

## 基于小波描述子的染色体自动识别方法研究

王海龙, 王文义

(中原工学院 计算机学院, 河南 郑州 450007)

**摘要:**提出了一种对染色体图像进行自动分析的方法. 首先基于边缘检测的区域生长法从细胞染色体显微图像中分割出单个染色体, 然后利用 Haar 小波描述子刻画单个染色体形状特征, 最后再据此对染色体进行配对. 实验结果证明: 利用该方法完成一张染色体显微照片的分析工作仅需要大约 10 min, 较人工作业大大提高染色体图像的分析效率, 并且在染色体分散良好、不存在交叉染色体的情况下, 染色体的正确配对率达到 90% 以上.

**关键词:**染色体分析; 边缘检测; 区域生长; Haar 小波变换

**中图分类号:** TP391.4

**文献标志码:** A

### 0 引言

染色体作为主要的遗传物质 DNA 载体, 研究它的结构和功能是细胞遗传学必不可少的关键环节, 其重要性受到了越来越多科学家的关注.

对染色体的分析与识别, 主要是从生物方法培养的细胞涂片的显微镜图像出发, 获取一个细胞全部染色体的配对核型图. 传统的人工染色体分析是一项非常繁重的工作: 医务人员必须把染色体的显微图片进行复制, 剪下每个染色体, 一个一个地进行比较、配对. 近几年来, 国外相继出现了一些染色体自动分析系统, 虽然其中不乏效果较好的, 但都不同程度的存在这样那样的问题, 如对染色体重叠图像分割, 存在染色体特征提取和分类困难等等, 都直接影响了分析系统的效果. 另外, 国外的这类设备价格一般都比较昂贵, 我们国内的中、小医院很难承受.

笔者采用区域生长法分割出单个染色体, 用小波描述子刻画单个染色体的形状特征, 最后完成染色体分类, 大大缩短了染色体的分析与识别所需时间.

### 1 系统概述

典型的染色体分析系统<sup>[1]</sup>工作流程如图 1 所

示. 其分析过程描述为: ①染色体分割. 通过分割染色体图像来获取单条染色体, 以便进一步分析和匹配染色体. ②特征提取. 染色体特征主要有: 染色体的形状特征描述(包括边界特征和区域特征), 染色体的长、短臂长以及着丝粒指数等. ③染色体配对. 根据提取的染色体特征, 采用合适的相似度计算方法对染色体进行配对分类, 然后再由人工交互纠正.

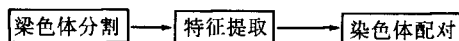


图 1 染色体图像分析系统结构

Fig. 1 Chromosome image analysis system structure

### 2 边缘检测的区域生长法分割染色体

由于 Canny 算子具有定位精度高、检测性能好等优点, 笔者采用该算子来检测单条染色体边缘, 提取独立染色体图像轮廓<sup>[2]</sup>. 首先用  $3 \times 3$  高斯滤波模板与原始图像进行卷积, 以平滑图像消除图像噪声. 然后利用 Sobel 导数算子计算图像灰度沿  $x$ 、 $y$  方向的导数  $G_x$ 、 $G_y$ , 并求出梯度的大小和方向, 寻找图像中的可能边缘点. 最后利用双阈值检测通过双阈值递归寻找图像边缘点, 实现边缘提取. 图 2 是在光学显微镜下的细胞染色体图像. 图 3 则是对图 2 通过 Canny 算子得到的边缘检测.

收稿日期: 2011-04-04; 修订日期: 2011-06-12

基金项目: 河南省基础与前沿技术研究计划项目(082300410300)

作者简介: 王海龙(1976-), 男, 河南邓州人, 中原工学院讲师, 主要研究领域为数字图像处理技术, E-mail: whlseadragon@163.com.

对于图像灰度值变化不明显处,提取的边缘不一定连续,有断裂现象,不能形成封闭边缘.笔者采用传统的区域生长算法,选择染色体像素灰度值最大的点作为初始种子像素点,利用边缘像素点集合的平均灰度作为后续区域生长的判决条件来进行染色体图像分割<sup>[3]</sup>.

利用像素邻域的概念,通过标注种子像素八连通域分量进而标注单个染色体,实现染色体图像中各条不同染色体的自动分离,还可以对较小的染色体或杂质实行剔除,从而保留需要用于进行识别的染色体<sup>[4]</sup>.对图2经过分割处理后的单条染色体图像如图4所示.



图2 细胞染色体显微图像

Fig.2 Microscopic cell chromosome image



图3 Canny算子染色体边缘图像

Fig.3 The edge of chromosome using canny operator



图4 分离染色体图像

Fig.4 Detached chromosome images

### 3 染色体边界小波描述子提取

假设单条染色体图像的轮廓是XY平面上一条由N个点组成的数字化边界,其中每一个边界点用其坐标 $(x_k, y_k)$ 表示.从任意点 $(x_0, y_0)$ 开始,沿着逆时针方向遍历整个边界,依次经过坐标 $(x_1, y_1), (x_2, y_2), \dots, (x_{N-1}, y_{N-1})$ .考虑到单条染色体图像用64个点描述其形状已经可以满足.当边界点数大于64时,可通过等步长采样的方法得到64个点予以解决,而对于边界点N数小于64的情况,则需要插值处理.

Haar小波基实现简单,计算速度快,且滤波器更短,更容易刻画平面闭合曲线,因此笔者选择Haar小波基对单条染色体图像的轮廓曲线进行小波变换,求得小波系数可以作为服装轮廓的小波描述符<sup>[5]</sup>.

