

彗星实验图像分析中彗星分割阈值的选择

闫学昆 陈英 杜杰 张学清 骆亿生

(军事医学科学院 放射与辐射医学研究所 北京 100850)

摘要 彗星实验(单细胞凝胶电泳)图像自动分析中,彗星轮廓的合理分割是后续各项指标分析的前提。首先采用几种常用的图像边缘检测算法对彗星轮廓进行提取,并分析了彗星图像的特点;然后对文献中提到的和现有彗星分析软件(如CASP、TriTek CometScore™)中的彗星轮廓分割方法进行比较;最后采用一种基于图像内容的阈值分割方法,实现了最佳阈值的选择,并得出了彗星轮廓分割的几点原则。

关键词 彗星实验, 图像分析, 分割阈值, 剂量-效应曲线

中图分类号 Q345+2, TL72, R318

彗星实验(Comet assay, CA),又称单细胞凝胶电泳(Single cell gel electrophoresis, SCGE),是一种能够在辐射损伤早期快速、灵敏检测DNA损伤的有效方法,具有良好的剂量-效应关系^[1,2]。该方法具有灵敏、可靠、快速、所需样品少等优点,已被广泛应用于环境生物样品检测、化学毒物检测、辐射生物学、遗传毒理学等研究领域。

对于彗星图像的自动分析,国外已有免费的CASP、TriTek CometScore™等软件,也有价格昂贵、可移植性差的KOMET和KS400等分析系统;国内籍涛等^[3]人还研制了IMI1.0自动分析系统。这些彗星图像分析软件、分析系统以及有关文献中,对于如何将彗星轮廓从彗星图像背景中合理地分割出来并未作详细说明。国内多数科研人员也只是使用这一自动分析工具,尤其是免费的CASP自动分析软件使用较多,而对彗星轮廓分割阈值的选取问题尚欠深入研究。然而,彗星轮廓的合理分割是DNA测量的关键,其分割效果直接影响到剂量-效应曲线和实验结果的可比性。

图像分割是图像分析中的一项基本技术,其目的是确定对象区域并将对象区域从图像背景中分割出来。图像分割方法可分为三类^[4]:(1)区域法,将各像素划归到相应的物体或区域的像素聚类方法;(2)边界法,通过直接确定区域间的边界来实现分割的方法;(3)边界—区域法,首先用边缘检测算子生成初始边缘,然后经细化、修补、去分支、边界跟踪得到封闭的边缘,最后进行边界填充形成

分割区域。在彗星实验中,如何准确、合理地从背景中提取彗星轮廓、确定彗星头尾的分界线,是后续分析彗星图像各类指标的基础。本文采用不同的阈值分割方法对彗星图像进行阈值分割,并进行比较。

1 彗星图像及其特点

图1(a)为一幅典型的彗星图像^[5],已采用中值滤波的方法进行了降噪预处理。图1(b~f)为分别采用Sobel、Prewitt、Roberts、Canny算子^[6,7]、Otsu^[8]方法对其进行边缘检测的效果图。可以看出,由于彗星尾部与背景的界线不明显,直接采用边缘检测算子检测并不能将彗星轮廓从图像中分割出来;即便是普遍认为较为优秀的Otsu阈值分割法,也没能检测到完整的彗星轮廓。又考虑到彗星各项指标分析的统一性,应采用同一标准将彗星头部和尾部从图像背景中分割出来,故局部阈值法也不宜采用。

对图1(a)的原始彗星图像进行三维数值化,见图2,其中X轴和Y轴分别为彗星图像的长度和宽度,Z轴表示彗星图像中各像素点的灰度值。从图2可以看出,彗星头部的灰度值较大,形成一个明显的特征峰,故彗星头部易于从背景中分割出来,见图1(b, c, d, f);但彗星尾部的灰度值较低,且非常平缓地逐渐过度到了背景当中,即彗星尾部边缘和背景之间的灰度差较小。

