

基于图像矩的板材细胞图像取样方法研究

王建华¹ 高巍巍² 赵磊²

(哈尔滨师范大学计算机科学与信息工程学院 哈尔滨 150025)¹

(黑龙江外国语学院信息科学系 哈尔滨 150025)²

摘 要 精确识别细胞的前提是细胞采样。利用图像矩的旋转不变性和平移不变性,提出利用图像矩的概念对细胞进行采样处理,进而提取细胞的相关数学参数。首先利用动态阈值的方法分割灰度化后的细胞图像;然后利用图像细化算法对粘连细胞进行分割处理;最后利用图像矩提取样本细胞。实验证明该方法有较好的稳定性和鲁棒性。

关键词 图像矩,粘连,分割,采样

中图分类号 TP391.4 文献标识码 A

Research of Wood Cell Sampling Method Based on Image Moments

WANG Jian-hua¹ GAO Wei-wei² ZHAO Lei²

(College Computer Science and Information Engineering, Harbin Normal University, Harbin 150025, China)¹

(The Department of Information Science, Heilongjiang International University, Harbin 150025, China)²

Abstract Accurate cell identification is based on accurate sampling of cell images. This paper put forward that using the concept of image moments samples cell image and extracts mathematical features of wood cells based on rotational and translation invariance. We first grayed colored images and splited cell images with dynamic threshold segmentation method preliminarily, then implemented segmentation process for adherent cells with adherent cell segmentation method based on image thinning, finally extracted sample cells with the concept of image moments and completed the sampling work. The experiment proves that this algorithm has a good stability and robustness.

Keywords Image moments, Adherent, Segmentation, Sampling

1 引言

准确识别板材材种是现代工业具有实用价值的研究课题,基于木材细胞特征参数的材种识别方法的关键是获取高质量的木材细胞图像^[1],并且在分割木材细胞图像后,提取例如细胞横向排列的数量、细胞壁的纤维分布、管胞长度、管胞直径、管胞长度比、年轮密度、类圆度等等相关的数学参数,而准确提取细胞参数的前提条件是木材细胞图像取样。本文利用图像矩的概念对木材细胞图像中的对象集进行取样。M. K. Hu 于 1962 年提出了基于几何矩的图像不变矩。M. R. Teague 首先用 Zernike 矩来描述图像。Y. L. Sheng 等在 1994 年提出正交 Fourier-Mellin 矩。Z. L. Ping 等提出了切比雪夫-傅里叶矩、圆谐傅里叶矩和雅可比傅里叶矩等正交图像矩。除了 M. K. Hu 提出的几何矩外,这些矩都具有旋转不变性,并且都是正交矩,都可以用来重建图像^[2,3]。但由于关于原点旋转的表示不变量的多项式都应当符合,因此有旋转不变性的矩都是在极坐标系下定义的,而在计算机上的图像大都是以直角坐标系作为参照来进行显示存储的,这就给计算具有旋转不变性的图像矩带来了麻烦,即在求图像矩时,常常不得不先把图像中的每一个像素的坐标转换到极坐标系

下,然后再求图像的矩。在转换的过程中,不仅会使算法变得复杂,极大地增加计算量,而且当像素的坐标从直角坐标系向极坐标系进行转换时需要量化而导致舍入误差,从而使得到的图像矩不准确,在利用求得的图像矩进行图像识别时就会导致错误率大幅上升,而在利用图像矩重建图像时也会导致重建图像的质量下降^[4]。针对这个问题,本文利用圆谐傅里叶矩对分割好的板材细胞图像进行取样处理,经过取样处理后,细胞数字参数提取的结果更准确^[5]。

2 圆谐傅里叶矩

圆谐傅里叶矩的定义,如

$$\Phi_{mn} = \int_0^{2\pi} \int_0^1 f(r, \theta) T_n(r) \exp(-jm\theta) r dr d\theta \quad (1)$$

式中

$$T_n(r) = \begin{cases} \frac{1}{\sqrt{r^2}} n \\ \sqrt{\frac{2}{r}} \cos(n\pi r) \\ \sqrt{\frac{2}{r}} \cos(n\pi r) \end{cases} \quad (2)$$