对表示轮廓的64个点序列提取小波半径描述子算法的具体描述为<sup>[6]</sup>:①求单条染色体轮廓的形心;②求各个点对应形心的距离(半径);③对半径序列基于Haar函数进行6层的小波分解;④对小波变换后的结果,保留尺度系数,对每层补充细节的小波系数进行有选择的截断,作为单个染色体图像的描述符.

如果选择所有的小波变换系数作为半径描述符,计算量特别庞大.在实验中对得到的6层Haar小波变换系数,保留尺度系数,对补充细节的小波系数进行有选择的截断,以一定精度描述了原轮廓的特征,在这里取32个系数点,该特征矢量定义为: $M = [c_1, x_2, \dots, c_{32}]$ .

### 4 染色体的配对

采用欧氏距离来计算单个染色体之间的相似度<sup>[7]</sup>.使用小波半径描述子作为染色体分类的依据.两条染色体的匹配相似度计算公式如下:

$$\text{similar}(A, B) = 1 - \sum_{i=1}^N (C_A(i) - C_B(i)) \quad (1)$$

式中: $A, B$ 表示任意两条染色体; $C_A(i)$ 表示A染色体小波描述矢量的第i项; $C_B(i)$ 表示B染色体小波描述矢量的第i项; $\text{Similar}(A, B)$ 表示A染色体与B染色体的相似度;当 $\text{Similar}(A, B)$ 接近1时表示A染色体与B染色体越相似,匹配度越高.

通过上述公式算出每条染色体与其它染色体的匹配相似度,实验结果如表1.从表中可以看出,染色体A1与染色体A2的匹配相似度最大,视为两者匹配;染色体A3与A5相匹配;染色体

A4与A10匹配;染色体A6与A7匹配;染色体A8与A9匹配.实验结果表明,利用图像处理技术提取染色体的相关特征,可以比较满意地实现染色体的有效、正确匹配.

表1 图4中染色体匹配相似度

Tab. 1 Chromosome matching in figure 4

代号	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
A1	1.000	0.998	0.244	0.458	0.239	0.127	0.120	0.754	0.741	0.469
A2	0.998	1.000	0.249	0.459	0.240	0.128	0.121	0.755	0.743	0.471
A3	0.244	0.249	1.000	0.877	0.997	0.899	0.849	0.765	0.768	0.879
A4	0.458	0.459	0.877	1.000	0.862	0.859	0.849	0.753	0.848	0.996
A5	0.239	0.240	0.997	0.862	1.000	0.902	0.879	0.562	0.564	0.850
A6	0.127	0.128	0.899	0.859	0.902	1.000	0.995	0.513	0.518	0.831
A7	0.120	0.121	0.849	0.849	0.879	0.995	1.000	0.508	0.511	0.827
A8	0.754	0.755	0.765	0.753	0.562	0.513	0.508	1.000	0.997	0.813
A9	0.741	0.743	0.768	0.848	0.564	0.518	0.511	0.997	1.000	0.815
A10	0.469	0.471	0.879	0.996	0.850	0.831	0.827	0.813	0.815	1.000

#### 4 结论

利用图像分析技术进行染色体图像的自动分析和特征提取,最终实现染色体的自动配对.在人工作业情况下,一个熟练的工作人员完成一张染色体显微照片的分析工作至少需要1 d时间.若用笔者提出的分析技术则仅需要10 min,大大地提高了工作效率,减少人为因素对特征提取的干扰、有效提高染色体配对的效率和准确性.

笔者将小波变换运用到染色体的轮廓特征分析上,提取染色体轮廓半径序列,利用harr小波基对半径序列进行小波变换提取小波描述子作为染色体特征.结果表明:对于一幅染色体图像,在其单条染色体不出现交叉的情况下,利用所开发的染色体自动识别系统可以使染色体的正确配对率达到90%以上.

#### 参考文献:

[1] 陈晓华,俞昌,袁志强.人类染色体识别系统[J].

计算机工程与应用,2002(10):231-233.

[2] 李牧,闫继红,李戈.自适应Canny算子边缘检测技术[J].哈尔滨工程大学学报,2007,28(9):1002-1007.

[3] 张泽彪,李式巨,程敏.图像区域标记和边沿检测的两步法[J].计算机工程与设计,2004,25(4):625-628.

[4] 章毓晋.图像分析——图像工程:中册[M].北京:清华大学出版社,2005:87-91.

[5] 王瑜,穆志纯,付冬梅.基于小波变换和规范型纹理描述子的人耳识别[J].电子学报,2010,38(1):241-243.

[6] 刘仰龙,王从庆,高珏,等.基于小波描述子的水果果形分类[J].浙江大学学报:农业与生命科学版,2010,36(3):322-328.

[7] PIPER J, GRANUM E. On fully automatic feature measurement for banded chromosome classification[J]. Cytometry, 2005 10(3):242-255.

### Automatic Chromosome Recognition Research Based on Wavelet Descriptor

WANG Hai-long, WANG Wen-yi

(School of Computer Science, Zhongyuan University of Technology, Zhengzhou 450007, China)

**Abstract:** This paper introduces a method to implement the automatic chromosome analysis. The regional growth based on the chromosome detection is adopted to get the single chromosomes, and then Haar wavelet descriptor based on the chromosome shape is extracted to chromosome pairing. The work of chromosome analysis is completed within ten minutes, which improves efficiency greatly, and the correct rate of chromosome pairing reach more than 90% where the chromosomes slide isolation each other, without overlapping.

**Key words:** chromosome analysis; edge detection; regional growth; Haar wavelet transform