

牛膝和川牛膝药材的特征提取与图像识别[△]

王耐*, 卢文彪#, 凌秀华, 梁丽金, 李熙灿, 李睿(广州中医药大学中药学院, 广州 510006)

中图分类号 R282.5 文献标志码 A 文章编号 1001-0408(2017)12-1670-04

DOI 10.6039/j.issn.1001-0408.2017.12.24

摘要 目的:提取牛膝和川牛膝药材的特征,并建立其图像识别方法。方法:采用MATLAB软件编程拼接牛膝和川牛膝药材的横切面显微图像,提取颜色、不变矩、纹理和横切面维管束组织特征;将数据整理成数据矩阵,通过Zscore函数对数据矩阵进行标准化,通过Princomp函数进行主成分分析;采用BP神经网络识别模式。结果:药材样品显微图像拼接处的组织细节保持完整;测得每组药材样品图像的27个特征参数,经主成分分析,选取11个主成分参数建立BP神经网络,两种近缘药材样本($n=50$)的BP神经网络平均识别率为100%。结论:该方法可用于中药材显微图像自动拼接,及牛膝和川牛膝药材的图像识别。

关键词 牛膝;川牛膝;图像拼接;模式识别;MATLAB软件

Feature Extraction and Image Recognition of *Achyranthes bidentata* and *Cyathula officinalis*

WANG Nai, LU Wenbiao, LING Xiuhua, LIANG Lijin, LI Xican, LI Rui (College of TCM, Guangzhou University of TCM, Guangzhou 510006, China)

ABSTRACT OBJECTIVE: To extract the feature of *Achyranthes bidentata* and *Cyathula officinalis*, and to establish image recognition method. METHODS: The microscopic image stitching of *A. bidentata* and *C. officinalis* was implemented by MATLAB. The color, invariant moment, stripes and the features of vascular bundle in cross section were extracted. The data was organized into data matrix, and then data matrix was standardized by Zscore function; principal components were analyzed through Princomp function. BP nerve network recognition mode was adopted. RESULTS: The microstructures in the micro images of the samples were kept integrated. The measured data of 27 characteristics were acquired in each group of sample. Through principal component analysis, the parameters of 11 main components were selected to establish BP never network. The average recognition rate of BP nerve network was 100% between 2 medicinal material relatives ($n=50$). CONCLUSIONS: The method can be used for micro-image auto stitching of Chinese medicinal materials and image recognition of *A. bidentata* and *C. officinalis*.

KEYWORDS *Achyranthes bidentata*; *Cyathula officinalis*; Image stitching; Pattern recognition; MATLAB

在医学图像分析等研究中经常会涉及到图像拼接问题,这就需要获取一定范围的完整图像才易于全面地观察和分析研究对象,从而获取重要信息^[1]。在中药显微鉴别中,若通过逐一视野对显微图像进行分析统计,使得图像中不完整的部分易产生误差,且有许多组织特征难于实现参数化表征,不利于批量图像的快速、准确分析。2015年版《中国药典》(一部)将牛膝 *Achyranthes bidentatae* RL.和川牛膝 *Cyathulae officinalis* 作为两种药材收录^[2]。目前,牛膝和川牛膝药材的传统鉴别方法主要是性状鉴别和显微鉴别^[3]。因此,笔者以牛膝和川牛膝这两种近缘药材为试样,针对它们的性状特征、维管束特征等差别,采用MATLAB编程实现药材横切面显微图像的自动拼接以及药材特征的定量描述与识别,为中药材的快速鉴别与质量评价奠定基础。

1 材料

△ 基金项目:国家自然科学基金资助项目(No.81573558)

* 硕士研究生。研究方向:中药及其制剂质量标准。E-mail: 18819170315@163.com

通信作者:副教授,硕士生导师。研究方向:中药及其制剂质量标准。E-mail: 962212925@qq.com

1.1 仪器

XSP-35TV型生物显微镜(成都一科仪器设备有限公司);TDCE型电子目镜[江南光电(集团)股份有限公司];D3200型数码相机(日本Nikon公司);YL3型回旋式切片机(上海仪表厂);MATLAB软件(R2013a版,美国MathWorks公司)。

1.2 试剂

聚乙二醇、水合氯醛均为化学纯。

1.3 药材

牛膝(批号:160401、160709、160728)、川牛膝(批号:151003、160317、160210)均购自广州清平药材市场,经广州中医药大学黄海波副教授鉴定为真品。

