜色差异。

## 1 图像获取

烟叶数字图像采集在烟叶分级标准实验室进行<sup>[14]</sup>。采用佳能 EOS40D 型数码单反相机(约 1 050 万像素)。RGB 色彩空间,自动白平衡,ISO100~800。为模拟人眼观测烟叶的角度,相机镜头与工作台面呈 45° 夹角;为保证拍摄角度及焦距的一致性,相机三脚架固定后定角度、定焦距拍摄。图像分辨率为 3 888 × 2 592, JPEG 格式存储。由 USB 数据线及接口将图像直接传至计算机。图像处理软件用 Visual C++、Matlab 6.5 编程语言设计。

## 2 图像预处理

由于烟叶在烘烤过程中失水引起皱缩,所以即使在烟叶自然平铺的舒展状态时,烟叶表面也并非平整,拍摄的烟叶图像会在皱缩的地方出现局部的暗光;同时,需要剔除微小病斑、破损以及主支脉。因此,需要将获取的烟叶数字图像进行分割预处理。首先,用类差影法从图像中去除背景;然后,采用基于子区域的区域生长方法<sup>[15]</sup>,对去除背景的烟叶图像进行分割。烟叶原图像及预处理后图像如图 1 所示。

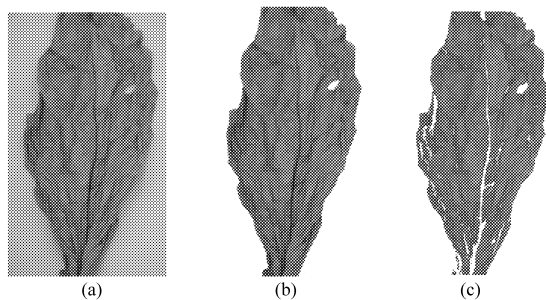


图 1 烟叶原图像及处理后图像

Fig. 1 Original and processed digital images of tobacco leaf

(a) 烟叶原数字图像 (b) 去掉背景后的烟叶图像

(c) 经图像分割去掉暗光、微小病斑以及主支脉后的烟叶图像

## 3 颜色特征参数的提取

### 3.1 基于 RGB 颜色直方图的特征参数

由直方图获取红 ( $R$ )、绿 ( $G$ )、蓝 ( $B$ ) 各个通道

的图像均值。为降低光照变化对颜色特征值的影响,消除 RGB 三色相关性,采用  $r = R/(R + G + B)$ 、 $g = G/(R + G + B)$ 、 $b = B/(R + G + B)$ 、 $G/R$ 、 $G/B$ 、 $R/B$ 、色调  $H$ 、饱和度  $S$  和亮度  $I$  等 9 个参数作为烟叶颜色特征的参数<sup>[16]</sup>。

### 3.2 烟叶颜色分量计盒维数计算

烟叶表面的呈色特性主要是由质体色素类等显色物质在构成烟叶表面的最小单元(叶细胞)内组成比例的不同而呈现出来的。烟叶表面的这种呈色特性,用计盒维数可以确切地反映出来,因为基于计盒维数原理计算出的数字图像盒维数本质上是反映图像(数据点)的空间几何特性。

对于具有严格自相似性的分形,相似维数

$$D_s = \frac{\ln N(r)}{\ln(1/r)} \quad (1)$$

式中  $N(r)$ ——相似元数目

$r$ ——相似元的相似比

通常近似计算计盒维数<sup>[17]</sup>

$$D_B = \lim_{k \rightarrow \infty} \frac{\ln N_{\delta_k}(F)}{-\ln \delta_k} \quad (2)$$

式中  $N_{\delta_k}(F)$ ——与  $F$  相交的  $\delta_k$  网立方体的个数

$\delta_k$ ——图像划分网格的边长

$F$ ——Hausdorff 维数严格大于其拓扑维数的集合

烟叶颜色各分量计盒维数的计算流程如图 2 所示。借助 Fractal Fox 软件<sup>[18]</sup>获得了定量表征烟叶颜色表面分布状态的系列指标: $D_{R1}, D_{R2}, \dots, D_{Rn}$  (描述红色分量的分形分布特征); $D_{G1}, D_{G2}, \dots, D_{Gn}$  (描述绿色分量的分形分布特征); $D_{B1}, D_{B2}, \dots, D_{Bn}$  (描述蓝色分量的分形分布特征); $D_{I1}, D_{I2}, \dots, D_{In}$  (描述亮度分量的分形分布特征)。

