

滤法(筛孔大小为 1500 微米/孔)每平方毫米内约有 35 个细胞,成群细胞团数在 3 个以下,重叠少,背景清晰,涂片基本符合计算机分析细胞的要求。若应用计算机进行细胞筛选,则嫌这种涂片方法的细胞量太少,影响检出率。可采用调整剩下的林格氏溶液量的方法来改变细胞浓度。

3. 制片过程是否标准化,例如稀释液倍数,离心速度,染色过程等是否一致,是获取细胞数据是否标准与准确的一个重要因素。

细胞涂片的制备质量直接影响后续环节(如信息输入与数据获取,细胞数字图象的区域划分与提取,细胞特征的抽提与选择,细胞的诊断与分类等),因此必须认真按严格要求作好。

参 考 文 献

[1] Tanaka, N. et al.: Field Test and Experimental Use of CYBEST Model 2 for Practical Gynecologic Mass Screening, In printing, 1979.
 [2] 菊地等: 日临会志, 16 (2), 205, 1977.
 [3] Tanaka, N. et al.: Acta. Cytol., 21, 531, 1977.

细胞图象的预处理方法

——食管癌细胞自动分类研究专题之二

王 今 著

(中国科学院生物物理研究所)

如前文所述由于涂片制备和染色工艺方面的问题,在视野范围内除观察到待识的主要图象外,往往在背景上还夹杂着一些次要图象。当应用电子计算机对前者进行研究时,必须经过预处理的步骤将后者去掉,以便提出待识的主要图象,使计算机能够进行细胞图形的区域划分、特征抽提,和分类判决。

预处理的方法很多,本文只给出我们初步

认为对细胞数字图象来说是行之有效的一种。

一、细胞图象预处理的步骤和方法

1. 平滑和边缘增强技术 目前所用方法大都是将输入的二维数字图象与模板窗口(例如 3×3 或 5×5 的窗口)进行离散卷积运算,结果就是处理后的某个象素。模板窗口的系数选得不同,得到的处理效果也不同,如下图所示:

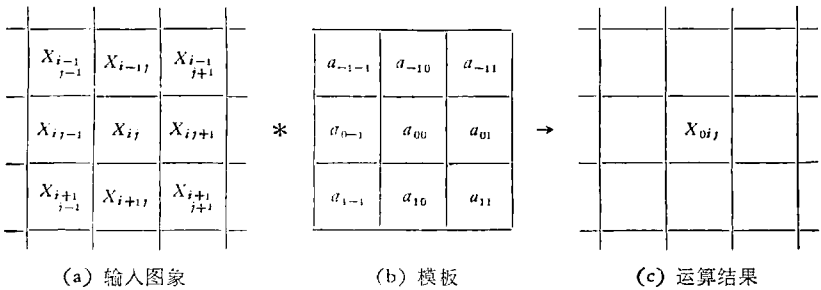


图 1 二维数字图象与模板的卷积运算示意图

运算方式如下:

$$X_{0i,j} = \sum_{k=-1}^1 \sum_{l=-1}^1 a_{kl} \cdot X_{i+k,j+l} \quad (1)$$

当模板中系数 a 均取正值时,效果相当于

低通滤波器,处理结果使图象“平滑”以消除高频噪声;如果系数 a 有一部分取负值时,其效果相当于高通滤波器,处理结果使图象的边缘反差增加。

在平滑处理中采用的模板系数如下:

1	1	1
1	4	1
1	1	1

运算形式是

$$X_{0ij} = (4X_{ij} + X_{i+1,j} + X_{i-1,j} + X_{i,j+1} + X_{i,j-1} + X_{i+1,j+1} + X_{i-1,j-1})/12 \quad (2)$$

在边缘增强处理中采用的模板系数如下:

-1	-2	-1
-2	64	-2
-1	-2	-1

运算形式是

$$X_{0ij} = 2X_{ij} - (X_{i-1,j-1} + X_{i+1,j+1} + X_{i-1,j+1} + X_{i+1,j-1} + 2(X_{i,j-1} + X_{i,j+1} + X_{i-1,j} + X_{i+1,j}))/32 \quad (3)$$

