

# 一种基于PC机的肺癌细胞自动识别分类系统

陆新泉 陈世福 (南京大学计算机软件国家重点实验室 210093)

叶玉坤 (中国人民解放军八一医院 210002)

**摘要:**本文介绍了一种基于PC机的肺癌细胞自动识别分类系统的设计思想和处理流程,给出了形态特征提取及细胞分类的算法,并给出了专家系统技术在系统中的作用和处理方法。

**关键词:**图象处理 形态特征提取 细胞识别 专家系统

## 一、引言

近几年来,一些发达国家已将图象处理、模式识别及科学计算可视化技术应用于医疗领域,取得了一些可喜的成果,计算机断层扫描(CT)及核磁共振(MRT)等技术已普遍使用于医疗领域,对医生发现病灶,作出诊断有极大的帮助。最近二年来已有商品化的医学图象处理软件用于脑部结构的显示等方面。但医学图象处理软件用于肺癌早期诊断系统还未见报道。而目前全世界肺癌发病率正在逐年上升,国内外的资料表明,肺癌病例一经确诊80%已属晚期,即失去手术治疗的机会。因此及早的发现癌变,以达到准确的早日诊断和治疗已成为迫切解决

的问题。

本文介绍了一种基于PC机的肺癌细胞自动识别分类系统。该系统对肺癌穿刺标本进行采集,利用数学形态学(Mathematical Morphology)的方法进行滤波、平滑及特征提取,找出癌变信息,对癌细胞实现自动分类,根据专家知识完成判决、分类统计等任务,以达到快捷、准确和及早断、治疗的目的。

## 二、系统结构与功能特点

系统结构如图1,它由图象采集及预处理模块、形态处理模块及识别分类等模块组成,该系统的功能特点是:

(1)利用 CCD 及图象板对肺癌穿刺标本进行采集, 从而得到待处理的图象, 图象格式采用 BMP 的格式。

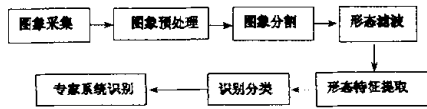


图1 系统结构图

(2)对采集到的图象进行滤波及 RGB 到 GRAY 的转换等处理。在 RGB 转换成 GRAY 时, 根据 NTSC 视频标准按下列公式进行计算:

$$\text{gray} = 0.3 \times \text{red} + 0.59 \times \text{green} + 0.11 \times \text{blue}$$

(3)根据显微细胞图象的特点, 用阈值分割的方法对细胞图象进行分割。

(4)用数学形态学的方法进行形态滤波和形态特征提取。

(5)根据所提取的特征及肺癌细胞病理学的专家知识对肺癌细胞进行识别分类。

### 三、系统的关键技术及算法

#### 1. 图象分割

图象分割模块的主要功能是将肺癌细胞逐一从背景中分离出来, 而在通常情况下, 显微细胞图象目标比例小, 背景较复杂, 存在有干扰和噪声, 因而显微细胞图象的分割显得十分复杂和困难。图象分割的基本方法目前主要根据基于区域内部的特性具有某种均匀性的原则或者是基于区域之间的特性具有某种不连续性的原则进行分割, 目前常见的分割技术有: 阈值分割、聚类分割、区域生长、区域的分裂与合并和边缘检测与边界跟踪等, 其中阈值分割因其简单、有效以及占用内存小而受到推崇。以往的阈值分割方法是基于灰度直方图统计的搜索法: 谷值点搜索以及最优搜索, 该方法的优劣直接取决于灰度直方图的形状的优劣, 实验结果表明, 该方法的鲁棒性不强, 速度不快。根据实验表明边界处的灰度级变化信息与阈选择密切相关, 为此, 本文提出了基于图象灰度梯度的直方图和显微细胞图象的特点的双阈值自动分割方法。该方法根据显微细胞图象的特点, 给出约束, 抑制噪声和减弱杂散物的影响, 改善直方图的谷点和峰值, 用迭代方法求出类内距离意义下的最优阈值, 并对图象作阈值分割, 效果很好。

#### 2. 形态滤波

滤波是抑制噪声, 传统的方法是通过空域或时域变换来完成。由数学形态学的开闭运算组成的滤波算法是一种非线性滤波算法, 它的一个重要特点就是根据目标和噪声之间的形态或拓扑结构上的差异去伪存真, 既可以去掉孤立的噪声点, 又可以对细胞核实现平滑处理, 以便于下一步的形态特征提取。本文根据显微细胞图象的特点, 采用了开运算滤波法。所谓开运算, 就是对图象先腐蚀、后膨胀, 以消除散点和毛刺, 并起到对图象的平滑作用。