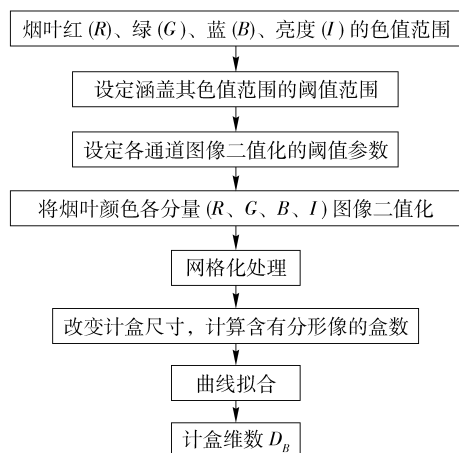


图 2 烟叶颜色分量计盒维数的计算流程图

Fig. 2 Calculation flow chart of box-counting dimension related to each color component of tobacco leaf digital image

## 4 分类原理

设有  $n$  个样本,  $X = \{x_1, x_2, x_3, \dots, x_n\}$  是烤烟样本集, 每个样本测得  $p$  项颜色特征指标(变量),  $Y = \{y_1, y_2, \dots, y_p\}$  是描述烤烟颜色特征的指标集。  $A = (a_{ij})_{n \times p}$  表示烤烟样本集  $X$  关于指标集  $Y$  的指标矩阵, 其中  $a_{ij} = x_i(y_j)$  ( $i = 1, 2, \dots, n; j = 1, 2, \dots, p$ )。任何两个样本之间的颜色相似性, 可以通过欧氏距离来描述。欧氏距离

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_i y_k - x_j y_k)^2} \quad (3)$$

烟叶颜色分类基本步骤为: 各样本颜色自成一类, 采用欧氏距离, 先求出各类间的距离矩阵  $D(0)$ , 然后采用离差平方和法, 每次缩小一类, 选择使  $S$  增加最小的两类合并, 直到所有的样本归为一类为止。最后绘制聚类图<sup>[19]</sup>。

## 5 颜色特征参数分析及烟叶颜色分类

### 5.1 烟叶颜色特征统计描述

样本烟叶的 RGB 颜色特征参数值的统计数据如表 1 所示, 样本烟叶的颜色分量值均表现为  $R > G \gg B$ ; 不同产地烟叶的颜色之间存在一定差异, 其中差异最大的为蓝色分量值。

表 1 基于 RGB 的烟叶颜色特征参数描述统计

Tab.1 Characteristic parameters descriptive statistics of tobacco leaf color based on RGB color model

颜色特征参数	最小值	最大值	平均	标准偏差	极差	变异系数/%
$R$	167	212	188	14.36	45	7.64
$G$	108	167	135	16.99	59	12.63
$B$	13	28	19	4.48	14	23.34
$r$	0.53	0.58	0.55	0.02	0.05	3.21
$g$	0.37	0.41	0.39	0.01	0.04	3.41
$b$	0.04	0.08	0.06	0.01	0.04	22.22
$G/R$	0.65	0.78	0.71	0.04	0.14	6.07
$G/B$	4.83	10.43	7.35	1.61	5.61	21.88
$R/B$	6.50	14.70	10.38	2.46	8.20	23.69
$H$	0.66	0.81	0.73	0.05	0.16	6.86
$S$	0.76	0.89	0.83	0.04	0.13	4.47
$I$	115	164	138	14.35	49	10.41

注: 供试样本是从全国 7 个主要烤烟省区选取的 2008 年初烤 C3F 等级烟叶。

### 5.2 烟叶颜色分量计盒维数分析

根据烟叶 RGB 颜色特征参数统计值(表 1), 设

定各颜色分量的阈值范围, 并将其分为 8 个小阈值范围, 作为烟叶不同颜色分量进行二值化的阈值参数, 计算得到定量表征烟叶颜色表面分布状态的系列指标:  $D_{R1}, D_{R2}, \dots, D_{R8}; D_{G1}, D_{G2}, \dots, D_{G8}; D_{B1}, D_{B2}, \dots, D_{B8}; D_{I1}, D_{I2}, \dots, D_{I8}$ 。