2 方法与结果

2.1 图像采集

2.1.1 药材样品图像 两种药材样品随机各抽取25根,置于白纸上,保持相机与药材样品每次的拍摄距离(约10 cm)和光照条件(自然光)相同,摄取50张药材样品图像。采用MATLAB软件将摄取的图像转至HSV(Hue、Saturation、Value,即颜色、饱和度、亮度)颜色空间,在S

通道上用最大类间方差法设置分割阈值,实现目标物与背景的分割,经裁剪后备用,详见图1。

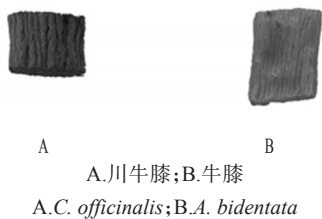


图1 药材样品图像

Fig 1 Image of medicinal material

2.1.2 药材横切片样品显微图像 取药材样品中段,以聚乙二醇包埋制片法制成横切片,装片。在显微镜下按一定顺序移动视野,覆盖整个横切片,保持每次放大倍率(40×)相同,控制图像在HSV空间的V均值范围为0.4~0.8,并使各视野间有一定程度重叠(10%~30%),以电子目镜摄取各视野图像,经小波除噪,编号保存。将采集的彩色图像(RGB)转换为灰度图像,在两幅图像的重叠区域中选定匹配区域,采用基于灰度相关的区域匹配法^[4],用MATLAB软件内置函数Normxcorr2计算匹配区域中对应像素点之间的灰度值相关系数,以最大相关系数对应的坐标为两幅图像的配准坐标;计算两幅图像配准区域的平均灰度差值,用直接平均灰度调整法^[5]调整两幅图像的灰度值,从而消除图像拼接处较

大的灰度梯度;根据配准坐标自动裁去图像拍摄时平移错位引起的冗余部分,合并两图像矩阵完成拼接。若旋转错位,则增加对待拼接图像进行角度扫描,以得到最大相关系数对应的配准坐标及偏转角度校正值。依照两幅图像的拼接方法,设定循环依次读入待拼接图像,先按行/列依次拼接,再按列/行依次拼接,得到牛膝与川牛膝横切片的拼接图像,详见图2。

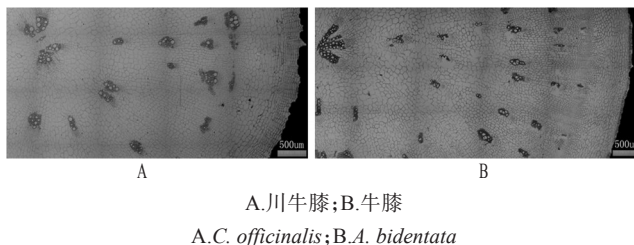


图2 药材切片的部分拼接图像

Fig 2 Some stitched images of medicinal material sections

2.2 药材样品特征提取

2.2.1 颜色特征 利用图像的颜色矩可以简单有效地描述图像的颜色分布。图像的颜色信息主要分布于低阶矩中,所以用颜色的一阶矩(均值)、二阶矩(方差)、三阶矩(斜度)足以表达图像的颜色分布^[6]。将药材样品图像转为HSV颜色空间,分别计算H、S、V分量相应的3个颜色矩,结果见表1。

表1 药材样品的颜色参数测量结果($\bar{x} \pm s, n=25$)

Tab 1 Results of color parameters measurement of medicinal materials($\bar{x} \pm s, n=25$)

药材	H颜色矩			S颜色矩			V颜色矩		
	一阶矩	二阶矩	三阶矩	一阶矩	二阶矩	三阶矩	一阶矩	二阶矩	三阶矩
牛膝	0.001±0.0003	0.01±0.001	0.02±0.002	0.006±0.001	0.06±0.006	0.12±0.009	0.006±0.002	0.05±0.008	0.11±0.013
川牛膝	0.0009±0.0002	0.009±0.001	0.02±0.001	0.006±0.001	0.06±0.007	0.13±0.010	0.003±0.001	0.03±0.004	0.08±0.007

2.2.2 几何不变矩(Hu矩)特征 Hu不变矩描述了图像形状的全局特征,具有缩放、平移、旋转不变性,在图像识别的特征提取中被广泛应用^[7]。将药材样品图像转化

为灰度图像,提取目标物的7个Hu矩,将Hu矩绝对值的对数作为测量结果,结果见表2。

2.2.3 纹理特征 灰度共生矩阵是描述图像纹理特征

表2 药材样品的Hu不变矩参数测量结果($\bar{x} \pm s, n=25$)