图3为图1(a)的直方图(纵坐标为对数坐标),

军队十一五重大专项项目子课题

第一作者:闫学昆,男,1981年6月出生,2006年于海军工程大学辐射防护与环境保护专业获工学硕士学位,现在军事医学科学院攻读博士学位,放射医学专业

通讯联系人:陈英,yingchen29@yahoo.com.cn

收稿日期:初稿 2008-05-05,修回 2008-06-12

可以看出,整幅图像的灰度值较低(分布在 7~130 范围内),且灰度直方图的双峰特性不明显,故难以直接采用直方图阈值分割的方法来确定其分割阈值。

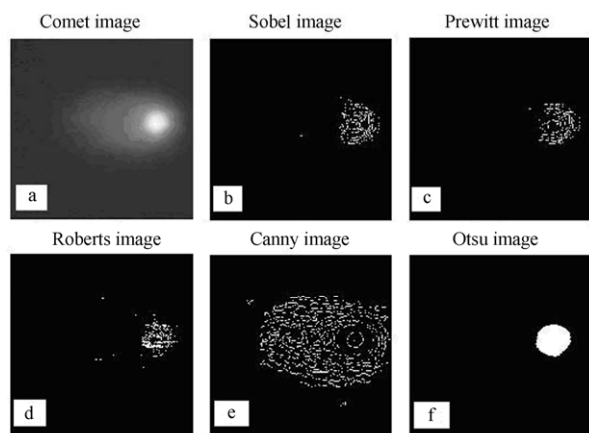


Fig.1 Original comet image (a) and the edge detection results in different arithmetic operators, such as Sobel (b), Prewitt (c), Roberts (d), Canny (e) and the Otsu method (f)

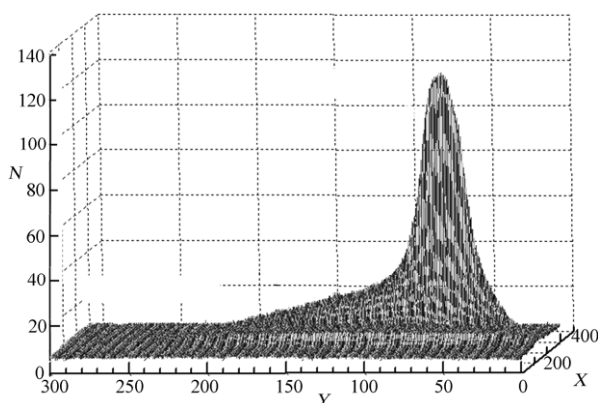


Fig. 2 3-D visualization of the comet image Fig.1(a)

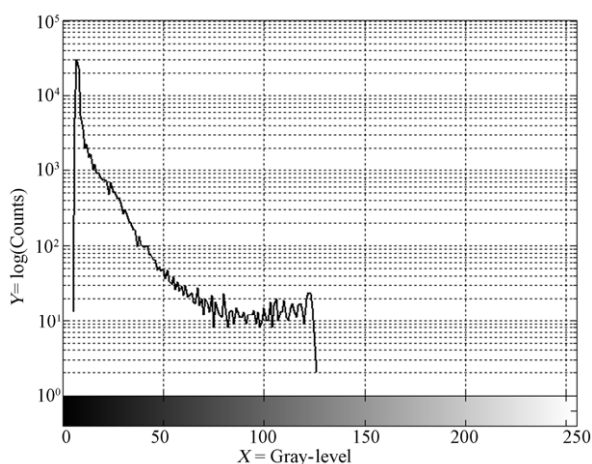


Fig. 3 Gray-level histogram for the comet image Fig.1(a)

2 彗星图像阈值的选择

对于图 2,若能合理地将彗星图像的背景部分剔除,那么剩余的部分即为彗星的轮廓(包括头部和尾部)。常用的阈值分割方法是把图像灰度分成不同的等级,然后通过设置灰度阈值的方法来确定有意义的区域或分割物体的边界。阈值(T_h)的选择是一个关键而较困难的问题:阈值越大,背景和噪声剔除得越干净,但丢失的彗星信息也越多;阈值越小,保留的彗星信息越多,但背景数据和噪声也会随之增加。这是一对矛盾,故有必要对阈值的选择进行深入研究^[9]。

需要指出的是,采用阈值分割法来分割彗星图像时,该阈值一旦确定、并在软件中设定之后,对于所有采用该阈值进行分割的彗星图像来说,这个标准都是一致的,相当于系统误差。如果该系统误差太大,将会导致彗星各项分析指标的误差较大。

文献[10]中对于一幅大小为 512pixel × 512pixel、灰度级为 0~255 的彗星图像,简单地采用公式(1)设置阈值得到二值图像,进而由公式(2)完成彗星图像的二值化,从而实现彗星轮廓的分割。但是,对于阈值 T_h 选取,作者直接给出了公式(1),并未做详细说明,其科学性和合理性有待进一步验证。

$$T_h = \frac{Gray_{max} + Gray_{min}}{2.25} \quad (1)$$

$$g(x, y) = \begin{cases} 0, & f(x, y) < T_h \\ 255, & f(x, y) \geq T_h \end{cases} \quad (2)$$

