

# 有效的染色体分割新方法

敬雪,滕奇志,唐棠,王姝

JING Xue, TENG Qi-zhi, TANG Tang, WANG Shu

四川大学 电子信息学院 图像信息研究所, 成都 610064

Institute of Image Information, College of Electronics and Information, Sichuan University, Chengdu 610064, China

E-mail: cias@263.net

JING Xue, TENG Qi-zhi, TANG Tang, et al. Efficient method for segmentation of chromosome images. *Computer Engineering and Applications*, 2008, 44(35): 182-184.

**Abstract:** Getting the correct segmentation-point is a major step toward the segmentation of touching and overlapping chromosomes, when analyzing and recognizing chromosome images. This paper applies a calculating method of edge chain-code to find the concave points correctly which segmentation points belong to, that is to say the candidate segmentation points. And then the right ones are estimated and selected by the distance threshold among segmentation points, as well as the arc threshold of edge. At the same time, it introduces a proper theory for two chromosomes which touching head-to-head or head-to-end.

**Key words:** chromosome; image segmentation; chain code; concave points

**摘要:**在染色体图像分析与识别中,将粘连或是交叠的染色体分割开的关键技术是找到正确的分割点。通过使用一种边界链码的计算方法来准确定位分割点所属的凹点,即候选分割点;再利用候选分割点间的距离阈值和边界弧长阈值判断并筛选出正确的分割点。同时提出了两条粘连的染色体在其端部粘连或首尾粘连情况下的正确分割方法。

**关键词:**染色体;图像分割;链码;凹点

**DOI:** 10.3778/j.issn.1002-8331.2008.35.054 **文章编号:** 1002-8331(2008)35-0182-03 **文献标识码:** A **中图分类号:** TP391.4

核型分析是遗传病临床诊断的一种工具,运用图像处理手段,可对染色体的数目和形态特征进行定性和定量的描述<sup>[1]</sup>。尽管核型分析是选择在染色体分散良好、形态清楚的有丝分裂中期进行<sup>[1]</sup>,但染色体难免会粘连或交叠在一起,为每条独立染色体的特征提取及分析带来障碍。因此需要研究图像分割算法,将标本图片中粘连或交叠在一起的染色体分开,形成各个独立的单位个体。这一步对于真实准确的反映每条染色体的特征是非常重要的,而其中的关键就是要正确的找到分割点。

有丝分裂中期的染色体成长条状,所以在染色体的粘连或是交叠处通常会有凹点,而此时的凹点就是分割点。在做分割前,先将粘连或交叠在一起的染色体看作一条,那么其具体做法是:

(1)沿着一条染色体边界找到它的一系列凹点,作为候选分割点。

(2)在候选分割点中,找到其中几个满足一定的阈值条件(分割条件)的点作为分割点:粘连分割需要2个分割点,而交叠分割往往需要3~4个分割点。

(3)连接分割点,完成染色体分割。

对于寻找分割点,现在普遍的做法是沿着染色体的轮廓找到曲率的极小值点作为候选分割点<sup>[2-3]</sup>,通过求一系列极小值点的间距来找到满足一些阈值条件的点,此时筛选出的极小值点

就是分割点。在寻找曲率极小值点的过程中,人们通常先求得染色体轮廓曲线上两点间弧长 $\Delta S$ ,切线的夹角 $\Delta\theta$ ,再运用公式 $K = \frac{\Delta\theta}{\Delta S}$ 求得曲率 $K$ 。这种按照数学方法寻找凹点虽然易懂,但由于计算机进行的是数字化图像处理,在实现得到弧长和切线夹角时往往涉及到离散几何知识的运用,比如曲线拟和,因此在理论实现方面来说增加了复杂度。同时用离散方法求得的曲率不带有方向信息,为了能正确的描述染色体边界走势,还要对离散曲率进行矢量修正,这无疑又加大了工作量。

为了提高该图像分析识别系统的运行速率,同时能很好的反映染色体边界的曲率变化程度。采用对边界链码的一系列计算来获得染色体形状描述,在形状描述的基础上获得凹点,最后通过一系列判断条件准确找到分割点。实验证明该方法简明有效。