Tab 2 Results of Hu invariant moment parameters measurement of medicinal materials($\bar{x} \pm s, n=25$)

药材	lg Hu矩1	lg Hu矩2	lg Hu矩3	lg Hu矩4	lg Hu矩5	lg Hu矩6	lg Hu矩7
牛膝	6.37±0.09	15.67±1.33	24.36±1.90	25.65±1.12	51.23±2.39	34.22±2.07	51.86±2.49
川牛膝	5.96±0.07	14.77±1.23	22.98±1.04	23.45±1.16	47.12±2.21	31.49±2.03	47.71±1.78

的常用方法。由灰度共生矩阵产生的4个纹理特征参数[对比度(Contrast)、相关性(Correlation)、能量(Energy)、同质性(Homogeneity)]可以有效地描述目标的纹理特征^[8]。将药材图像转化为灰度图像,用MATLAB软件内置的Graycomatrix函数和Graycoprops函数计算图像灰度共生矩阵的4个纹理参数,结果见表3。

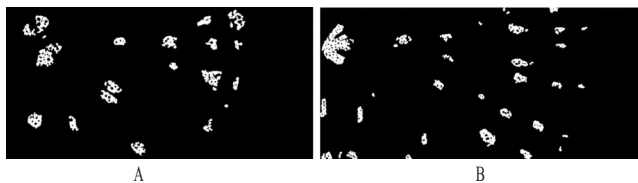
表3 药材样品的纹理参数测量结果($\bar{x} \pm s, n=25$)

Tab 3 Results of stripe parameters measurement of medicinal materials($\bar{x} \pm s, n=25$)

药材	对比度	相关性	能量	同质性
牛膝	0.004±0.0009	0.979±0.005	0.976±0.005	0.998±0.0003
川牛膝	0.003±0.0006	0.956±0.006	0.981±0.005	0.999±0.0003

2.2.4 横切面维管束组织特征 维管束组织特征是牛膝与川牛膝的重要显微鉴别特征^[9]。采用数学形态学运算和基于改进的4个方向边缘检测Sobel算子^[10]进行药材样品横切面拼接图像的二值化以及维管束的分割。其中,以维管束中的纤维、导管等厚壁组织为目标分割所得到的图像块(500像素以上)作为维管束组织特征提取的目标图像块,结果见图3。

由于药材样品切片整体形状呈类圆形,故选择极坐标系来描述各目标图像块的位置关系;将二值图中各目标图像块封闭、填充后,用MATLAB软件内置的Regionprops函数求出各图像块的质心、整个切片的质心



A.川牛膝;B.牛膝

A.C. officinalis;B.A. bidentata

图3 药材样品拼接图像中维管束分割结果

Fig 3 Results of vascular bundle segmentation in stitched images of medicinal materials

(作为极坐标的极点)及整个切片的总面积(以像素为单位,下同)。在上述基础上进行下列测量——(1)计算维管束间夹角,并计算均值及RSD;计算各维管束图像块的质心点在极坐标中的角度值,再依次求出各图像块与逆时针角向最近邻图像块质心点之间的夹角值,计算夹角平均值及RSD。(2)测量维管束间距离RSD:计算各图像块质心点到极点的距离,再依次求出各图像块与其最近邻图像块质心点之间的距离,统计图像块之间距离的RSD。(3)测量维管束的长宽比:测量各图像块的长与宽的比值,统计长宽比的平均值。(4)测量单个维管束面积的RSD:测量各图像块的面积,统计面积的RSD。(5)测量维管束的个数和总面积分数:测出切片中图像块的个数及其面积之和,用图像块面积之和除以切片总面积得到维管束的总面积分数,结果见表4。

表4 药材样品横切片图像中维管束特征参数测量结果($\bar{x} \pm s, n=25$)

Tab 4 Results of parameters measurement of features for vascular bundle in cross section image of medicinal materials($\bar{x} \pm s, n=25$)

药材	夹角均值,度	夹角RSD,%	间距RSD,%	长宽比均值	维管束面积RSD,%	维管束个数,个	维管束总面积分数
牛膝	4.5±1.1	0.86±0.12	4.7±0.6	1.47±0.09	1.49±0.59	78.4±15.7	0.03±0.01
川牛膝	4.3±0.3	0.84±0.09	4.8±0.3	1.39±0.04	1.95±1.20	83.3±5.80	0.08±0.03