式中, $Gray_{max}$ 和 $Gray_{min}$ 分别为彗星图像中像素的最大值和最小值。

2.1 TriTek CometScore™ 软件中阈值的选择

TriTek CometScore™ 软件^[11,12]中,软件研发人员并未给出阈值选择的详细说明和理论依据,而是在主界面的右侧设置了一个“Cutoff”滚动条(变化范围从 0%至 100%),用户需要根据彗星轮廓的分割效果来自行调整指示器(分割阈值),进而实现彗星轮廓的提取,图 4 是滚动条指示器为 77%时的彗星分割效果。对于不同的彗星图片,这具有很大的不确定性,彗星轮廓的分割效果因人而异,会导致较大的误差。

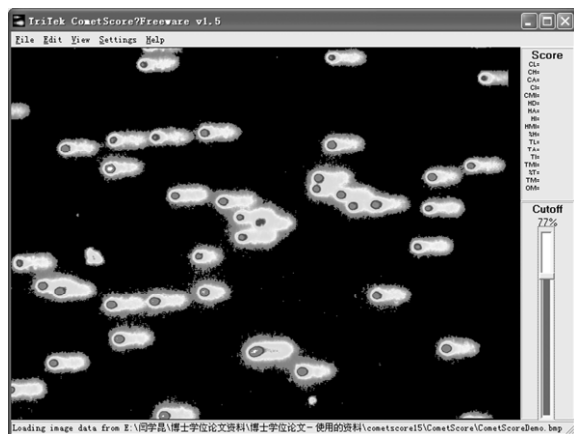


Fig.4 Different comet contours at different segmentation threshold in the TriTek CometScore™ software

2.2 CASP 软件中阈值的选择

目前使用较广的 CASP 软件，关于此阈值的设定问题，作者也未作理论推导和说明，而是给出了根据实验测得的最优阈值^[5,13]：

(1) 彗星头部中心阈值 (Head center threshold, HCT)，最优 HCT=0.8，其含义为：只有超过最大荧光强度 80% 的点才被认作是彗星头部中的点，再通过计算这些点的重心来确定彗星头部的中心；

(2) 彗星尾部阈值 (Tail threshold, TT)，最优 TT=0.05；

(3) 彗星头部阈值 (Head threshold, HT)，最优 HT=0.05；

(4) 彗星阈值 (Comet thresholds, CT)，最优 CT=0.05，CT 的取值与 TT 和 HT 有关，只有当 TT 和 HT 取值相同时，CT 达到最佳效果，一般使用时只设置 CT 即可。

其中，0 表示背景的荧光强度，1 代表最大荧光强度，即确定彗星头部中心的最佳阈值为最大荧光强度的 80%，彗星轮廓分割的最佳阈值为最大荧光强度的 5%。该最优值 (HCT、CT) 是 CASP 软件研发人员根据实验图像多次测量比较后得到的结果，并将其设置为 CASP 软件的默认阈值。

采用实验测定的方法来确定合理的阈值，这是合理可行的，且方法比较简单。因为对于一套固定的彗星实验图像自动分析系统，摄取彗星图像的条件是一致的。国内多数科研人员只是使用该软件，而对于阈值的设置是否适合于所拍摄彗星图像的分析，还缺乏深入考虑。

2.3 基于图像内容的阈值计算法

文献[14]给出了一种基于图像内容分析的两步

阈值分割方法。先将采集到的彗星图像进行高斯噪声污染估算，通过计算污染后图像中高斯噪声的污点个数来设置一个初始阈值 T_h ， T_h 和图像中高斯噪声的污点数有关，经验上把污点数为 307 时的阈值作为初始阈值 T_h ；然后设置一个调节阈值 T_2 ，以进一步提高分割效果，并根据实验测试结果设置 $T_2=T_h-5$ 。

设 $I(x, y)$ 为输入图像灰度、 $O(x, y)$ 为输出图像的灰度，则：

$$\begin{cases} \text{if } I(x, y) > T_h \rightarrow O(x, y) = 1 (\text{保留}) \\ \text{if } I(x, y) < T_2 \rightarrow O(x, y) = 0 (\text{剔除}) \end{cases} \quad (3)$$

对于边界点：如果该点与保留区域连接则 (a)，如果该点与保留区域分离则 (b)：

$$\text{if } T_2 \leq I(x, y) < T_h \begin{cases} O(x, y) = 1 (a) \\ O(x, y) = 0 (b) \end{cases}$$

该方法也属于实验测定法，但是该方法充分考虑了阈值边界部分的像素点，与直接设置阈值的方法相比，能够进一步提高彗星轮廓的分割效果。

在 MATLAB 软件中采用该阈值选择算法，对图 1 (a) 中的原始彗星图像进行阈值分割，并依次设置阈值 T_h 为 8、10、12、...、30，设置 $T_2=T_h-3$ ，分割效果见图 5。可以看出，采用不同的阈值 T_h 时，得到的彗星轮廓图像并不同：