经此处理后图形边缘处的灰度较原来有所提高,而背景和图形内部的灰度则变化不大。

2. 作直方图,确定阈值,进行图形分割

(1) 频数—灰度直方图 设图象的灰阶范围 $G = 0, 1, 2, \dots, L$, 对经过平滑和边缘增强的图象 $A(i, j)$ 求每个灰度间隔中象点的数目 $n(G)$, 就得到频数—灰度直方图, 如图 2 所示。

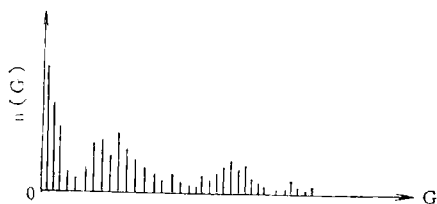


图 2 频数—灰度直方图

(2) 差分和一灰度直方图^[1] 对于图象 $A(i, j)$, 我们用 a_{ij} 表示位置在 (i, j) 处的象点的灰度值, a_{ij} 的周围 8 个相邻点用 a_{ij}^k 表示如下:

a_{ij}^4	a_{ij}^3	a_{ij}^2
a_{ij}^5	a_{ij}	a_{ij}^1
a_{ij}^6	a_{ij}^7	a_{ij}^8

取 a_{ij} 对第 k 个相邻点的差值 d_{ij}^k 为

$$d_{ij}^k = \begin{cases} \frac{a_{ij} - a_{ij}^k}{a_{ij} + a_{ij}^k + 1} \times 100 & \text{当 } a_{ij} - a_{ij}^k > 0 \text{ 时} \\ 0 & \text{其他} \end{cases} \quad (4)$$

对每个 a_{ij} , 求其差分和, 即令

$$q_{ij} = \sum_{k=1}^8 d_{ij}^k \quad (5)$$

则由 q_{ij} 组成的新数组 $Q(i, j)$ 表示对应 $A(i, j)$ 的每个点的灰度差分和, 再对 $Q(i, j)$ 求属于每个灰度间隔的差分之和 $S(G)$ 得

$$S(G) = \sum_{i=1}^m \sum_{j=1}^n q_{ij}(G) \quad (6)$$

$(G = 0, 1, 2, \dots, L)$

式中 $m \times n$ 为图象的总象点数。

用频数直方图对应的灰度间隔的频数 $n(G)$ 去除 $S(G)$ 便得到平均差分和一灰度直方图

$$NDM(G) = S(G)/(n(G) + T) \quad (7)$$

其中 T 是为了防止在一些孤立点上出现较大的 $NDM(G)$ 值所加的权重, 可根据处理的具体对象及先验知识加以规定。

最后得到的平均差分和一灰度直方图的形式如图 3 所示:

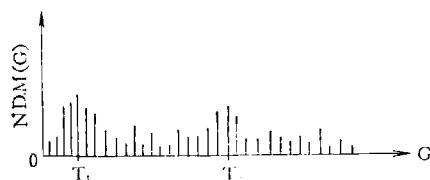


图 3 平均差分和一灰度直方图
 T_1, T_2 处为阈值

(3) 选取阈值, 进行图形分割 依上述方法作出的平均差分和直方图有时不能满足检测峰值

的条件,因此常常需要对直方图进行光滑处理。处理的方式和次数可根据具体对象及象点的多少而定,经过光滑处理后的直方图,可按求极大值的条件在直方图上依次取五点,在整个灰度范围内寻找峰值,如果根据先验知识规定出第一峰和第二峰的灰度范围,则选取的峰值可能更符合实际情况。例如灰阶范围为0—63时,对于我们所处理的细胞图象,第一峰值可限制在2—10灰度内,第二峰值可限制在16—48内。用选取的第一个阈值 T_1 (如图3)对图象 $A(i, j)$ 进行阈值处理,就可将背景和物体分开,将图形分割成面积不等的几块区域。

3. 利用找连通区的办法进行区域提取^[2]
 区域提取常用的方法是对经过图形分割后的各个区域求出各自的连通区,并将各连通区域的连通成份(象点)用不同记号标记,然后根据需提取出所感兴趣的部分。例如一幅细胞图象,经过阈值处理后有三个相互孤立的区域(用0, 1分别表示背景和物体区域),通过找连通区的办法,可提取出其中面积最大的那块图形,即所需观测的主要图象,如图4所示:

