

文章编号: 1005—0523(2006)05—0066—04

显微图像特征量的获取及分析

蒋先刚, 艾剑锋

(华东交通大学 基础科学学院, 江西 南昌 330013)

摘要: 研究显微图像中各连通域的几何特征计算和统计分析技术, 主要通过对细胞面积、周长、外接矩形及不变矩等的计算形成对细胞组织综合全面的描述和分析, 并在实际系统开发中得到验证。

关键词: 矩形度; 矩; 边缘复杂度; 均值聚类。

中图分类号: TP368.15 **文献标识码:** A

0 引言

医学图像分析与统计软件涉及到各特征量的计算和这些特征量的综合分析。组织细胞分类、红血球计数、病变细胞的判别在解决了细胞粘连的前提下, 其他都是对单个细胞的各个特征量的计算。单个细胞的几何特征就是各连通区域的几何特征, 各特征量之间存在着密切的联系和影响, 样片中各细胞的分布和统计就是图像中各连通区域的分布和统计。细胞各特征量的计算是统计与分析的基础, 而正确的判断和统计还需经过数理统计、均值聚类等进一步计算和分析而获得。

自主开发的医学图像分析与统计软件系统的软件处理模块包括显微摄像头的图像采集, 阈值优化选择(灰度图)或 HSL 选取掩膜处理(彩色图), 二值化处理, 孔洞填充, 腐蚀去噪, 细化和粘连细胞的分割, 各连通域的计算包括对几何特征如周长、面积、重心的运算和形态特征如圆度、矩形度、矩的应用、边缘复杂度的运算。

1 各特征值的定义与计算

1.1 连通区域面积、周长和重心的计算

形态特征的计算首先涉及到细胞连通区域的

标记, 区域标记是指将连在一起的同一区域像素赋上相同的标记的过程。不同细胞的连接部分赋上不同的标记。为了能更好地区分不同的细胞体, 本文采用颜色差值区域标记法, 给不同的连通体赋上不同的像素值, 而且其颜色差值比较大, 这样我们就可以很清楚地区分每个独立的细胞区域。

连通区域的面积通过连通区域的总像素的求和得到, 而连通区域的周长通过获得连通域的外轮廓, 然后求取累积边界的长而获得, 对于边界对角线上的点, 其长度应计为 $\sqrt{2}$ 。重心的行坐标和列坐标分别按下列公式计算:

$$\bar{x} = \frac{\sum xf(x, y)}{\sum f(x, y)}, \bar{y} = \frac{\sum yf(x, y)}{\sum f(x, y)} \quad (1)$$

1.2 连通区域圆度的计算

对于面积一定的图形, 一般周长越小, 圆形外轮廓表面越光滑, 越接近圆; 反之, 周长越大, 则图形外轮廓表面褶皱越多, 形状也就越复杂。由此可采用圆度 C 来衡量图形的复杂程度, 其最常用的计算公式如下:

$$c = \frac{P^2}{4\pi A} \quad (2)$$

其中, P 为图形周长, A 为周长所围面积。

另外, 圆度还有两个计算公式:

$$c = \frac{A}{A'} \quad (3)$$

其中, A 为区域面积, A' 为区域最小外接圆面

收稿日期: 2006—03—16

作者简介: 蒋先刚(1958—), 男, 湖南永州人, 教授。主要研究领域为计算机图形学、机械、电气 CAD/CAM。

积,该指标同样可认为圆形区域最紧凑时,其紧凑度为1.在计算中采用最小外接圆面积作为衡量区域形状的标准.

$$c=\frac{1.273A}{L^2}$$

(4)

其中, A 为区域面积, L 为区域最长轴长度.该指标也认为圆形为标准形状,但它只考虑了最长轴的长度,它只能概略地反映连通区域的形状.实验表明,连通区域越是接近圆,其圆度值越接近于1.

1.3 连通区域矩形度的计算

当单个连通区域给定后,即可求出其最小外接矩形,而反映物体矩形度的一个参数是矩形拟合因子 R ,即

$$R=A_0/A_r$$

(5)

其中, A_0 是物体的面积, A_r 是物体最小外接矩形的面积.