当阈值较低 (如 $T_h=8$) 时，彗星轮廓图中混入了较多的背景；

当阈值较高 (如 $T_h=30$) 时，彗星轮廓图明显减小，剔除了较多的尾部信息；

只有当阈值适当 (如 $T_h=16$ 、 $T_h=18$) 时，得到的彗星轮廓图像较为理想；

设置不同分割阈值时彗星尾部并不同，这将会导致慧尾参数测量的差异。

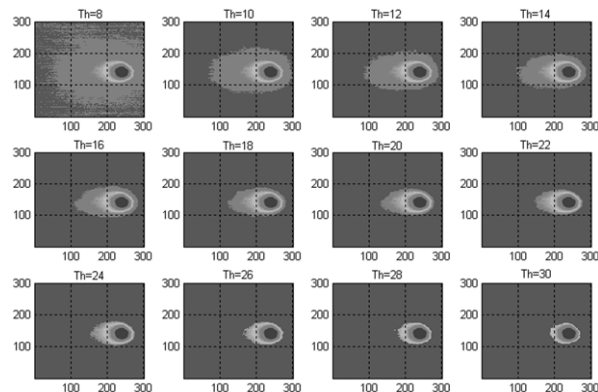


Fig.5 Different comet contour images at different segmentation threshold

采用阈值 ($T_h=17$) 分割得到彗星图像, 再分别采用 Sobel、Prewitt、Roberts、Canny 算子和 Otsu 方法对其进行边缘检测, 检测效果见图 6。与图 1 比较可以看出, 几种算子均能较好地检测出彗星的轮廓, 这是由于经分割得到的彗星轮廓边缘处的灰度级具有了明显的间断, 即有了鲜明的边缘。

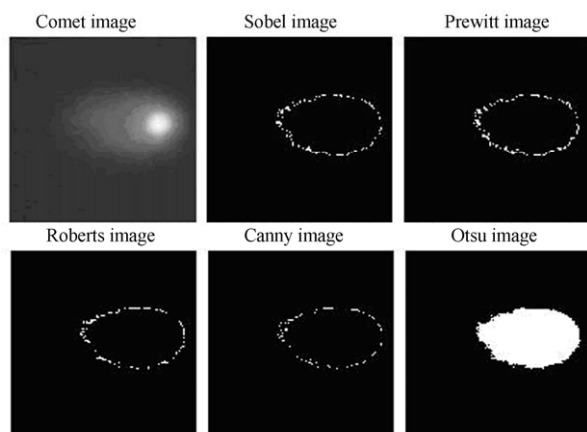


Fig.6 Edge detection results in different arithmetic operators when the segmentation threshold is 17

此外, 通过利用 MATLAB 对最大熵原则法、迭代法、流域分割法等几种图像分割方法^[15]编程测试发现, 流域分割算法是一种可靠、理想的彗星图像分割方法^[16], 可采用该方法选择合理的分割阈值, 作为彗星图像自动分析系统的默认阈值。

3 讨论

在采用彗星实验进行辐射生物剂量估算时, 使用 CASP、IMI1.0 等自动分析软件可以快速、准确地分析彗星的各项指标, 建立剂量-效应曲线。通过分析以上分割方法和分割效果可以得知, CASP 软件中采用的实验测定最佳分割阈值的方法是合理、可行的; 基于图像内容分析的双阈值计算方法, 不但能继承实验测定方法的优点, 还充分考虑了阈值边界部分的像素点, 能够进一步提高彗星轮廓的分割效果; 流域分割算法可作为一种科学的最佳分割阈值选择方法。针对彗星图像中彗星轮廓的提取问题, 由以上分析可以得到如下几点结论:

(1) 彗星轮廓的合理分割, 是后续确定彗星头尾、分析彗星各类指标的前提; 将图像中的彗星轮廓(头部和尾部)从背景中提取出来时, 应采用同一阈值标准剔除背景。

(2) 彗星的头部各点的灰度值(荧光强度)较大, 形成了一个明显的特征峰, 故彗星头部易于从背景中分割出来; 而彗星尾部区域的灰度值较低, 尾部边缘非常平缓地逐渐过度到背景当中, 即彗星