## 1 链码在形状分析中的应用

### 1.1 链码

链码是表示数字图像中当前一个像素到另一个像素的方向值,有4-方向链码和8-方向链码,如图1所示,使用8-方向链码。

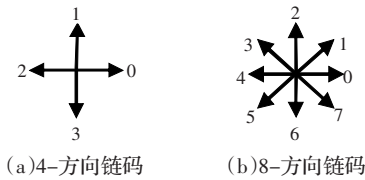


图1 方向链码

在染色体图像预处理时,先进行图像的二值化和去噪,然后就要用 freeman 链码跟踪算法对二值化的染色体进行边缘跟踪,得到边界链码。由于链码具有方向性,它可以仅仅通过链码数组准确描述出染色体轮廓各个相邻像素的位置关系,这种表达方式可以大大减少边界表示所需要的数据量。也正是由于链码的方向性,链码才能够很好的反映出图像目标的形状,为后面运用链码进行染色体图像分割提供理论依据。不过由于噪声的影响,染色体边界会出现毛刺等不平滑的现象,所以在用 freeman 链码算法进行边界跟踪好,要首先要对染色体的边界进行平滑,同样也是通过链码来完成,采取基于链码的轮廓平滑模板<sup>[4]</sup>,实验证明平滑结果良好。

## 1.2 链码计算

采用陆宗骥提出的三点链码计算方法<sup>[5]</sup>。

先用 freeman 链码跟踪算法获得的  $n$  条染色体的边界链码,设边界有  $N$  个点。且本文使用的是顺时针跟踪路线,然后进行三点链码计算,计算步骤如下:

(1)求相对链码  $RelativeCode[i]$ ,即出该点的链码值减去进入该点的链码值。

(2)求绝对链码  $AbsoluteCode[i]$ ,即将前一点的绝对链码加上当前点的相对链码赋予当前点的绝对链码。

(3)求相对链码和  $SumCode[i]$ ,即将当前点与相邻前两点的绝对链码和相加。链码和可以表示染色体边界上当前点的切线方向。

(4)链码差  $DiffCode[i]$ ,即将当前点的后第三个点的链码和减去当前点的链码和链码差  $DiffCode[i]$ 反映了染色体边界上当前点的曲率变化,虽然它并不等于曲率值,但却与该处的曲率成正比。当链码差小于 0 时,表示该点为凹点,且负值越小越凹陷当链码差大于 0 是,表示该点为凸点,且正值越大越凸出。

## 1.3 链码计算结果的使用

对于图 2 所示的染色体,现在通过分别使用三阶 B 样条逼近轮廓的方法计算该染色体的边界曲率和使用三点链码方法计算边界曲率,所得到的曲率分布曲线图如图 3 所示。



图2 粘连在一起的染色体

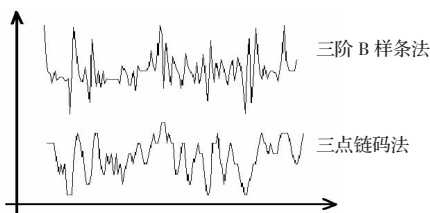


图3 两种方法计算图2染色体的曲率对比图

通过图 3 的两种曲线对比,可以看出用三点链码计算得到的曲率与传统算法得到的结果大体一致,尤其对于曲率极值点

的定位都相同。既然链码差等效与曲率值,那么在寻找候选分割点时,就可以直接通过对链码差的阈值判断,就能找到合乎要求的点

## 2 染色体粘连分割

染色体分割最主要是利用其粘连或交叠处的轮廓凹陷的特征,再结合分割路径的长度以及以及分割点间的边界弧长等判断条件来筛选出正确的分割点。

### 2.1 候选分割点

在分割之前暂且将粘连或交叠在一起的染色体作为一条染色体。对每条染色体的边界链码进行三点链码计算,求得边界上每个点的链码差  $DiffCode[i]$ ,然后与阈值  $T_c$  比较。对于第  $k$  条染色体:

$$D_k = \{i | i = 0, 1, 2, \dots, N\} \quad (1)$$

$$C_k = \{j | DiffCode[j] < T_c, j \in D_k\} \quad (2)$$

其中,  $D_k$  表示该条染色体轮廓上每个点的序号集合,  $C_k$  表示该条染色体所有满足阈值条件的点集,  $T_c$  为负值。

由于染色体的长条状,会出现边界有连续点的链码差都满足阈值的情况。如果随便挑选其中一个作为候选分割点,就会发生偏离实际分割的误差。因此要对点集  $C_k$  中的点再进行一次筛选。算法描述如下:

#### 步骤 1

$$C_k = \bigcup_{p=0}^P E_k^{(p)}, p=0, 1, \dots, P \quad (3)$$

$$E_k^{(p)} = \{e_0, e_1, e_2, \dots, e_j, \dots\}, e_j - e_{j-1} = 1 \quad (4)$$

其中,  $E_k^{(p)}$  为  $C_k$  划出来的第  $p$  个子集,子集内的点都满足在染色体边界上连续分布。

#### 步骤 2

$$R_k^{(p)} = \{r | \max\{|DiffCode[r]|\}, r \in E_k^{(p)}\} \quad (5)$$

$$R_k = \bigcup_{p=0}^P R_k^{(p)} \quad (6)$$

其中,  $R_k^{(p)}$  为第  $k$  条染色体的第  $p$  个候选分割点集,如果该集合内的元素不止一个,即同时出现相等的最大值,则保留  $R_k^{(p)}$  内最后出现的那个元素作为候选分割点,删除另外的元素。  $R_k$  为第  $k$  条染色体的候选分割点集。