到稿日期:2013-07-03 返修日期:2013-11-11 本文受国家自然科学基金:面向对象的高分辨率影像水土保持措施检测项目(41071262)资助。

王建华(1956—),女,硕士,教授,CCF 高级会员,主要研究方向为人工智能应用,E-mail: wjh@vip.163.com;高巍巍(1976—),硕士,副教授,主要研究方向为人工智能应用;赵磊(1980—),男,博士,讲师,主要研究方向为模式识别与智能控制。

坐标系中,图像函数 $f(r, \theta)$ 可以分解为

$$f(r, \theta) = \sum_{n=0}^{\infty} \sum_{m=-\infty}^{+\infty} \Phi_{nm} T_n(r) \exp(jm\theta) \quad (3)$$

对训练集中中板材细胞图像求圆谱傅里叶矩时应先将样本图像归一化到单位圆中^[6],式(1)中的积分区域如图1所示。

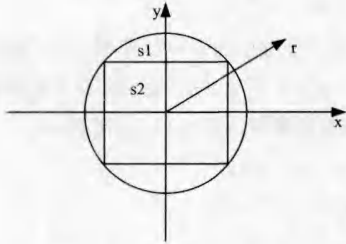


图1 圆谱傅里叶矩的积分区域

其中单位圆 s_1 是式(1)中的积分区域,矩形 s_2 区域是归一化后的图像。则由式(1)可得

$$\begin{aligned} \Phi_{nm} &= \int_0^{2\pi} \int_0^1 f(r, \theta) T_n(r) \exp(-jm\theta) r dr d\theta \\ &= \iint_{s_2} f(r, \theta) T_n(r) \exp(-jm\theta) ds + \int_{s_1} \int_{s_2} f(r, \theta) T_n(r) \exp(-jm\theta) ds \end{aligned} \quad (4)$$

由图2可知, $s_1 - s_2$ 的区域不是图像所在的区域,其中没有图像的像素值,这一部分区域对矩的贡献为零,则图像的圆谱傅里叶矩可写为

$$\Phi_{nm} = \iint_{s_2} f(r, \theta) T_n(r) \exp(-jm\theta) ds \quad (5)$$

式(5)还可以改为

$$\Phi_{nm} = \iint_{s_2} f(r, \theta) T_n(r) \exp(-jm\theta) dx dy \quad (6)$$

式(6)即为在直角坐标系下图像的圆谱傅里叶矩的计算公式。

在计算图像的圆谱傅里叶矩时,利用式(6),就不必再将图像每个像素的坐标从直角坐标系下转换到极坐标系下,这不仅降低了计算量,简化了算法,而且避免了对像素坐标的四舍五入的近似计算,从而彻底避免了转换坐标时带来的误差。

这个结论可以推广到所有其他的具有旋转不变性的正交图像矩(如正交 Fourier-Melin 矩^[7,8]、切比雪夫-傅里叶矩、圆谱傅里叶矩和雅可比-傅里叶矩等)^[9]中,这是因为具有旋转不变性的正交图像矩的公式都具有如下的形式

$$\varphi_{nm} = \int_0^{2\pi} \int_0^1 f(r, \theta) J_n(r) \exp(-jm\theta) r dr d\theta \quad (7)$$

式中, $J_n(r)$ 为径向函数, $f(r, \theta)$ 为图像函数。比较式(7)与式(1),可以很容易得到具有旋转不变性正交图像矩在直角坐标系下的计算公式:

$$\varphi_{nm} = \iint_s f(x, y) J_n(r) \exp(-jm\theta) dx dy \quad (8)$$

式中, s 为图像所在的区域。

3 细胞图像预处理

本文实验主要以以下4种树种切片作为处理样本,并根据

实验需要分别截取实验样本中的图片作为处理对象。如图2所示。

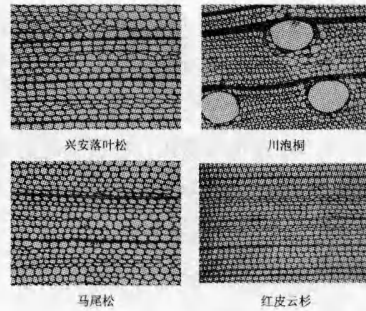


图2 实验样本图

首先对切片进行灰度化处理,然后对灰度化后的图片进行分割。木材细胞实体检测图像受多种因素的影响,其灰度分布不均匀,有些目标区域的灰度值甚至低于背景区域,针对这种情况,根据细胞图像的质量特点,本文采用动态阈值法分割木材细胞实体图像,灰度图像分割后的效果如表1所列。