2.3 主成分分析与模式识别

2.3.1 主成分分析 将表1~表4中27项特征参数的原始数据($n=50$)整理成数据矩阵(50×27),用MATLAB软件内置的Zscore函数对数据矩阵标准化,用Princomp函数进行主成分分析,得到载荷矩阵(27×27)、主成分空间表示原数据的矩阵(50×27)、协方差矩阵的特征值向量(27×1),样本协方差矩阵特征值和累积贡献率见表5。

表5 主成分分析协方差矩阵特征值和累积贡献率

Tab 5 Feature value and accumulative contribution rate for covariance matrix of principle component analysis

项目	1	2	3	4	5	6	7	8	9	10	11	12	...	27
特征值	12.05	5.06	1.75	1.62	1.33	1.06	0.93	0.71	0.53	0.49	0.43	0.41	...	0.000
累积贡献率,%	44.6	63.4	69.9	75.9	80.8	84.7	88.2	90.8	92.7	94.6	96.2	97.7	...	100

2.3.2 模式识别 在主成分空间矩阵(50×27)中选取特征值累计贡献率 $\geq 95\%$ 的前11个成分(50×11)作为模式识别的输入数据。BP神经网络是目前应用广泛的神

经网络模型,它能学习和存贮大量的输入-输出模式映射关系。将两种药材样品主成分分析后的特征数据整理成数据矩阵(50×11),并将两种样本的性状特征数据矩阵(50×20)以及维管束组织特征数据矩阵(50×7)作为对照,从中随机选取60%的样本(保持两种样本的数目相等)作为训练样本,用MATLAB软件内置的BP神经网络工具箱进行模式识别:输入节点数为11(11个主成分),隐层节点数为15(性状特征识别隐层为25,维管束特征识别隐层为12),隐层节点数应大于输入节点和输出节点之和,输出节点数为2(两种药材样品),隐层的转移函数是Tansig,输出层的转移函数是Purelin。随机选取3次不同的训练样本构建BP神经网络,结果见表6。

表6 BP神经网络识别结果

Tab 6 Results of BP nerve network recognition

类别	总样本数,个	识别数,个	识别率,%	平均识别率,%
性状特征	50	50	100	98
	50	48	96	
	50	49	98	
维管束特征	50	47	94	95.3
	50	46	92	
	50	50	100	
性状特征+维管束特征	50	50	100	100
	50	50	100	
	50	50	100	

3 讨论

本研究采用的图像拼接算法精度高,简便实用。从拼接的药材样品切片显微图像中能较清晰地观察到各种组织及其分布情况,可满足图像后续处理的要求;该方法对拍摄过程中出现的平移错位、小幅度旋转情况具有良好的鲁棒性。在拍摄过程中,不需要使每行或每列的视野图像数目保持一致,但要保证各视野间有部分重叠;应注意获得清晰的图像以及尽可能消除显微镜聚光效应,以便简化图像的后续处理。

牛膝和川牛膝是两种近缘药材,它们的许多组织形态特征差别不大。本研究结果表明,结合药材样品颜色特征、不变矩特征、纹理特征和维管束组织特征等多种形态参数进行模式识别,可以客观、快捷、方便地鉴别两种药材样品。本研究是数字图像处理技术与中药鉴定方法结合的应用,编程实现药材样品图像的图像拼接、特征提取与模式识别有利于批量图像的快速处理。但对药材样品其他精细结构的表征与识别,还有待深入研究。

参考文献

- [1] 李娟,张宝昌,孙娟.浅谈基于MATLAB的图像拼接技术在医学图像处理中的应用[J].济宁医学院学报,2009,32(6):433-436.
- [2] 国家药典委员会.中华人民共和国药典:一部[S].2015年版.北京:中国医药科技出版社,2015:38,72.
- [3] 肖随安,王焱.怀牛膝与川牛膝的鉴别及其临床应用的异同[J].中国民间疗法,2013,21(12):69-70.

高产高油桂郁金种质的筛选[△]

靳雅惠*, 杨传强, 王 建[#](广西中医药大学药学院, 南宁 530001)

