

文章编号: 1000-2022(2003) 05-0712-05

基于图像区域特征的细胞识别方法及实现

程 勇, 傅德胜

(南京气象学院 计算机科学与技术系, 江苏 南京 210044)

摘 要: 先提取目标物体的面积、中心矩、圆形度、细长度、核浆比等区域特征。然后采用单原型模式表征方法, 利用集群分析技术中最小距离分类器, 构造判别函数, 并依此作为识别类型的依据。最后, 通过对鳞细胞的区域特征进行计算比较, 选择出能够较明显地辨别出类别的特征向量, 建立判别函数并对结果的可信度进行了讨论。

关键词: 细胞识别; 区域特征; 模式; 单原型

中图分类号: TP391 **文献标识码:** A

随着科学技术的发展, 影像技术已经被广泛应用在生物医学领域。以前, 对疾病的诊断仅仅通过表面现象, 即各种疾病的症状, 这不仅使得经验在诊断中起主导作用, 而且错判率始终占有一定比例; 而今, 通过对显微细胞图像的研究和分析, 即通过细胞形态, 来判断疾病的病因、研究医疗方案、并可以观测医疗疗效。因而对细胞的形态分析是现代医学诊断的重要环节。目前, 医学上对细胞的分析, 一般先是通过人工去辨识显微细胞图像, 而后做出判决, 所以医学专家们不断重复地做同一件事, 这不仅浪费了珍贵的人力资源, 而且延误了诊断时间, 耽误了对疾病的治疗, 给病人造成了更大的痛苦。因而, 本文提出了基于图像区域特征, 利用计算机技术对显微细胞图像进行自动识别和判断的思想。实验证明此方法可以准确、快速地对某些医学图像进行处理并做出相应的诊断结果。最后通过对鳞细胞的识别, 证明了此方法的有效性。

1 医学细胞图像特征

物体的特征包括形状、光密度(灰度)、颜色、结构、纹理等特征。根据医学图像的特点, 本文着重选取以下几种区域特征^[1]:

1) 区域的面积

对于 $N \times M$ 的二值图象, 可通过对图象进行扫描并计算灰阶为 1 的象点总数来表示区域 R 的面积 A 。

$$A = \sum_{i=1}^m \sum_{j=1}^n f(i, j)。$$
(1)

收稿日期: 2002-09-23; 改回日期: 2003-05-24

基金项目: 江苏省高校科学研究项目(02KJD520005)

作者简介: 程 勇(1975-), 男, 江苏徐州人, 硕士生, 研究方向: 图像集成处理与模式识别。

其中 $f(i, j) = \begin{cases} 1, & (i, j) \in R; \\ 0, & (i, j) \notin R. \end{cases}$

2) 中心矩

如区域 R 是以其内部点的形式给出的, 矩对该区域的大小、旋转和平移的变化都是不变的。给定二维连续函数 $f(x, y)$, 其中 $(p+q)$ 阶矩定义是

$$m_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^p y^q f(x, y) dx dy. \quad (2)$$

式中 $p, q = 0, 1, 2, \dots$ 。

如果 $f(x, y)$ 是分段连续的, 只在 xy 平面的有限部分中有非零值, 则所有各阶矩都存在, 并且矩序列 m_{pq} 唯一地被 $f(x, y)$ 确定, 反之 m_{pq} 也唯一确定了 $f(x, y)$ 。

中心矩表示以中心为圆点进行计算的不变矩 u_{pq} , 因此中心矩 u_{pq} 为

$$u_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - \bar{x})^p (y - \bar{y})^q f(x, y) dx dy. \quad (3)$$

$$\text{式中 } \bar{x} = \frac{\sum_{-\infty}^{+\infty} x f(x, y)}{\sum_{-\infty}^{+\infty} f(x, y)}, \bar{y} = \frac{\sum_{-\infty}^{+\infty} y f(x, y)}{\sum_{-\infty}^{+\infty} f(x, y)}.$$

本文采用二阶混合中心矩

$$u_{11} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - \bar{x})(y - \bar{y}) f(x, y) dx dy. \quad (4)$$

对于 $m \times n$ 数字图像可用求和代替积分

$$u_{11} = \sum_{x=0}^m \sum_{y=0}^n (x - \bar{x})(y - \bar{y}) f(x, y). \quad (5)$$

3) 圆形度

$$F_1 = \frac{|P_1|}{\sum_{n=1}^M (|P_n| + |P_{-n}|)}. \quad (6)$$

式中 p_n 为通过边界链码计算的傅立叶系数, 链码为 $c^1, c^2, \dots, c^i, \dots, c^M$, 并且是沿边界曲线的逆时针方向构成的。

$$P_n \cong \frac{1}{2\pi i j} \sum_{m=1}^M a_m e^{j \left[\frac{\pi}{4} C_m - 2\pi n \frac{\sum_{k=1}^m a^k}{\sum_{k=1}^M a^k} \right]}. \quad (7)$$

