DOI: 10.16016/j.1000-5404.1998.06.028

统计聚类分析方法在细胞图像分拆处理中的应用

易 东* (第三军医大学预防医学系卫生统计学教研室) 重庆,400038
李辉智 (西南政法学院刑侦系) 重庆,400031

提 要 目的:建立一种细胞图像自动分拆方法,以达到将原始细胞图最佳分拆为胞质、胞核图像之目的,而且此分析过程可自动进行;方法:利用概率统计聚类分析中组间距离与组内距离的思想,对一幅图像象素点灰度值的集合进行分类;结果:通过对淋巴细胞图像进行分析,其分解图像准确、结果可靠;结论:本研究的分拆方法合理、准确且可移植性强。

关键词 聚类分析;细胞图像分析;自动分析
中图法分类号 R319

在医学图像处理中,特别是对细胞图像的处理,往往需要利用分拆技术将原始图像分拆为胞质胞核图,然后分别对胞质、胞核图进行预处理,最后合并分拆图以达到图像分析的目的[1,2]。本研究利用统计聚类分析法,建立了一种快速、可靠的自动图像分拆模型。

1 聚类分析思想引入

淋巴细胞微核计算机图像自动识别研究中,需要将原始图像分拆为细胞核和质图,然后分别独立地进行去除噪声处理。但由于其涂片往往采用的是 Giemsa染色法,胞质是浅蓝紫时,胞核是深紫色,见图 1这样,不可能在色度上对之进行分拆,只能采用灰度强度对之进行分解。因此,如何准确地对图像进行分拆是以后处理中至关重要的一步。细胞分拆往往采用直方图图像分析法,但如何很好地选择阈值从而将物体从背景中提取出来是一个非常重要的课题。在此方面已有一些研究,在理想情况下其直方图有明显的峰值和谷底,这时阈值的确定就是一个非常简单的问题了。然而实际情况却并非如此,往往会出现图势比较平坦和一些噪声的存在。目前已有用于克服上述缺陷的常用方法,如谷底增加技术,它采用了 Laplacian法和梯度法[1];直方图差分法增强技术等[2],这些方法均采用原始图中象素点的邻域(或边界)的信息来修正直方图,以便有效地确定阈值。另外,也有采用参数化方法处理直方图,它借助了高斯分布和统计决策方法,然而此方法不足之处是计算量太大和不稳定性。

以上提出的各种阈值选择方法均未提出"最佳阈值"的判别标准,而且还需借助于图像的先验知识。为此,本研究利用统计分析中聚类分析的思想,建立了微核图像胞质和胞核的分解模型,此方法可实现自动分拆且不涉及图像的任何先验知识[3,4]。

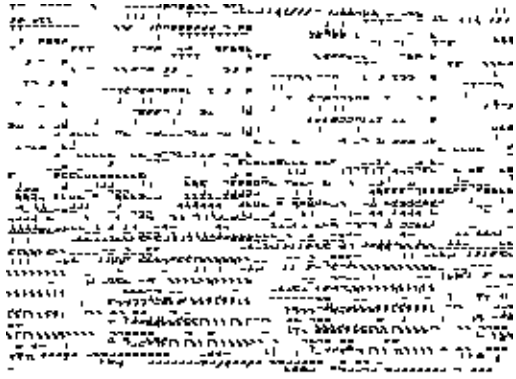


图 1 淋巴细胞图像

Fig 1 The image of human lyphocytes

2 基本原理

在微核图像分解中,其主要任务是将之分解为胞核、胞质两部分。因而我们不妨将图中的象素点分为三大类,即:胞核、胞质、背景,见图 2

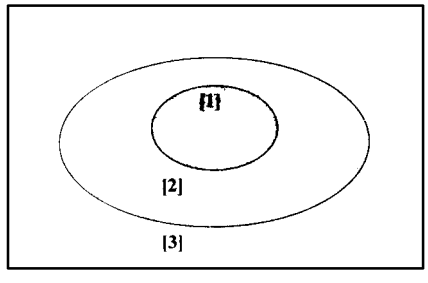


图 2 分割示意图
1:胞核 2:胞质 3:背景

Fig 2 Segmentation schematic diagram
1: nuclci 2: cytoplasm 3: background

* 易 东,男,34岁,副教授,博士

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

图像准确分析的关键是最佳判别函数的建立,本研究建立差别函数的主要思想是:使得灰度特性在类中相差最小,类间相差最大,即统计聚类分析的思想。下面,我们将主要讨论组内距离和组间距离的表达式,其具体实现如下。