这里, R 反映了一个物体对其最小外接矩形的充满程度,对于真正的矩形, R 取得最大值1;对于纤细的、弯曲的物体则其值变小.矩形拟合因子的值限定在0与1之间.包含外接矩形的面积、长宽比以及矩形拟合因子的对比分析如图1所示.

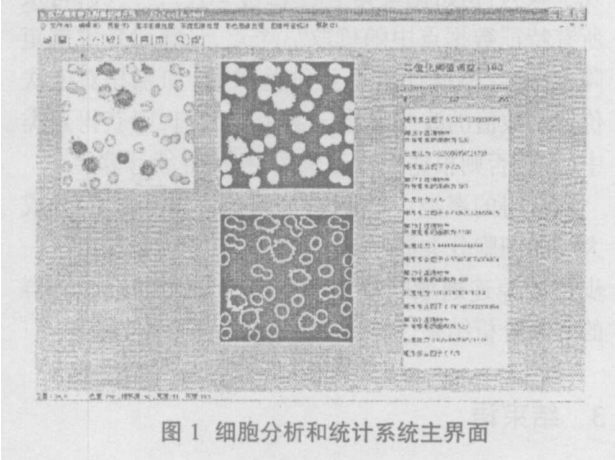


图1 细胞分析和统计系统主界面

1.4 连通区域的矩不变量的计算

矩不变量是指物体图像在经过平移、旋转、比例等变换后仍然不变的矩特征量.

设图像函数 $f(x,y)$, D 为目标区域,则目标的 $p+q$ 阶原点矩 m_{pq} 定义为:

$$m_{pq}=\int_D \int_D x^p y^q f(x,y) dx dy$$

(6)

其中 $p,q=0,1,2,\dots$, 目标的 $p+q$ 阶中心矩定义为:

$$\mu_{pq}=\int_D \int_D (x-\bar{x})^p (y-\bar{y})^q f(x,y) dx dy$$

其中 $(\bar{x},\bar{y})$ 为目标区域灰度重心,离散化后公式为:

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

为了描述物体内部灰度分布特征,在 $f(x,y)$ 分片连续且在有限区域内有非零值时,则其各阶细化次数与最小连通皆存在,且 $f(x,y)$ 唯一地确定一个矩序列 $\{m_{pq}\}$,反之,矩序列 $\{m_{pq}\}$ 也唯一地确定 $f(x,y)$.

二阶中心矩表征了图像的方向特性,设图像以具有最小和最大二阶矩的轴为主轴,则主轴的方向可表示为:

$$\varphi=(1/2)\tan^{-1}[2\mu_{11}/(\mu_{20}-\mu_{02})]$$

(7)

对细长类的图像可以用椭圆模型表示其尺寸和方向特性,利用图像的3个二阶中心矩 μ_{20} , μ_{11} 和 μ_{02} 构建图像的椭圆模型.

设椭圆长半轴为 a ,短半轴为 b ,则有:

$$a=\left[\frac{u_{20}+\mu_{02}+[(\mu_{20}-\mu_{02})^2+4\mu_{11}^2]^{1/2}}{\mu_{00}/2}\right]^{1/2}$$

(8)

$$b=\left[\frac{u_{20}+\mu_{02}+[(\mu_{20}-\mu_{02})^2+4\mu_{11}^2]^{1/2}}{\mu_{00}/2}\right]^{1/2}$$

(9)

由(8)、(9)和(10)式可见, φ 为椭圆长半轴或短半轴与坐标系 X 轴的倾角.

至此,通过矩的计算,我们可以获得椭圆模型的长半轴与短半轴,这将为图像特征的进一步计算提供更有效的计算参数.