烟叶颜色各分量不同阈值范围的计盒维数越接近于 2, 表明对应阈值范围的颜色点越趋向于充满整个烟叶图像范围分布; 计盒维数在 1 附近, 表明对应阈值范围的颜色点趋向于沿烟叶图像范围内某条直线分布; 计盒维数接近于零, 表明对应阈值范围的颜色点是在烟叶图像范围内离散分布的一些小区域。

如图 3 所示, 亮度分量值在 130 ~ 150 范围趋向分布在整个烟叶表面, 在 90 ~ 100 范围趋向于沿直线分布(图 3a); 红色分量值在 180 ~ 200 范围趋向分布在整个烟叶表面, 在 140 ~ 150、210 ~ 220 范围趋向于沿直线分布(图 3b); 绿色分量值在 130 ~ 150 范围趋向分布在整个烟叶表面, 在 90 ~ 100 范围趋向于沿直线分布(图 3c); 蓝色分量值在 10 ~ 25 范围趋向分布在整个烟叶表面(图 3d)。这与提取的颜色特征均值( $I: 144, R: 194, G: 142, B: 21$ )相吻合, 表明不同阈值下烟叶颜色分量的分形维数既能反映颜色的空间分布特征, 又包含其累计特性。这与冯斌等<sup>[11]</sup>的研究结论一致。

运用涵盖烟叶颜色信息的阈值范围分段计算其颜色信息对应的分形维数, 形成了一条分形维数变化曲线, 曲线上的每个数据点反映了对应阈值范围的颜色分形维数。产地之间烟叶颜色各分量的分形维数变化曲线存在着较明显差异(图 4)。福建及湖南产区烟叶图像亮度分量的分形维数变化曲线呈波浪型; 云南、贵州和四川产区烟叶图像亮度分量的分形维数变化曲线呈倒勾型, 而河南和山东产区烟叶图像亮度分量的分形维数变化曲线近似呈抛物线(图 4a)。不同产地烟叶红色分量(图 4b)和绿色分量(图 4c)图像对应的分形维数变化曲线线型与上述描述相似; 蓝色分量图像(图 4d)对应的分形维数变化曲线线型略有不同。

### 5.3 基于颜色分形维数的烟叶颜色分类

如图 5 所示, 采用颜色分形维数作为聚类参数得到了最好的分类结果(图 5a), 分类结果与分级专家的判定结果相吻合; 采用基于 RGB 的颜色特征参数作为聚类参数得到的分类结果(图 5b)存在一定的出入; 采用亮度和绿色分量图像的计盒维数作为聚类参数也得到了满意的分类结果(图 5c、图 5e)。由此可见, 采用颜色分形维数作为烟叶颜色分类参数是合理的。