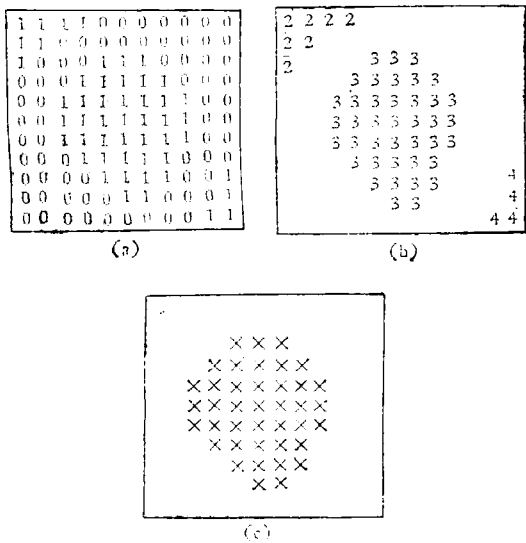


图4 用找连通区的办法进行区域提取
 (a) 经阈值处理后的二值图
 (b) 将找出的连通区用不同数码标记
 (c) 提取面积最大的区域

在实验中,我们用了两种找连通成份的方法都取得成功,一种是扫描增殖法,另一种是跟

踪法,每种方法中可采用4向连通也可采用8向连通。下面就这两种方法作一些简要说明:

(1) 扫描增殖法 对经过阈值处理并二值化的图形 $Q(i, j)$ 开始扫描搜索,当遇到某点 $q(i, j) = 1$ 时,在另一个数组 $C(i, j)$ 中相应的位置上置1,此点即为第一连通区的起始点。然后对数组 $C(l, m)$ 扫描搜索。当遇到某一点 $c(l, m) = 1$ 时,则在 $Q(l, m)$ 中相应点 $q(l, m)$ 的8个相邻点中找其是否有等于1的点,如有,则在 $c(l, m)$ 的8个相应邻点 $c(l + k, m + h)$ ($k, h = -1, 0, 1$)上置1,并将 $c(l, m)$ 点置2,将 $q(l, m)$ 点置0。继续对数组 $C(l, m)$ 中的下一点搜索,如发现有 $C(l, m + 1) = 1$,则重复上述过程使 $c(l, m + 1)$ 点置2, $q(l, m + 1)$ 点置0。如此继续搜索,直到 $C(l, m)$ 中不再有1为止,则此时在 $Q(l, m)$ 中的第一个连通区已全部置0,而在 $C(l, m)$ 中相应的区域全部置2。继续对 $Q(i, j)$ 的扫描搜索,当再次遇到新的 $q(i, j) = 1$ 的点时,说明它已是第二个连通区的第一点,重复上述过程,使 $q(l, m)$ 置0,使 $c(l, m)$ 置3,直到第二个连通区找完为止,这个过程一直进行到使 $Q(i, j)$ 全部置0为止。在 $C(i, j)$ 中则是以2, 3, 4, ……标记的连通区。

扫描增殖法,具体算法如下:

输入: 二值化图象 $Q(i, j)$ 。

输出: 目标图象 $C(i, j)$

步骤:

① N 赋以2; $C(i, j)$ 赋以0;

② 对 $Q(i, j)$ 扫描搜索,遇到 $q(i, j) = 1$ 时,对 $C(i, j)$ 中相应的点 $c(i, j)$ 赋以1;

③ 对 $C(l, m)$ 扫描搜索,遇到 $c(l, m) = 1$ 时,检查 $q(l, m)$ 点的8个邻点 $q(l, m + 1)$, $q(l, m - 1)$, $q(l + 1, m)$, $q(l - 1, m)$, $q(l - 1, m - 1)$, $q(l - 1, m + 1)$, $q(l + 1, m - 1)$, $q(l + 1, m + 1)$,对其中为1的点,在 $c(l, m)$ 相应的邻点上也赋1; $q(l, m)$ 赋以0; $c(l, m)$ 赋以 N ;