其中 $n = \pm 1, \pm 2, \dots; a_k = \begin{cases} 1, & \text{若 } C_k \text{ 为偶数;} \\ 2, & \text{若 } C_k \text{ 为奇数;} \end{cases} \quad k = 1, 2, \dots, M.$

将坐标原点移至形心位置, 曲线 C 的参数方程可写为

$$U(t) = \sum_{n=1}^{\infty} (P_n e^{jnt} + P_{-n} e^{-jnt}), \quad 0 \leq t < 2\pi. \quad (8)$$

当傅立叶系数 P_n 中除 P_1 之外其他项全为零时, 那么 $U(t) = P_1 e^{jt}$ 表示轮廓曲线的形状是以半径为 $|P_1|$ 的一个圆。即当 C 为一个圆时, 相应的圆形度特征 $F_1 = \frac{|P_1|}{|P_1|} = 1$; 当 C 为其他形状时有 $0 < F_1 < 1$ 。并且特征满足平移、旋转、尺寸、起始点等条件的不变性。

4) 细长度

设 $E(t) = P_1 e^{jt} + P_{-1} e^{-jt}$ 表示形状的拟合椭圆, 其长半轴的长度为 $|P_1| + |P_{-1}|$, 短半

轴的长度为 $|P_1| - |P_{-1}|$ 。长短半轴长度之比反映形状的椭圆度(或称细长度),公式如下

$$F_2 = 1 - \frac{|P_1| - |P_{-1}|}{|P_1| + |P_{-1}|} \quad (9)$$

当 C 接近于圆时,其长短长度之比接近于 1,因此 $F_2 = 0$; 当为其他形状时有 $0 < F_2 < 1$ 。

5) 形心偏差度

$$F_3 = \frac{|P_{0C} - P_{0N}|}{|P_{1C} - P_{-1C}|} \quad (10)$$

其中, P_{nC}, P_{nN} 表示两个区域边界曲线 C, N 的付立叶级数展开获得的付立叶系数。 P_{0C}, P_{0N} 表示付立叶系数的零次项系数,表示曲线 C 和 N 的形心位置。 $|P_{0C} - P_{0N}|$ 直接反映两个形心之间的距离,当 C 与 N 为同心圆时 $F_3 = 0$,当两曲线相对偏心度较大时, $F_3 > 1$ 。

6) 核浆比

$$F_4 = \frac{\sum_{n=1}^{\infty} n(|P_{nN}|^2 - |P_{-nN}|^2)}{\sum_{n=1}^{\infty} n(|P_{nC}|^2 - |P_{-nC}|^2)} \quad (11)$$

癌变细胞胞核的体积增大速度比胞浆体积的增大速度快得多。所以相对于正常细胞,癌变细胞的胞核与胞浆之间比例失调。

2 识别模式

为了有效地实现分类识别,就要对提取的特征进行比较,并得到最能反映分类本质的特征,这就是特征的选择。模式识别中应根据具体的图像选择不同的特征。特征选择后应对选择的特征进行计算、学习、作标准化处理、求取特征向量。假设经试验选取特征向量 $S(s_1, \dots, s_n)$, $1 < n \leq 11$, 模式向量 $X = S^T(\alpha_1, \dots, \alpha_n) = (s_1\alpha_1, \dots, s_n\alpha_n)$, α_i 为 s_i 相对应的权重系数。模式识别的关键问题是决策分类。本文采用集群分析技术的最小距离分类器,并使用单原型模式表征分类。单原型表示一个集群子域中有一个典型性模式。

假设 C 类模式,以原型模式 $Z_1, Z_2, \dots, Z_c$ 为代表,原型模式 Z_i 一般可用均值 $m_i = \frac{1}{N_{iX}} \sum_{\beta_i} X$ 表示,其中 β_i 是第 Z_i 类的样本集, N_i 是 β_i 集中的样本数。则任一模式向量 X 与第 i 个原型模式之间的欧氏距离为^[2]

$$D_i = \|X - Z_i\| = \sqrt{(X - Z_i)^T (X - Z_i)} \quad (12)$$

最小分类器是计算未知模式 X 到每一类原型之间的距离并赋予它最小的一类模式。即:如果 $D_i < D_j$, 且 $j \neq i$ 则有 $X \in \omega_i$ 类。