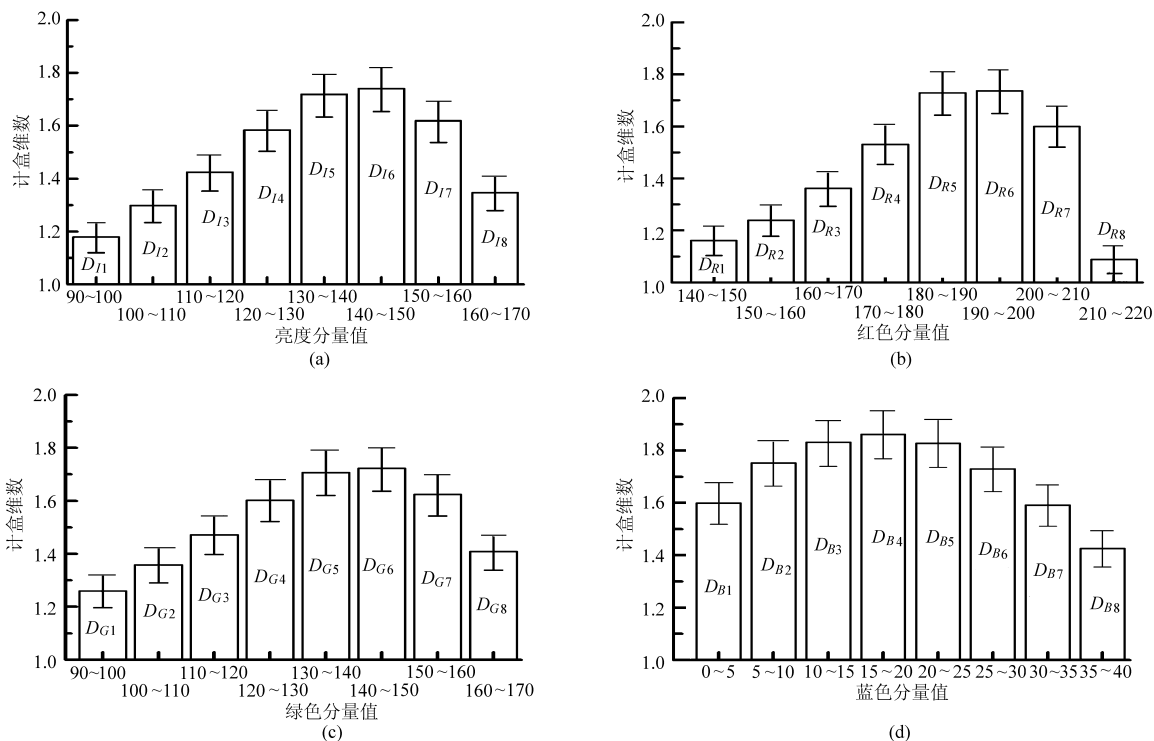


图 3 不同阈值参数下云南弥勒产 C3F 等级烟叶各颜色分量计盒维数

Fig. 3 Box-counting dimension related to each color component of tobacco leaf digital image under different threshold parameters of C3F grades tobacco leaf produced by Mile County in Yunnan Province