④ 继续对 $C(l, m)$ 扫描搜索,如遇 $c(l, m) = 1$ 则转③;

⑤逆向对 $C(l, m)$ 扫描搜索, 遇到 $c(l, m) = 1$ 时转③;

⑥正向对 $C(l, m)$ 扫描搜索, 遇到 $c(l, m) = 1$ 时转③;

⑦ $N + 1$ 赋给 N ; 继续对 $Q(i, j)$ 扫描, 当遇到 $q(i, j) = 1$ 时则转②;

⑧计算 $C(i, j)$ 中各连通区的面积, 找出最大面积连通区的标记符号 N ;

⑨对 $C(i, j)$ 扫描, 当 $c(i, j) = N$ 时 $c(i, j)$ 赋以原始图象的对应点的值 $a(i, j)$ 否则 $c(i, j)$ 赋以 0;

⑩输出 $C(i, j)$ 。

(2) 跟踪法 对经过阈值处理并二值化的图象 $Q(i, j)$ 以循环变量 i, j 进行扫描搜索, 当遇到 $q(i, j) = 1$ 时, 在 $C(i, j)$ 中相应的位置上赋 2, 并将坐标原点移到 $q(i, j) = 1$ 的点上 (用 $q(l, m)$ 表示), 按逆时针方向依次检查 $q(l, m)$ 点的 8 个邻点, 当遇到的邻点为 1 时, 将坐标原点移到该点上, 并在 $C(l, m)$ 相应位置上置 2, 将原来的 $q(l, m)$ 置 0, 然后以新原点为中心再依次检查它的 8 个邻点, 重复上述操作直到在 $Q(i, j)$ 中第一个连通区被“抹掉”(置 0)在 $C(i, j)$ 中用符号 2 标记该连通区为止, 重复上述操作, 并用 3 来标记在 $C(i, j)$ 中的第二个连通区。按此方式进行到 $Q(i, j)$ 中的点全部为 0 为止, 在 $C(i, j)$ 中则以 2, 3, 4, ……标记出 $Q(i, j)$ 中的所有连通区。

具体算法如下:

输入: 二值化图象 $Q(i, j)$;

输出: 目标图象 $C(i, j)$;

步骤:

① N 赋以 2; $C(i, j)$ 赋以 0;

②对 $Q(i, j)$ 扫描, 遇 $q(i, j) = 1$ 时 $c(i, j)$ 赋以 N ; 变换坐标原点即将 i 换为 l, j 换为 m ;

③以 $q(l, m)$ 为坐标中心点, 按逆时针方向依次检查其 8 个相邻点 $q^k(l, m)$, ($k = 1, 2, 3, \dots, 8$), 如遇到第 k 个邻点 $q(l + g, m + h) = 1$ 时 ($g, h = -1, 0, 1$), 则 $c(l + g, m + h)$ 赋以 N ; $q(l, m)$ 赋以 0; 转移中心点

坐标到 $q(l + g, m + h)$ 上来即 l 赋以 $l + g, m$ 赋以 $m + h$; 转③;

④ N 赋以 $N + 1$; 转②;

⑤计算 $C(i, j)$ 中各连通区的面积, 找出最大面积连通区的标记符号 N ;

⑥对 $C(i, j)$ 扫描, 当 $c(i, j) = N$ 时 $c(i, j)$ 赋以原始图象对应点的值 $a(i, j)$ 否则 $c(i, j)$ 赋以 0;

⑦输出 $C(i, j)$ 。

二、结 果

一幅数字化图象经过平滑、边缘反差增强, 阈值处理后基本上能将背景和图形正确地分离开, 再用找连通区的办法找出各个连通区, 进行区域提取, 可选出面积最大者作为待观测图形。经过这一系列的处理后, 为特征提取作好了准备。

对一些白血球, 食管上皮细胞涂片的显微镜数字图象在国产 TQ-16 计算机和 013 计算机上进行处理, 证明处理方法可行, 基本上达到了预定要求。

三、讨 论

只要对图象进行正确的分割, 用找连通区的办法进行区域提取, 总是可取出所要的图形的。

扫描增殖法和跟踪法各有特点, 前者编写程序简单, 能适应各种图形, 但对于较大的或形状复杂的, 可能要多次进行扫描检查, 因此所需的处理时间较长。后者只需一次扫描检查, 但对形状较复杂、有单层线状突出的, 没有闭合边界的图形, 由于要加上各种限制, 编写程序较复杂一些。