设 A 为原灰度肺癌细胞图象, B 为如图 2 所示的结构元素, 则开运算为:

$$A \circ B = (A \ominus B) \oplus B$$

其中: A ∘ B 为实施开运算后肺癌细胞图象, ⊖: 腐蚀运算; ⊕: 膨胀运算。

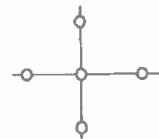


图2 结构元素 B

#### 3. 形态特征提取

形态特征提取实质上就是“图象图形化”的过程, 即计算反映分析对象的各种参数, 为下一步识别分类或判决作准备。由病理学的知识可知, 肺癌细胞与正常细胞相比具有如下特点: (1)细胞核大; (2)核浆比例倒置; (3)核更近似于圆或椭圆。根据这些特点, 本文提出了用 Freeman 链码(简称链码)对二值图象进行边缘跟踪的方法实现了如下特征的提取: ①核面积; ②圆形成度; ③形状奇异因子。根据不同的连通性定义, 链码有四连通链码和八连通链码之分。在边缘跟踪时, 本文采用了图 3 所示的八连通链码表示, 这样上述的形态特征可用下式求得:

$$\cdot \text{核面积: } S = \sum_{i=0}^n a_{y_0} (y_{i-1} + \frac{1}{2} a_2)$$

$$\cdot \text{圆形成度: } CG = \frac{4\pi S}{C^2}$$

$$\cdot \text{形状奇异因子: } F = \frac{2S^2}{\pi^2 L_{\max}^2 L_{\min}^2}$$

其中 n 为总的链码数,  $y_0$  为初始点的纵坐标, 且  $y_i =$

$y_{i-1} + a_{j2}$ ;  $j$  表示 07 八个方向, 且有:

$$\{a_{00}, a_{10}, \dots, a_{70}\} = \{1, 1, 0, -1, -1, -1, 0, 1\}$$

$$\{a_{02}, a_{12}, \dots, a_{72}\} = \{0, 1, 1, 1, 0, -1, -1, -1\}$$

$C$  为细胞核的周长, 并且  $C = n_e + n_o \sqrt{2}$

$n_e, n_o$  分别为链码中偶数与奇数方向的链码数;  $L_{\max}$

和  $L_{\min}$  分别表示细胞核在各方向上投影长度的最大值和最小值。

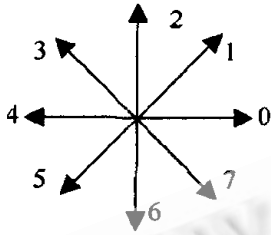


图3 八连通链码图

#### 4. 癌细胞识别分类

(1) 识别分类算法。根据上述所提取的形态特征, 利用下面的算法可以完成对细胞的粗分类:

For 每个细胞核

if ( $STH1 < S < STH2$ )

if ( $CG > CGTH$ ) and ( $F < FTH$ )  
then 该细胞为可疑的小细胞癌细胞

\* else if ( $S > STH2$ )  
then 该细胞为可疑小细胞癌细胞  
else 该细胞为正常细胞

其中  $S_{TH1}, S_{TH2}, CG_{TH}$  及  $F_{TH}$  分别是  $S, CG$  和  $F$  的阈值, 它们是在训练阶段由实验获得的。

经过上述粗分类后, 去除了大多数的正常细胞核, 而只保留少量的可疑癌细胞作为后继分类对象。值得一提的是, 由于小细胞癌具有聚堆的特点, 使得在分割过程中很难将聚堆的小细胞癌的细胞分割为单个细胞, 而只能将它们分割为一个“大细胞核”, 并且其面积往往大于其

他任何一种细胞核的面积。因此, 采用了另一面积上限阈值  $STH2$  来选择小细胞癌的细胞, 即当某一细胞核的面积  $S$  大于阈值  $STH2$  时, 则被视为可疑小细胞癌的细胞。

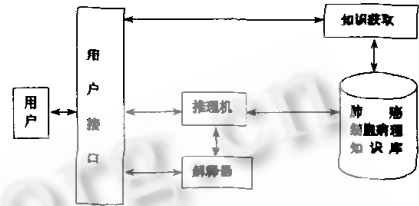


图4 肺癌细胞病理诊断专家系统

(2) 专家系统判别诊断。经过粗分类后, 得到了可疑的癌细胞。根据这些细胞的形态特征, 系统利用有关肺癌病理方面的专家知识, 借助于专家系统的推理机对细胞进行最终的识别分类。该系统的肺癌细胞病理诊断专家系统如图 4 所示, 其中知识库中的事实和规则是根据医学知识和病理学专家长期的临床诊断经验总结而成的, 推理机利用知识库中的知识对抽取的可疑细胞的形态特征进行推理, 判断出结果, 并由解释器给出相应的解释, 其中本部分各功能模块的具体实现技术因篇幅有限, 这里不再赘述。

#### 四、结束语

本系统在 586 高档微机的 Windows 平台上, 用 Visual C++ 编程实现。经实验表明, 系统的设计和采用的有关算法是成功的, 达到了预期的效果。进一步的研究工作正在进行中。

#### 参考文献

- [1] J. Serra. "Image analysis and Mathematical morphology", Academic Press, 1982.
- [2] S. R. Sternberg. Grayscale Morphology, CVGIP, Vol. 35, 3, 1986, pp. 333 - 355.
- [3] J. Serra. Introduction to mathematical morphology, CVGIP, Vol 35, 3, 1986, pp. 283 - 305.

(来稿时间: 1998 年 5 月)