1.5 连通区域边缘复杂度的计算

所谓边缘复杂度,意指图像不规则边缘的长度同与其边缘拟合的规则图像周长的比值.在所研究的椭圆模型中,我们通过原图像的不变矩求其长、短半轴,利用这两个值的均值近似代替椭圆模型的拟合图像的半径.则可得边缘复杂度的计算公式为:

$$c=\frac{L}{L'}$$

(10)

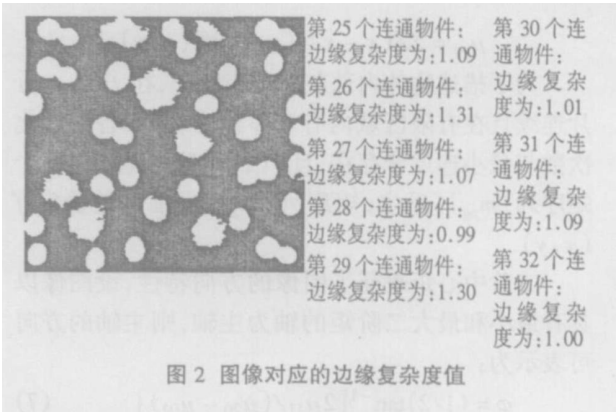
其中 L 为不规则图像的实际周长, L' 为其拟合图像的周长,而在系统中我们采用的计算公式为:

$$c=\frac{0.8 * L}{2 * \pi * (a+b) / 2}$$

(11)

这样便可以通过对边缘复杂度的计算,判断出那些病变的不规则细胞.

从图2和测量结果中,我们可以明显分辨出病变细胞和它们所对应的复杂度的值.边缘越是规则的图形,其复杂度值越接近于1,相反,其值则越远离1.这样我们就可以通过计算目标图像的复杂度来判断出那些不规则的图形.



2 特征值的综合分析与统计

考虑到排除图像噪音的需要,在对细胞的几何与形态特征量的运算中,通过对细胞特征中周长、面积、圆度、长短半轴等量的数学期望以及它们的方差的计算,可以排除掉过小的噪声点而使有效的连通域进入进一步的计算.统计量期望的计算公式为:

$$E=\frac{1}{n}\sum x_n \tag{12}$$

其中, n 为细胞图像中细胞的个数, x_n 为第 n 个细胞的特征量(如:周长、面积等).

统计量方差的计算公式为:

$$S^2=\frac{1}{n}\sum (x_n-E)^2 \tag{13}$$

其中, n 为细胞图像中细胞的个数, x_n 为第 n 个细胞的特征量(如:周长、面积等). E 为给特征量的期望.在一定方差范围要求内被视为有效的连通区域.

为了对已计算的连通域进行合理的归类,如需统计红细胞和白细胞的个数,正常细胞和非正常细胞的数量等,应继续采用 K 均值聚类算法获得. K 均值聚类核心思想是把 n 个向量 $X_j(j=1,2,\dots,n)$ 分为个组 $G_i(I=1,2,\dots,c)$, 并求每组的聚类中心,使得非相似性(或距离)指标的价值函数(或目标函数)达到最小.当选择欧几里德距离为组 j 中向量 X_k 与相应聚类中心 G_i 间的非相似性指标时,价值函数可定义为:

$$J=\sum_{i=1}^c J_i=\sum_{i=1}^c (\sum_{k, x_k \in G_i} \|x_k - c_i\|^2) \tag{14}$$

如用一个通用距离函数 $d(x_k, c_i)$ 代替组 i 中的向量 x_k , 则相应的总价值函数可表示为:

$$J=\sum_{i=1}^c J_i=\sum_{i=1}^c (\sum_{k, x_k \in G_i} d(x_k, c_i)) \tag{15}$$

以欧几里德距离作为向量的非相似性指标,距离计算用细胞特征中周长、面积、圆度、长短半轴、边缘复杂度等量值与 c 个聚类中心的值分别做计算,距离的计算式为 $Distance[i]=\sqrt{\sum (P_x-c_i)^2}$. 分群组数 i 的取值要大于需要得到的识别细胞的类型数以进一步排除杂信号,而距离计算中的维数为参加判别分析的细胞特征的种数.