表1 处理效果图

样本名称	样本截图	灰度图片	动态阈值分割	反色后效果图
兴安落叶松				
川泡桐				
马尾松				
红皮云杉				

(1)将图像分割成 10×10 块子图像。

(2)对子图像中所有像素进行迭代运算,将其作为本块子图像的阈值。

$$T_{k+1} = 0.5 \times \left\{ \frac{\sum_{i=0}^{T_k} h_i \times i}{\sum_{i=0}^{T_k} h_i} + \frac{\sum_{i=T_k+1}^{L-1} h_i \times i}{\sum_{i=T_k+1}^{L-1} h_i} \right\} \quad (9)$$

(3)利用邻近区域子图像阈值平均的方法,计算相邻3块子图像阈值的平均值,根据这个平均值,对各子图像进行阈值分割。

$$g_i(x, y) = \frac{f_{i-1}(x, y) + f_i(x, y) + f_{i+1}(x, y)}{3} \quad (10)$$

动态阈值分割细胞图像的方法具有快速、准确性高等优点,但是有时切片图像的质量影响也会造成分割不彻底的问题,也就是细胞存在粘连现象,因此对粘连细胞的正确分割又是下一步解决的重点问题。本文采取基于图像细化的木材粘连细胞分割方法对个别出现粘连的细胞图像进行分割。

(1)首先利用类圆度的思想确定粘连细胞。细胞图像的形态特征参数有:面积 S (连通区域内像素个数之和)和周长 P (连通区域轮廓上像素个数之和),用形状因子 $F(X)$ 描述图

像 X 与圆形的偏离程度, 定义为:

$$F(X) = \sqrt{\|P\|^2 / 4\pi S} \quad (11)$$

其中, 设 X 是标准的圆形, 半径为 r , 则周长 $P = 2\pi r$, 面积 $S = \pi r^2$, 于是 $F(X) = 1$ 。在一般情况下, X 与圆的偏离程度越大, 则 $F(X)$ 越大。

(2) 对倾斜细胞图像进行校正。根据细胞的倾斜模式, 对水平倾斜的粘连细胞进行双线性插值旋转校正, 对垂直倾斜的细胞进行双线性插值错位偏移校正^[5]。

(3) 提取细胞图像的骨架, 并进行模板遍历。

4 样本细胞的特征提取

特征提取就是要找到一组尽可能小的能代替图像的数据集合。对图像的识别实际上是对代表图像的数据集中数据的判别, 所谓对图像特征的提取, 就是将细胞的特征转化为数据集中的数据, 代表一幅图像的数据集中的所有数据就构成了一个特征向量, 每个特征向量对应于空间中的一个点, 而同类细胞的特征向量应该位于空间中的相同区域。所以对图像的识别实际上是对向量归属哪一个空间区域的划分问题^[6]。

在选择图像的特征时有两个基本的要求就是准确性和鲁棒性, 而且要具有缩放、平移、旋转等多畸变不变性。因为在很多情况下, 图片不仅有噪声的干扰, 而且相同细胞的图片可能具有不同的尺寸、位置和旋转方向, 这就要求所选择的特征量不能随着图像的平移、旋转、缩放而发生改变。

在对木材样本细胞识别中, 采用圆谐傅里叶矩作为图像的特征, 这是因为圆谐傅里叶矩在各种图像矩中对图像的描述性能最好, 计算速度最快。圆谐傅里叶自身不是畸变不变量, 但是归一化之后, 可获得平移、灰度、尺度、旋转不变性。归一化方法如下:

在计算图像矩之前首先计算细胞图像的重心, 对于二维的 MXM 图像 $f(x, y)$ 来说, 图像的重心可由公式得到:

$$\bar{x} = \frac{m_{10}}{m_{00}} \quad (12)$$

$$\bar{y} = \frac{m_{01}}{m_{00}} \quad (13)$$

其中

$$m_{pq} = \sum_{x=0}^{M-1} \sum_{y=0}^{M-1} x^p y^q f(x, y) \quad (14)$$

然后进行坐标变换, 将坐标原点平移到图像的重心。则在新坐标系中采用这种方法计算出的圆谐傅里叶矩具有平移不变性。

由于圆谐傅里叶矩的角向函数为 $e^{jm\varphi}$, 将图像旋转角度 φ 后, 所有矩 E_{jm}^i 都增加相同的相位因子 $e^{jm\varphi}$, 圆谐傅里叶矩的模 $|E_{jm}^i|$ 是旋转不变的, 因此在新坐标系中计算的圆谐傅里叶矩具有旋转不变性。