(a) 亮度分量 (b) 红色分量 (c) 绿色分量 (d) 蓝色分量

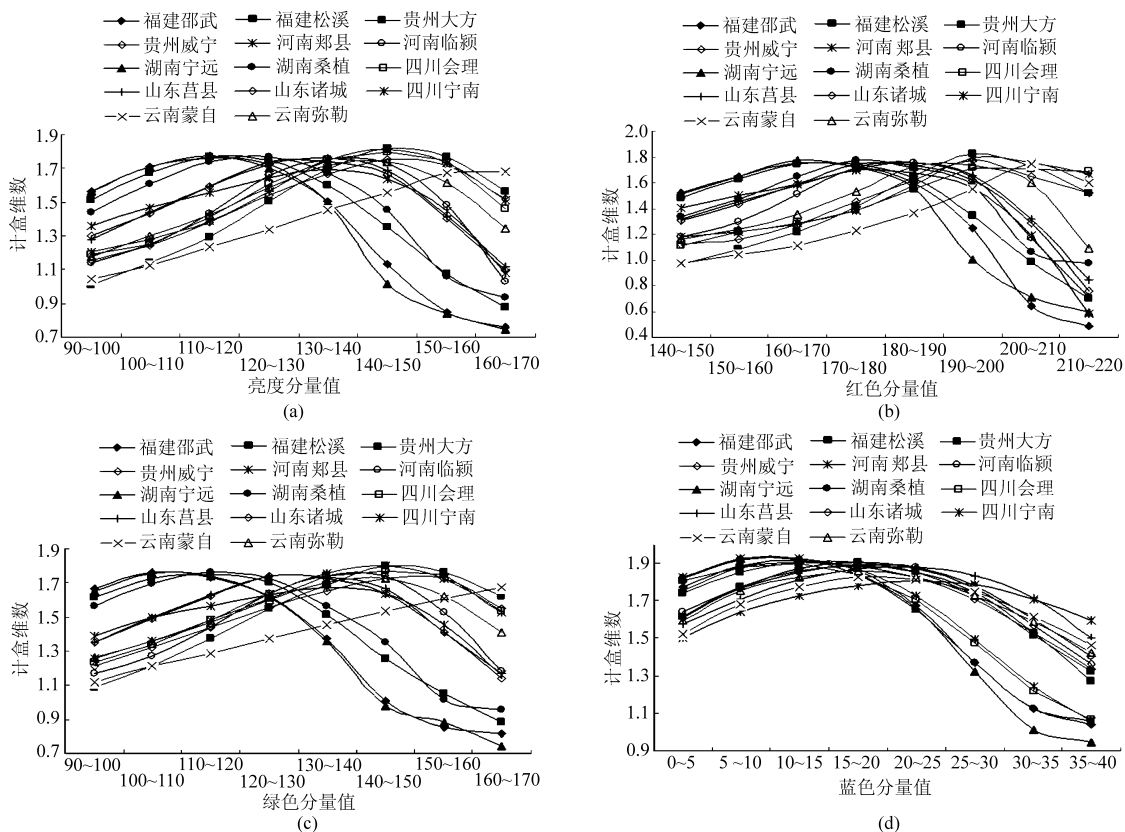


图 4 不同阈值参数下不同产地 C3F 等级烟叶各颜色分量计盒维数变化曲线

Fig. 4 Curves of box-counting dimension related to each color component of tobacco leaf digital image under different threshold parameters of C3F grades tobacco leaf from different tobacco-producing areas

(a) 亮度分量 (b) 红色分量 (c) 绿色分量 (d) 蓝色分量

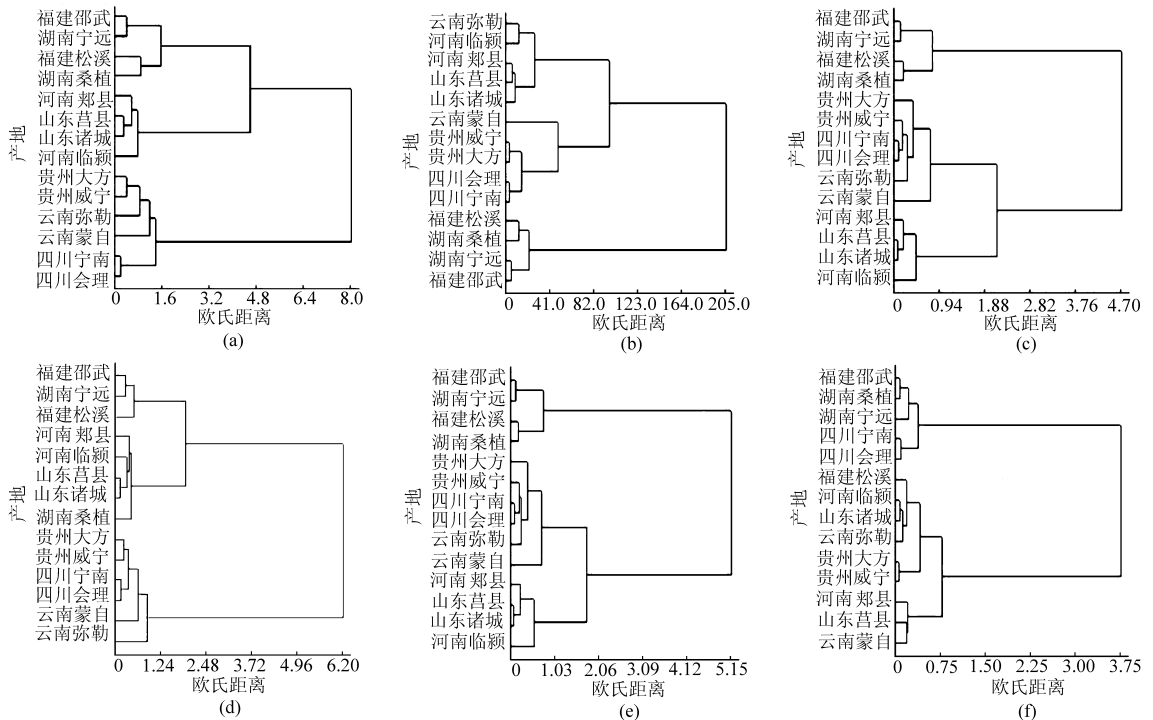


图5 不同颜色特征参数组合下系统聚类结果(软件:DPS9.50,聚类距离:欧氏距离,聚类方法:离差平方和法)