另外欧式距离可用决策函数形式表示,将式(12)两边平方得:

$$D_i^2 = X^T X - 2X^T Z_i + Z_i^T Z_i = X^T X - 2(X^T Z_i - \frac{1}{2} Z_i^T Z_i) \quad (13)$$

因为距离为正值,所以 D_i^2 选择最小等价于 D_i 最小。式(13)中 $X^T X$ 是独立的,与类别无关。选择 D_i^2 最小等价于选择 $X^T Z_i - \frac{1}{2} Z_i^T Z_i$ 最大,因此决策函数可定义为

$$d_i(X) = X^T Z_i - \frac{1}{2} Z_i^T Z_i, \quad i = 1, 2, \dots, c \quad (14)$$

如 $d_i(X) > d_j(X)$, 且 $j \neq i, j = i, 2, \dots, c$, 有 $X \in \omega_i$ 类

本文对 128 例三种鳞细胞样本的上述区域特征进行计算, 通过标准化处理计算出各特征分量, 建立三类单原型模型, 并利用上述最近距离分离器对细胞进行识别。

3 检测结果实例

图 1 为待检测的三种鳞细胞的图像^[3]。第一种是正常鳞细胞, 形状较为规则, 核成圆形或椭圆形, 位于细胞中央, 胞浆分布均匀; 第二种是蝌蚪状的鳞癌细胞, 胞浆向外侧延伸, 核成圆形, 位于细胞的头部, 极不规则; 第三种是多核鳞癌细胞, 粗颗粒状染色不均匀, 核仁及核模清晰可见, 不规则。因此在对 128 例三种鳞细胞的识别中, 通过实验本文选择了圆形度、细长度、核浆比三种特征向量作为重要分类依据。对三类鳞细胞的上述特征进行计算, 建立三类原型模型 Z_1, Z_2, Z_3 , 并利用最小距离分类器构造如下判别函数

$$d_i(X) = X^T - Z_i^T Z_i, \quad i = 1, 2, 3. \tag{15}$$

如 $d_i(X) > d_j(X)$, 且 $j \neq i, j = 1, 2, 3$, 有 $X \in \omega$ 类。



图 1 三种鳞细胞 a. 正常鳞细胞; b. 蝌蚪状的鳞癌细胞; c. 多核的鳞癌细胞

Fig. 1 3 types of squamous cells

a. natural squamous cell; b. tadpole-like squamous cell carcinoma;

c. poly-nucleus squamous cell carcinoma

三种鳞细胞形状特征值见表 1。实验中回顾性检验没发生错判的现象, 但是从理论的角度分析, 样本不可能包括同类细胞的所有形态, 即原型模型不能代表一切同类细胞, 如待测的样本增加, 将有错判的可能。另外由于显微镜图像有时模糊, 这将导致在进行边界跟踪时, 边界变形, 也有产生错判的可能。本文在实验中未发生错判, 有可能是样本不够丰富。

表 1 鳞细胞形状特征

Table 1 Morphological-character of squamous-cell

区域特征	正常鳞细胞		蝌蚪状鳞癌细胞		多核鳞癌细胞	
	胞核	胞浆	胞核	胞浆	胞核	胞浆
圆形度	0. 7716	0. 7538	0. 7844	0. 4069	0. 6530	0. 6426
细长度	0. 1327	0. 1536	0. 9676	0. 1280	0. 3641	0. 3312
核浆比	0. 0438		0. 4109		0. 8610	

4 结 论

基于区域特征, 采用最小距离分类器可有效地识别三类鳞细胞, 这对提高细胞分类速度和精度、减少人的因素干扰有重要意义。实际中, 细胞的组成是复杂的。选择更多的特征, 建立更为完善的判别函数, 可能会进一步提高分类精度。对此, 笔者将做进一步的研究。

参考文献:

- [1] 边肇祺. 模式识别[M]. 北京: 清华大学出版社, 1990.
- [2] 夏良正. 数字图象处理[M]. 南京: 东南大学出版社, 1999.
- [3] 刘树范. 临床细胞学[M]. 北京: 人民卫生出版社, 1990.

A Cell Recognition Approach Based on Image Area-Character and Its Realization

CHENG Yong, FU De-sheng

(Department of Computer Science and Technology, NIM, Nanjing 210044, China)

Abstract: Firstly, this paper picks up area characters of the object including area, barycentric moment, the round degree, slightness degree and nucleoplasm ratio of the geometrical modality. The sign-method of single antetype and the least space sort-implement are used to construct a distinguish function. Finally, by comparing and computing area character of Squamous-cell, the author elects characteristic vectors through which sorts squamous cells can be discerned obviously, establishes the distinguish function and discusses the reliability of results.

Key words: cell recognition; area character; pattern; single prototype