中图分类号 R284.1 文献标志码 A 文章编号 1001-0408(2017)12-1673-04

DOI 10.6039/j.issn.1001-0408.2017.12.25

摘要 目的:建立分析高产高油桂郁金挥发油的方法,以期为桂郁金良种选育繁育提供参考。方法:采用气相色谱-质谱联用法。色谱条件:色谱柱为HP-5MS石英弹性毛细管柱,载气为高纯氦气(99.999%),载气流量为1.0 mL/min,进样口温度为250℃,色谱柱初始温度为50℃(程序升温),分流进样,分离比为10:1。质谱条件:电离方式为电子轰击离子源,电离能量为70 eV,离子源温度为230℃,四极杆温度为150℃,传输线温度为280℃,电子倍增器电压为1 588 V,质量扫描范围为 m/z 45~500。从100株不同种质样品中筛选产量大、含油量高的桂郁金较好种质,比较筛选单株的挥发油成分及各成分的相对含量,采用SPSS 22.0统计软件进行聚类分析。结果:共筛选出10个桂郁金高产高油种质,鉴定出54种化合物,10个不同种质的桂郁金挥发油共有成分为樟脑、1-石竹烯、 γ -榄香烯、莜术烯、吉马酮、新莜术二酮,大多数种质含有龙脑、异龙脑、 δ -榄香烯、大根香叶烯、白菖烯。C78、C104、药用2的产量、挥发油含量以及挥发油中有效成分相对含量均较高。10个桂郁金高产高油种质可以分为3类。结论:该研究基本明确了高产高油桂郁金挥发油的主要化学成分,C78、C104、药用2为高产高油桂郁金种质中较优良的品系。

关键词 桂郁金;种质;挥发油成分;高产;高油

Screening of High-yield and High-oil *Curcuma kwangsiensis*

JIN Yahui, YANG Chuanqiang, WANG Jian (School of Pharmacy, Guangxi University of Chinese Medicine, Nanning 530001, China)

ABSTRACT OBJECTIVE: To establish a method for analyzing the volatile oil of high-yield and high-oil *Curcuma kwangsiensis*, and to provide reference for its breeding. METHODS: GC-MS was performed on the column of HP-5MS quartz elastic capillary column, carrier gas was high purity helium (99.999%), flow rate was 1.0 ml/min, inlet temperature was 250℃, the initial temperature of column was 50℃ (temperature programmed), split injection with split ratio of 10:1. Mass spectrometry conditions: ionization mode was the electron impact ion source, ionization energy was 70 eV, the ion source temperature was 230℃, the quadrupole temperature was 150℃, transmission line temperature was 280℃, the electron multiplier voltage was 1 588 V, and mass scanning range was m/z 45 to 500. High-yield and high-oil germplasm of were screened from 100 different germplasms, the volatile oil composition of single plant and relative percentage contents of each composition were compared, SPSS 22.0 software was used for cluster analysis. RESULTS: Totally 10 high-yield and high-oil germplasm were screened, 54 kinds of compounds were identified, the common compositions of 10 different germplasms were camphor, 1-caryophyllene, γ -elemene, curcumene, gemma ketone and new eurdione, most germplasms contained borneol, isoborneol, δ -elemene, germacrene and calamine. The C78, C104, purpose 2, volatile oil content and relative percentage contents of active compositions in volatile oil were high. The 10 high-yield and high-oil germplasms can be divided into 3 groups. CONCLUSIONS: The study basically clears the main chemical

[4] 林陆君,孙玲玲,李训根,等.一种改进的基于模板匹配的显微细胞图像拼接算法[J].计算机应用与软件,2010,27(1):109-110.

[5] 黄梅,唐琨,肖建新.基于特征点的图像拼接技术在动漫中的应用[J].湖南师范大学自然科学学报,2016,39(1):50-54.

[6] 杨红菊,张艳,曹付元.一种基于颜色矩和多尺度纹理特征的彩色图像检索方法[J].计算机科学,2009,36(9):

274-276.

[7] 田秋红,孙政荣.基于Hu不变矩和BP网络的条形码图像识别方法[J].计算机工程与设计,2012,33(4):1563-1567.

[8] 高程程,惠晓威.基于灰度共生矩阵的纹理特征提取[J].计算机系统应用,2010,19(6):195-198.

[9] 吴智高.怀牛膝与川牛膝的鉴别[J].海峡药学,2007,19(6):77-78.

[10] 李灿灿,孙长辉,王静,等.基于改进的Sobel算子和色调信息的叶脉提取方法[J].农业工程学报,2011,27(7):197-199.

[△] 基金项目:国家自然科学基金资助项目(No.81160500);广西科技厅课题(No.桂科攻1099063-2);广西研究生教育创新计划资助项目(No.YCSZ2015181)

* 硕士研究生。研究方向:药用植物栽培和育种。E-mail: sxyajyh@126.com

[#] 通信作者:教授。研究方向:药用植物栽培和育种。E-mail: gxzyywj@126.com

(收稿日期:2016-11-17 修回日期:2016-12-21)

(编辑:张 静)