Fig.5 Hierarchical clustering results under combinations of different characteristics parameters of the tobacco leaf digital image color (software: DPS9.50, clustering distance: Euclidean distance, clustering method: ward method)

- (a) 聚类参数为RGB三色及亮度图像不同阈值下计盒维数( $D_{R1}, D_{R2}, \dots, D_{R8}; D_{G1}, D_{G2}, \dots, D_{G8}; D_{B1}, D_{B2}, \dots, D_{B8}; D_{I1}, D_{I2}, \dots, D_{I8}$ ) (b) 聚类参数为RGB三色均值以及相关换算颜色参数( $R, G, B, r, g, b, G/R, G/B, R/B, H, S, I$ ) (c) 聚类参数为亮度图像不同阈值下计盒维数( $D_{I1}, D_{I2}, \dots, D_{I8}$ ) (d) 聚类参数为红色分量R图像不同阈值下计盒维数( $D_{R1}, D_{R2}, \dots, D_{R8}$ ) (e) 聚类参数为绿色分量G图像不同阈值下计盒维数( $D_{G1}, D_{G2}, \dots, D_{G8}$ ) (f) 聚类参数为蓝色分量B图像不同阈值下计盒维数( $D_{B1}, D_{B2}, \dots, D_{B8}$ )

## 6 结论

- (1) 颜色分形维数既能够反映颜色的空间分布特征,又包含其累计特性。
- (2) 不同产地烟叶的颜色各分量分形维数变化

曲线存在较明显差异。

- (3) 研究表明颜色分形维数可以作为颜色分类的特征指标,且更准确地描述图像颜色的分形分布特征。

## 参 考 文 献

- 1 张建平,吴守一,方如明,等. 农产品质量的计算机辅助检验与分级(第II报)—烟叶自动分级模型的建立与训练[J]. 农业工程学报,1997,13(4):179~183.  
Zhang Jianping, Wu Shouyi, Fang Ruming, et al. Computer aided quality inspecting and grading for agricultural products (2nd half)—construction and training of grading model for tobacco leaf[J]. Transactions of the Chinese Society of Agricultural Engineering, 1997, 13(4): 179~183. (in Chinese)
- 2 张惠民,韩力群,段正刚. 基于图像特征的烟叶分级[J]. 武汉大学学报:信息科学版,2003,28(3):359~362.  
Zhang Huimin, Han Liqun, Duan Zhenggang. Tobacco grading based on image features [J]. Geomatics and Information Science of Wuhan University, 2003, 28(3):359~362. (in Chinese)
- 3 阎瑞琼. 烤烟烟叶的计算机辅助分级系统研究[D]. 北京:北京工商大学,2001.  
Yan Ruiqiong. The computer-aided classification system of flue-cured tobacco [D]. Beijing: Beijing Technology and Business University, 2001. (in Chinese)
- 4 李浩. 基于数字图像处理技术的烤烟烟叶自动分组模型研究[D]. 武汉:华中农业大学,2007.  
Li Hao. The research on auto-grouping model of flue-cured tobacco leaves based on digital image processing [D]. Wuhan: Huazhong Agricultural University, 2007. (in Chinese)
- 5 谭旭,唐云岚,陈英武. 基于粗糙集的烤烟烟叶智能分级方法[J]. 农业机械学报,2009,40(6):169~174.