如何正确地选取阈值以便正确地进行图形分割是一个并未很好解决的问题。用上述选取阈值的方法对一些图象进行处理, 其中大部结果尚能与实际情况符合, 但对某些细胞图象的处理结果则与实际情况相差较大。因此有待进一步总结经验, 提高阈值选取的准确度。例如可根据实际经验对阈值的灰度范围及其它有关条

件加上适当限制,或者寻找选取阈值的新方法。

参 考 文 献

[1] Watanabe, S. (渡边贞一) and CYBEST Group: Co-

mputer Graphics and Image Processing, Vol. 3, 350—358, 1974.

[2] Rosenfeld, A.: J. A. C. M., 17 (1), 146—160, 1970.

用阈值法对细胞数字图象进行区域划分

——食管癌细胞自动分类的研究专题之三

汤之永 杨存荣

(清华大学自动化系)

应用电子计算机自动识别细胞首先需将细胞样本片上的细胞图象(图1见封2)由带有计算机控制的扫描显微分光光度计转换成一组反映各位置的透光度或吸光度(即光密度)大小的细胞灰度数字图象 f (图2见封2)。我们采用的扫描步距为 4μ 、 2μ 、 1μ 、 0.5μ ,光度分辨率为7 bit,光波长约为500毫微米。为要准确地抽提细胞的特征,首先需对细胞灰度数字图象 f 进行找连通区的预处理,已如前文所述。预处理后,要对数字图象进行区域划分(又称边界检测)。划分的方法很多,本文介绍我们用“阈值法”进行初步试验的结果。

一、用三种直方图求“阈值”的方法

这里的“阈值”是指细胞与背景,胞核与胞浆的光密度分界值。

1. 频数直方图法 假设有一细胞灰度数字图象 f ,按其灰度值分布情况大体可分为三个区域,即背景区 f_B 、胞浆区 f_C 、胞核区 f_N 。各区域内的灰度值基本符合正态分布,而各区域的灰度均值(期望值)又是可分的。

现用 H_j 表示灰度等级, $j=0,1,2\cdots 255$ 。

用 E_j 表示灰度值 Z_i 等于 H_j 的象点 e_i 的集合

$$E_j = \{e_i: Z_i = H_j\}.$$

定义:灰度值 $Z_i = H_j$ 的全部象点总个数(即集合 E_j 中全部元素 e_i 的个数)为灰度值 H_j

所出现的频数,记为 F_j 。

F_j 是 H_j 的函数,作灰度频数直方图如图3(a)。

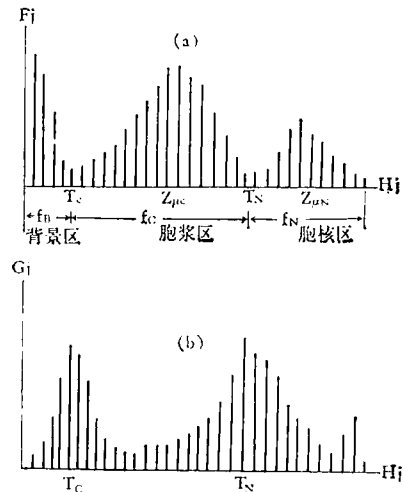


图3 (a) 理想的灰度频数
(b) 理想的灰度差分直方图

图中 $Z_{\mu C}$ 和 $Z_{\mu N}$ 分别相当于胞浆和胞核的灰度均值,对应直方图中两个谷的灰度值 Z_C 和 Z_N 分别选作为胞浆和胞核的“阈值”。

2. 差分(梯度和)直方图法 假设有一细胞灰度数字图象 f ,按其灰度值可分为三个区域,处于各区域内的相邻象点间的灰度差值是比较小的。而处于任一区域边界两边的相邻象点间的灰度差值则是比较大的,如图4所示。

现用 Z_i 表示第 i 个象点的灰度值