如果 c_i 是 x_j 的最近的聚类中心,那么 x_j 属于组 i . 由于一个给定数据只能属于一个组,所以隶属矩阵 U 具有如下性质:

$$\sum_{i=1}^c u_{ij}=1, \forall j=1,\dots,n \tag{16}$$

且

$$\sum_{i=1}^c \sum_{j=1}^n u_{ij}=n \tag{17}$$

另一方面,如果固定 U_{ij} , 则使最小的最佳聚类中心就是组 i 中所有向量的均值:

$$c_i=\frac{1}{|G_i|} \sum_{k, x_k \in G_i} x_k \tag{18}$$

对于数据集 $X_i(i=1,2,\dots,n)$, 一般 K 均值聚类算法用来确定聚类中心 G_i 和隶属矩阵 U 的实现步骤为:初始化聚类中心 $G_i(I=1,2,\dots,c)$, 可从所有特征数据点中随机取 c 个点,计算确定隶属矩阵 U , 然后根据计算得到价值函数.如果相对上次价值函数值的改变量小于某个控制值,则获得聚类中心值.否则根据式修正聚类中心,继续计算.

获得的聚类中心就是等到的不同特性的分成大类的细胞分类数.均值聚类算法可从宏观上进一步过滤掉连通域噪声并给出更为全面的细胞分群的判断分析结果.

3 结束语

根据连通域特征的定义和数学模型,各特征量的单独计算很容易得到,为了全面分析它们,需综合分析这些特征量.细胞大小、细胞的圆度、细胞的长宽、细胞的矩等在不同目标判别中,其占有的权重不同,系统而全面的分析还可通过各特征的期望与方差,可区分或筛选出图像中有效的连通区域.同时,用 K -均值聚类和 FCM 算法可从宏观上进一步过滤掉连通域噪声并给出更为全面准确的细胞分类和判断分析结果,得到精确的分析统计结果.

参考文献:

[1] David F. Rogers. Procedural Elements for Computer Graphics

[M]. 北京:机械工业出版社,2002.

国外医学—生物医学工程分册,1994,17(3):125~136.

[2] 蒋先刚. 基于 Delphi 的数字图像处理工程软件设计[M]. 北京:中国水利水电出版社,2006

[4] 马 华,王 清,张 永. 基于改进FCM 的医学图像分割[J]. 微计算机信息. 2006, (7):249~250.

[3] 郑 玲,等. 数学形态学及其在生物医学中的应用[J].

The Gaining & Analysis of Micro-Cell Image Properties

JIANG Xian-gang, AI Jian-feng

(School of Basic Science, East China Jiaotong University, Nanchang 330013, China)

Abstract:The paper studies the micro-cell image connected region 's geometrical characteristics calculation and statistical analysis technologies. It mainly deals with the calculation of integrated description and analysis. It also validates the methodology by coding of a medical image analysis system.

Key words:rectangular degree;moment;fringe complexity;k-means

(上接第 26 页)

3)高低和轨向谱中含有较为明显的与 25m 轨长相关的复杂周期波的尖峰谱线,应线加强钢轨焊接接头的质量控制,否则列车运行中在这些结构处不可避免地将出现周期振动.

参考文献:

[1] 铁科院. 我国干线轨道不平顺功率谱的研究系列报告[R]. 北京. 1999.

[2] 陈宪麦,杨凤春,吴旺青,柴雪松. 秦沈客运专线轨道谱的研究[J]. 铁道学报. 2006, (8):97~100.

[3] 杨文忠,练松良,刘扬. 轨道不平顺功率谱拟合分析方法[J]. 同济大学学报, 2006, (3):83~87.

The Comprehensive Study of Nanjing Changjiang River Bridge's Change Rule in Track Geometry Status

YAO Xiang-jing, XU Yu-de

(Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 200092, China)

Abstract:The Nanjing Changjiang River bridge which is located in Beijing-Shanghai line is the key route way that crosses the river from south to north. Owing to the differences in its rail structure, the change of track geometry is different. This paper contrasts the means of amplitude value with the means of power spectrum, and studies the bridge 's change rule in track geometry status synthetically, it can provide tech support for establishing repair and maintain standard reasonably.

Key words:track geometry irregularities;power spectrum;change rule