- Tan Xu, Tang Yunlan, Chen Yingwu. Intelligent grading of flue-cured tobacco leaves based on rough set theory [J]. Transactions of the Chinese Society for Agricultural Machinery, 2009, 40(6): 169 ~ 174. (in Chinese)
- 6 张帆, 方如明, 蔡健荣. 标准烟叶数据库的图像检索 [J]. 农业机械学报, 2001, 32(1): 66 ~ 68.  
Zhang Fan, Fang Ruming, Cai Jianrong. Image retrieval of standard tobacco leaf database [J]. Transactions of the Chinese Society for Agricultural Machinery, 2001, 32(1): 66 ~ 68. (in Chinese)
- 7 刘华波, 贺立源, 马文杰, 等. 透射图像颜色特征在烟叶识别中应用的探索 [J]. 农业工程学报, 2007, 23(9): 169 ~ 171.  
Liu Huabo, He Liyuan, Ma Wenjie, et al. Color features of transmittance images applied in the classification of tobacco leaves [J]. Transactions of the Chinese Society of Agricultural Engineering, 2007, 23(9): 169 ~ 171. (in Chinese)
- 8 梁魏峰. 应用 MATLAB 提取烟叶图像的基本特征 [D]. 济南: 山东大学, 2003.  
Liang Weifeng. Extraction the basic features of the tobacco image application with MATLAB [D]. Ji'nan: Shandong University, 2003. (in Chinese)
- 9 Mandelbrot B B. The fractal geometry of nature [M]. New York: W. H Freeman & Co., Ltd, 1983.
- 10 Liu J, Furuno T. The fractal estimation of wood color variation by the triangular prism surface area method [J]. Wood Science and Technology, 2002, 36(5): 385 ~ 397.
- 11 冯斌, 汪懋华. 基于颜色分形的水果计算机视觉分级技术 [J]. 农业工程学报, 2002, 18(2): 141 ~ 144.  
Feng Bin, Wang Maohua. Computer vision classification of fruit based on fractal color [J]. Transactions of the Chinese Society of Agricultural Engineering, 2002, 18(2): 141 ~ 144. (in Chinese)
- 12 李志臣, 姬长英. 基于图像分析的杂草分形维数计算 [J]. 农业工程学报, 2006, 22(11): 175 ~ 178.  
Li Zhichen, Ji Changying. Calculation of weed fractal dimension based on image analysis [J]. Transactions of the Chinese Society of Agricultural Engineering, 2006, 22(11): 175 ~ 178. (in Chinese)
- 13 张济忠. 分形 [M]. 北京: 清华大学出版社, 1995: 111 ~ 141.
- 14 YC/T 291-2009. 烟叶分级实验室环境条件 [S].  
YC/T 291-2009. Ambient condition of tobacco grading laboratory [S]. (in Chinese)
- 15 王艳娟, 陈晓红, 黄晓欣. 图像感兴趣区域自动提取算法 [J]. 科学技术与工程, 2007, 7(12): 2 867 ~ 2 871.  
Wang Yanjuan, Chen Xiaohong, Huang Xiaoxin. Automatic extraction of regions of interest [J]. Sci. Tech. Engng., 2007, 7(12): 2 867 ~ 2 871. (in Chinese)
- 16 蔡鸿昌, 崔海信, 宋卫堂, 等. 黄瓜初花期叶片光合色素含量与颜色特征的初步研究 [J]. 农业工程学报, 2006, 22(9): 34 ~ 38.  
Cai Hongchang, Cui Haixin, Song Weitang, et al. Preliminary study on photosynthetic pigment content and color feature of cucumber initial bloom stage [J]. Transactions of the Chinese Society of Agricultural Engineering, 2006, 22(9): 34 ~ 38. (in Chinese)
- 17 彭瑞东, 谢和平, 鞠扬. 二维数字图像分形维数的计算方法 [J]. 中国矿业大学学报, 2004, 33(1): 19 ~ 24.  
Peng Ruidong, Xie Heping, Ju Yang. Computation method of fractal dimension for 2-D digital image [J]. Journal of China University of Mining & Technology, 2004, 33(1): 19 ~ 24. (in Chinese)
- 18 Francesco Chiappelli, Joseph Alwan, Paolo Prolo, et al. Neuro-immunity in stress-related oral ulcerations: a fractal analysis [J]. Frontiers in Bioscience, 2005, 10(Supp.): 3 034 ~ 3 041.
- 19 唐启义, 冯明光. DPS 数据处理系统 [M]. 北京: 科学出版社, 2007.
- 20 吴兰兰, 刘俭英, 文友先. 基于分形维数的玉米和杂草图像识别 [J]. 农业机械学报, 2009, 40(3): 176 ~ 179.  
Wu Lanlan, Liu Jianying, Wen Youxian. Image identification of corn and weed based on fractal dimension [J]. Transactions of the Chinese Society for Agricultural Machinery, 2009, 40(3): 176 ~ 179. (in Chinese)