尾部边缘和背景之间的灰度值差别较小且连续, 使得彗星轮廓图像的分割较为困难。

(3) 对于同一彗星图像采用不同阈值分割得到的彗星轮廓并不同, 这将导致彗星总长度、尾长、尾矩、尾惯量等一系列彗星参数测量的差异: 当阈值选择过高时(见图 5, $T_h=30$), 由于慧尾信息丢失过多, 将导致测得的彗星参数值明显偏小; 当阈值选择过低时(见图 5, $T_h=8$), 由于慧尾中混入了过多的背景信息, 将导致测得的彗星参数值明显偏大、失真。即便其它实验条件完全一致, 由于分割阈值的不同或改动, 都将会导致实验结果的差异。

综上所述, 对于一套固定的彗星实验图像自动分析系统, 首先应合理、科学地选择和设定彗星的分割阈值, 尽可能减小系统误差, 以提高实验结果的可比性; 该分割阈值一旦设定后, 对所有图像中彗星轮廓的提取都使用同一标准, 不宜随意改动; 在估算辐射剂量时选用的分割阈值, 至少要和建立剂量-效应标准曲线时使用的阈值一致。

参考文献

- 1 蒋亚齐, 陈英, 刘芬菊, 等. 辐射防护, 2006, 26(6): 353-358
JIANG Yaqi, CHEN Ying, LIU Fenju, *et al.* Radiat Prot, 2006, 26(6): 353-358
- 2 刘强, 姜恩海, 李进, 等. 中华放射医学与防护杂志, 2006, 26(6): 573-576
LIU Qiang, JIANG Enhai, LI Jin, *et al.* Chin J Radiol Med Prot, 2006, 26(6): 573-576
- 3 籍涛, 黄耀熊, 谭润初, 等. 中国医学物理学杂志, 2001, 18(2): 80-82
JI Tao, HUANG Yaoxiong, TAN Runchu, *et al.* Chin J Med Phys, 2001, 18(2): 80-82
- 4 何东健. 数字图像处理. 西安: 西安电子科技大学出版社, 2003: 1-15
HE Dongjian. Digital Image Processing. Xi'an: XiDian University Press, 2003: 1-15
- 5 CASP-Comet assay software project
<http://www.casp.of.pl>
- 6 阮秋琦, 阮宇智. 数字图像处理(冈萨雷斯, 第二版). 北京: 电子工业出版社, 2003: 460-513
RUAN Qiuqi, RUAN Yuzhi. Digital image processing (Gonzales R C, 2nd ed.). Beijing: Publishing House of Electronics Industry, 2003: 460-513
- 7 罗军辉, 冯平. MATLAB7.0 在图像处理中的应用. 北京: 机械工业出版社, 2005: 201-223
LUO Junhui, FENG Ping. MATLAB7.0 Application in

- image processing. Beijing: China Machine Press, 2005: 201-223
- 8 Otsu N A. IEEE Trans on System, Man, and Cybernetics, 1979, **9**(1): 62-66
- 9 Helma C, Uhl M. Mutat Res, 2000, **466** (1): 9-15
- 10 Harrison C, Pearson J D, Bilton R F, *et al.* 11th IEEE Symposium on Computer-Based Medical System (CBM'98): 234-239
- 11 <http://tritekcorp.com>
- 12 <http://autocomet.com>
- 13 Konca K, Lankoff A, Banasik A, *et al.* Mutat Res, 2003, **534**(1-2): 15-20
- 14 Böcker W, Rolf W, Bauch T, *et al.* Cytometry, 1999, **35**(2): 134-144
- 15 章毓晋. 图象分割. 北京: 科学出版社, 2001: 46-91
- ZHANG Yujin. Image Segmentation. Beijing: Science Press, 2001: 46-91
- 16 Rivest J F, Tang M, McLean J, *et al.* Instrumentation and Measurement Technology Conference, IEEE, 1996(1): 111-114

Reasonable threshold value used to segment the individual comet from the comet assay image

YAN Xuekun CHEN Ying DU Jie ZHANG Xueqing LUO Yisheng

(Institute of Radiation Medicine, Academy of Military Medical Sciences, Beijing 100850, China)

ABSTRACT Reasonable segmentation of the individual comet contour from the Comet Assay (CA) images is the precondition for all of parameters analysis during the automatic analyzing for the CA. The Otsu method and several arithmetic operators for image segmentation, such as Sobel, Prewitt, Roberts and Canny were used to segment the comet contour, and characters of the CA images were analyzed firstly. And then the segmentation methods which had been adopted in the software for CA automatic analysis, such as the CASP, the TriTek CometScoreTM, were put forward and compared. At last, a two-step procedure for threshold calculation based on image-content analysis is adopted to segment the individual comet from the CA images, and several principles for the segmentation are put forward too.

KEYWORDS Comet assay, Image analysis, Segmentation threshold, Dose-responses curve

CLC Q345+.2, TL72, R318