相对于旋转和平移不变性, 我们采用下述方法来获得缩放不变性:

将要识别的细胞图像放入到一个重心边缘圆中, 这个圆的圆心就是图像的重心, 也就是新坐标系的原点, 圆的半径就

是离原点最远的图像像素到原点的距离。然后将这个重心边缘圆归一化为单位圆。在这个单位圆中计算出的图像矩就有了缩放不变性。这种方法中需要找到重心边缘圆的半径, 具体步骤如下:

(1) 计算出单个细胞的边界和重心 (x_c, y_c) ;

(2) 设 $r_{\max}^2 = 0$;

(3) 对于单个细胞中的每一行像素, 执行如下步骤:

找到这一行中最左边和最右边的不为零的像素: x_1, x_2 ,

如果该行中所有的像素值都为零, 跳到下一行。

$$x_i' = |x_i - x_c|, i = 1, 2 \quad (15)$$

$$x_0 = \max[x_1', x_2'] + 0.5 \quad (16)$$

$$y_0 = |y - y_c| + 0.5 \quad (17)$$

$$r^2 = x_0^2 + y_0^2 \quad (18)$$









如果 $r^2 > r_{\max}^2$, 则 $r^2 = r_{\max}^2$, 继续计算下一行。

(4) 将求得的每个细胞的 r_{\max} 求平均值 R , 用 R 作为重心边界圆的半径并将 R 作为缩放因子。这时求得的图像矩就具有了缩放不变性。

5 实验结果

利用上述算法分别对兴安落叶松、川泡桐、马尾松、红皮云杉进行样本细胞提取。根据细胞图像的闭合性和类圆度, 首先剔除由于截图样本细胞造成的残缺细胞。实验结果如表 2 所列。

表 2 图像矩处理结果分析

样本名称	处理前	处理后	处理前细胞数量	处理后细胞数量	样本提取有效率
兴安落叶松			176	135	130%
川泡桐			208	116	179%
马尾松			69	52	133%
红皮云杉			125	88	142%

在实验中, 对彩色图像进行灰度化, 然后利用动态阈值分割方法把细胞图像分割成单个的细胞, 由于拍摄图像质量的影响, 个别细胞出现粘连现象, 利用基于数学形态的方法分割粘连细胞, 并对图像使用圆谐傅里叶矩作为图像特征的样本采集方法。实验结果表明, 该算法对样本细胞图像的提取有较好的处理效果。

结束语 本文使用了一种新的以圆谐傅里叶矩为图像特征的新的图像样本提取算法来分割木材细胞样本图像。在这个算法中, 我们采用动态阈值分割方法对灰度化后的木材细胞图像进行分割, 并利用基于图像细化的木材粘连细胞分割方法, 对分割后出现粘连的细胞进一步处理, 以得到较好质量

的细胞图像。在这个样本提取的算法中,由于利用图像矩作为样本图像的特征,避免了对不同的细胞要采用不同的方法来测量图像特征的问题,因此,本文所介绍的算法可以适用于更多需要采样的木材细胞图像,可以推广到其他细胞图像的处理中。本算法由于只能对二值化后的细胞图像进行处理,因此一定程度上对细胞图像有错误处理的情况发生,故将本算法应用到彩色图像中将是下一步工作研究的重点。

参考文献

[1] 关涛,周东翔,刘云辉,等.基于自适应阈值分割的宫颈细胞图像分类算法[J].信号处理,2012(09):1262-1270
 [2] 王晓华,王绍虎,余增亮.细胞图像的采集区域划分及细胞识别[J].计算机工程,2003(20):35-37
 [3] 张燕红,李瑛,张燕宁,等.基于数学形态学的细胞图像分割方法