设图中象素点的灰度变化范围为:[1,2,3,…,L],

各组灰度水平上象素点数对应为:$n_1, n_2, n_3, \cdots, n_L$,

象素点总数为:$N = n_1 + n_3 + n_3 \cdots + n_L$,

其直方图的概率分布为:

$$P_i = n_i / N,\ P_i \geqslant 0,\ \sum_{i=1}^{L} P_i = 1 \tag{1}$$

假设图像中的象素点分为三类,即

C_0 胞核,其灰度变化范围为:[1,2,3,…,K_1],

C_1 胞质,其灰度变化范围为:[K_1+1, K_1+2,…,K_2],

C_2 背景,其灰度变化范围为:[K_2+1, K_2+2,…,L],

于是各类出现的概率为:

$$W_0 = P_r(C_0) = \sum_{i=1}^{K_1} P_i = W(K_1) \tag{2}$$

$$W_1 = P_r(C_1) = \sum_{i=K_1+1}^{K_2} P_i = W(K_2) - W(K_1) \tag{3}$$

$$W_2 = P_r(C_2) = \sum_{i=K_2+1}^{L} P_i = 1 - W(K_2) \tag{4}$$

其中 $W(K) = \sum_{i=1}^{K} P_i$ (5)

因而类中的平均灰度水平为:

$$\mu_0 = \sum_{i=1}^{K_1} i P_r(i/c_0) = \sum_{i=1}^{K_1} i P_i / W_0 = \frac{\mu(K_1)}{W(K_1)} \tag{6}$$

$$\mu_1 = \sum_{i=K_1+1}^{K_2} i P_r(i/C_1) = \sum_{i=K_1+1}^{K_2} i P_i / W_1 = \frac{\mu(K_2) - \mu(K_1)}{W(K_2) - W(K_1)} \tag{7}$$

$$\mu_2 - \sum_{i=K_2+1}^{L} i P_r(i/C_2) = \sum_{i=K_2+1}^{L} i P_i / W_w - \frac{\mu(L) - \mu(K_2)}{1 - W(K_2)} \tag{8}$$

其中 $\mu(K) = \sum_{i=1}^{K} i P_i$ (9)

容易验证:$W_0 + W_1 + W_2 = 1$,

且 $W_0\mu_0 + W_1\mu_1 + W_2\mu_2 = \mu(L)$ (10)

于是,下面就可定义组内、组间方差了。

组内方差的定义为:

$$\sigma_0^2 = \sum_{i=1}^{K_1} (i - \mu_0)^2 P_r(i/C_0) = \sum_{i=1}^{K_1} (i - \mu_0)^2 P_i / W_0 \tag{11}$$

$$\sigma_1^2 = \sum_{i=K_1+1}^{K_2} (i - \mu_1)^2 P_r(i/C_1) = \sum_{i=K_1+1}^{K_2} (i - \mu_1)^2 P_i / W_1 \tag{12}$$

$$\sigma_2^2 = \sum_{i=K_2+1}^{L} i - \mu_2)^2 P_r(i/C_2) = \sum_{i=K_2+1}^{L} (i - \mu_2)^2 P_i / W_2 \tag{13}$$

于是 $\sigma_W^2 = W_0\sigma_0^2 + W_1\sigma_1^2 + W_2\sigma_2^2$ 即是组内方差。

组间方差定义为:

$$\sigma_B^2 = W_0(\mu_0 - \mu(L))^2 + W^1(\mu_1 - \mu(L))^2 + W(\mu_2 - \mu(L))^2 \tag{14}$$

总方差定义为:

$$\sigma_T^2 = \sum_{i=1}^{L} (i - \mu(L))^2 P_i \tag{15}$$

不难验证:

$\sigma_W^2 + \sigma_B^2 = \sigma_T^2$

利用聚类分析中的可分离性测度函数可知

$$\lambda = \sigma_B^2 / \sigma_W^2 \qquad K = \sigma^2(L) / \sigma_W^2 \qquad \eta = \sigma_B^2 / \sigma_T^2 \tag{16}$$

均可作为其测度标准,下面考察 η 函数

$$\eta(K_1, K_2) = \alpha_B^2(K_1, K_2) / \sigma_T^2 \tag{17}$$

由(15)式可知,其 σ_T^2 并非是 K_1, K_2 的函数。

因而,"最佳阈值" K_1^*, K_1^* 的选择可归结为使测度函数达到极大,即

$$\sigma_B^2(K_1^*, K_2^*) = \max \sigma_B^2(K_1, K_2) \qquad 1 \leqslant K_1 < K_2 < L \tag{18}$$

这里

$$\sigma_B^2(K_1, K_2) = W(K_1)\left(\frac{\mu(K_1)}{W(K_1)} - \mu(L)\right)^2 + (W(K_2) - W(k_1))\left(\frac{\mu(K_2) - \mu(K_1)}{W(K_2) - W(K_1)} - \mu(L)\right)^2 + (1 - W(K_2))\left(\frac{\mu(L) - \mu(K_2)}{1 - W(K_2)} - \mu(L)\right)^2$$