研究及实现[J].计算机与现代化,2013(07):135-137
 [4] 刘艳丽,孟朝晖.层次聚类在细胞图像分析中的应用[J].计算机应用与软件,2013(05):287-290
 [5] 杨建菊,张贵英.基于自适应滤波的淋巴细胞图像分割算法研究[J].计算机仿真,2012(08):257-260
 [6] Colantonio S. Automatic fuzzy-neural based segmentation of microscopic cell images [J]. International Journal of signal and Imaging systems Engineering, 2008, 1(1): 18-24
 [7] 梁肖,胡贞,吕晓玲,等.基于自适应阈值的活体细胞图像分割改进方法[J].长春理工大学学报:自然科学版,2013(z2):138-140
 [8] 张鑫,陈伟斌.基于形态学重构的多结构元细胞图像边缘检测[J].计算机仿真,2009(08):216-219
 [9] 汪婧,曹益平,程旭升.一种基于H直方图变换的白细胞图像分割方法[J].光学与光电技术,2013(02):74-78

(上接第265页)

右特征向量矩阵,新序列所含的信息更多,用于异常检测变量的参数个数减少,异常检测结果会更精确。证明了该方法的有效性和准确性。

多变量时间序列长度值和 k 的取值可能会影响算法的效率。当 $k=5$ 时,数据样本的数量从 100 增加到 300,执行算法消耗的 CPU 时间如图 5 所示。从图 5 可以看出,消耗的时间与样本个数 N 呈近似二次方关系;当固定 $N=100$ 时,让相邻值 k 从 3 增加到 13,花在算法的 CPU 时间如图 6 所示,可以看到当 k 增加时算法消耗时间逐渐增加。

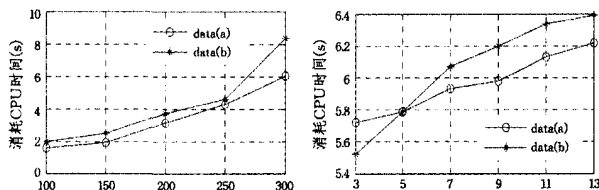


图5 序列个数对算法的影响 图6 k 近邻个数对算法的影响

结束语 相似性度量方法是提高多变量时间序列异常检测算法性能的关键。主成分加权 Euclid 范数距离度量算法与 k -近邻局部异常检测算法相结合,能够有效去除异常检测中的多变量之间相关性的影响,减少参与异常检测变量的个数,提高异常检测的准确性,并且对于时间序列的长度没有特殊要求,不仅适用于多变量时间序列,如果将 $D_{Ews}(A, B, w)$ 距离替换为欧氏距离或动态时间弯曲距离,还可以计算单变量时间序列之间的相似性,用于单变量时间序列异常检测。针对数据的异常检测实验验证了异常检测的有效性和合理性。

参考文献

[1] Tsai C Y, Chuang C-C. Developing a feature weight self-adjustment mechanism for a K-means clustering algorithm[J]. Com-

putational Statistics & Data Analysis, 2008, 52(10): 4658-4672
 [2] Kumpulainen P, Kimmo. Local anomaly detection for mobile network monitoring[J]. Information Sciences, 2008, 178(20): 3840-3859
 [3] Jiang F, Yuan J, Tsafaris S, et al. Anomalous video event detection using spatiotemporal context[J]. Computer Vision and Image Understanding, 2011, 115(3): 323-333
 [4] 肖辉. 时间序列的相似性查询与异常检测[D]. 上海: 复旦大学, 2005: 33-40
 [5] Cong Y, Yuan J, Liu J. Sparse reconstruction cost for abnormal event detection[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2011: 3449-3456
 [6] Baragona R, Battaglia F. Outlier detection in multivariate time series by Independent Component Analysis [J]. Neural Computation, 2007, 19(1): 1962-1984
 [7] Yamanli K, Takeuchi J. A unifying framework to detecting outliers and change-points from nonstationary data [C]// Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD2002). New York: ACM Press, 2002: 676-681
 [8] 林果园, 郭山清. 基于动态行为和特征模式的异常检测模型[J]. 计算机学报, 2006, 29(9): 1553-1559
 [9] 张力生, 杨美洁, 雷大江. 时间序列重要点分割的异常子序列检测[J]. 计算机科学, 2012, 39(5): 183-186
 [10] 郭小芳, 张锋丽. 基于加权范数的多维时间序列相似性主成分分析[J]. 江苏科技大学学报: 自然科学版, 2011, 25(5): 466-469
 [11] Agyemang M, Ezeife C I. LSC-Mine: Algorithm for Mining Local Outliers[C]// Proceedings of the 15th Information Resource Management Association (IRMA) International Conference. New Orleans, 2004, 1: 5-8
 [12] Yang K, Shahabi C. An efficient k nearest neighbor search for multivariate time series [J]. Information Computation, 2007, 205(1): 65-98