3 算法过程

下面,我们将给出算法的框图,见图 3

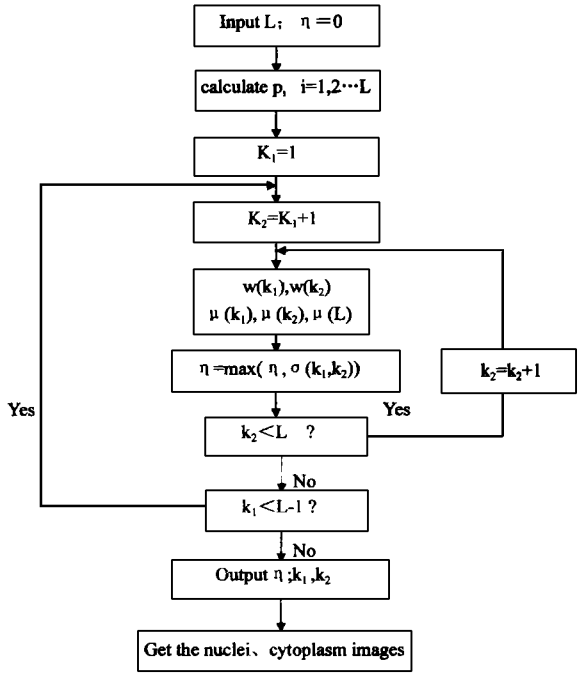


图 3 分解算法框图

Fig 3 The flow chart of segmentation algorithm

4 实验结果

下面,我们对上述算法进行实验验证,其分解结果见图 4

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

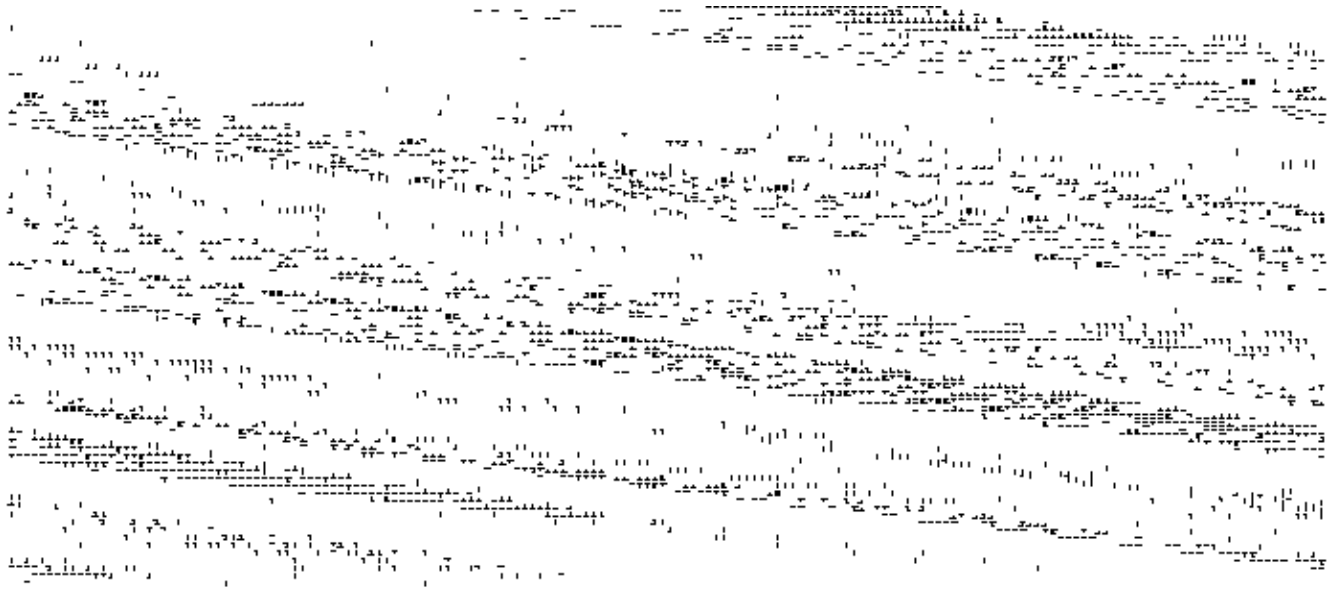


图 4　算法验证结果　(a)原图　(b)分解图

Fig 4　Test results by segmentation algorithm　(a: original image　b: segmentation)

5　讨论

从大量实验结果可知,本方法对淋巴细胞微核图像的分割与实际情况非常吻合,而且运行速度快。

本方法的自动化程度很高,只需输入原始图像即可。另外,本系统也设置了人工输入门限法,其目的是为了处理一些特殊情况。

此方法在分类类数不太多时(如 2、3类时)非常有效(如本研究中分为背景、胞质、胞核三大类),当类数增多,其可靠性将减弱。这是因为作为判别函数的 σ_B 随着分类类数的增多,其函数曲线将变得平缓,从而作为判别的灵敏度将降低。即在计算时其优化问题的收敛性和收敛过程将变得更为复杂。

本方法要求分类目标之间在灰度强度上有一定的差别。如目标灰度过于接近,且所占面积太小,则其识别结果与视觉的结果会有一定的差别[5]。

参考文献

1　Weszka J S, Nagel R N, Rosenfeld A. A threshold selection technique. IEEE Trans. Computer, 1974, (C-23): 1322

2　Watanabe S, CYBEST Group. An automated apparatus for cancer prescreening: CYBEST Comp graph. Image Process, 1974, (3): 350

3　Adadasun M, King R A. Low-level segmentation of multispectral image via agglomerative clustering of uniform neighborhoods. Pattern Recognition, 1995, 21: 261

4　F U K S, Mui J K. A survey of image segmentation. Pattern Recognition, 1996, 23(1): 3

5　Harms H, Aus H W, Hauck M, *et al*. Segmentation of stained blood cell images measured at high scanning density with high magnification and high numerical aperture optics. Cytometry, 1986, 7(6): 522

(收稿: 1998-04-15;修回: 1998-10-16)

(编辑　刘　鹏)

Application of statistical clustering in segmentation of image

Yi Dong, Li Huizhi (Department of Medical Statistics, Third Military Medical University, Chongqing, 400038)

Abstract　Objective: To explore a method of automatic segmentation for cell image. Methods: Using the distance within one group and that between two groups, the set of pixels were divided into two types composed of nuclei and cytoplasm respectively. Results: Segmentation tests indicated that the model was very useful and accurate. Conclusion: This method is useful for not only the segmentation but also image processing.

Key words　cell image; automatic segmentation

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net