### 2.2 轻度粘连分割

轻度粘连的染色体之间的粘连少,分割路径短,往往比染色体的宽度要小的多,不易与那些严重弯曲或是在着丝粒处凹陷的染色体混淆,所以容易区分。因此要分开轻度粘连的染色体,只要满足下列的阈值条件即可:

$$(\exists x)(\exists y)(x, y \in R_k)$$

$$S = \{(x, y) | L(x, y) < T_d \wedge Arc(x, y) > T_a\} \quad (7)$$

其中,  $S$  为最终确立的分割点集,  $L(\cdot)$  计算两点间的距离,  $T_d$  为分割路径阈值,  $Arc(\cdot)$  计算两点间的弧长,  $T_a$  为染色体边界弧长阈值。连接  $S$  内每个二元元素所表示的点,完成粘连分割。

### 2.3 重度粘连分割

这种情况有时被认为是 T 型交叠或 h 型交叠<sup>[6]</sup>。但考虑到这是发生在染色体之间的粘连部分很多的时候,故这里将此归类为重度粘连。如图 4(a)所示。大面积的融合使候选分割点间

的距离不能再满足于轻度粘连分割时使用的阈值  $T_d$ 。同时考虑到染色体着丝粒处也有凹陷,如图 4(b)所示;或者由于所采集的染色体图像样本中,染色体可能自身的严重弯曲凹陷,如图 4(c)所示。因此为了有别与这两种情况,避免误分割,重度粘连不再考虑候选分割点的距离,而利用其它特点作为条件进行分割。



(a)重度粘连染色体 (b)着丝粒凹陷 (c)严重弯曲凹陷的染色体  
图4 容易混淆判断的染色体

### 2.3.1 中轴有分叉的情况

由于大部分的重度粘连染色体经细化后所提取的中轴都有分叉情况,分叉点出现在它们粘连处的附近,如图 5 所示。

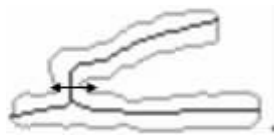


图5 中轴有分叉的染色体轮廓

因此只要在两个候选分割点的附近有中轴的分叉点,就能排除掉因着丝粒凹陷而误判成重度粘连的情况(亚端着丝粒染色体除外)。表示如下:

$$(\forall q)(q \in Q_k)$$

$$(\exists x)(\exists y)(x, y \in R_k)$$

$$S = \{(x, y) | L(x, q) < T_q \wedge L(y, q) < T_q \wedge Arc(x, y) > T_a\} \quad (8)$$

其中,  $S$  为最终确立的分割点集,  $Q_k$  为第  $k$  条染色体所有分叉点的集合,  $T_q$  为距离阈值,  $T_a$  为染色体边界弧长阈值。

### 2.3.2 中轴无分叉情况

样本图像中可能会有染色体端部粘连或首尾粘连的情况,在重度粘连时所提取的中轴往往没有分叉点,对此提出了新的方法。

#### 情况 1 端部粘连

端部粘连的染色体如图 6(a)所示,它的粘连处的宽度几乎与染色体等宽。将其细化后,得到的中轴并没有分叉,如图 6(b)所示。将这类染色体分开可以做两个方面的测试判断。



(a)端部粘连的染色体原图 (b)轮廓中轴图 (c)轮廓图  
图6 端部粘连的染色体

(1)在候选分割点中,寻找是否存在这么一个候选分割点,它的两端附近的边界有凸点出现,如图 6(c)的圆圈标记所示,凸点的出现是因为染色体端头本身形态所决定。此时凸点的链码差将会达到正值极大。因此在该条染色体链码差数组中,当前候选分割点的前  $T_p$  个点和后  $T_p$  个点中,只要有一点的链码差值大于 1,那么该候选分割点满足分割条件。实验中将距离阈值  $T_p$  设为该染色体周长的  $1/9$ 。在确定第一个分割点的情况下,再在候选分割点中寻找一个与该分割点距离最短的点

作为另外一个分割点。表示如下:

$$D_k = \{d | DiffCode[d] > T\}$$

$$(\exists x)(x \in R_k)$$

$$S_r = \{x | (\exists x_{d1})(\exists x_{d2})(x_{d1} \in D_k) \wedge (x_{d2} \in D_k) \wedge$$

$$Arc < x_{d1}, x > < T_p \wedge Arc < x, x_{d2} > > T_p\}$$

$$S_r \rightarrow S = \{(x, y) | (x \in S_r) \wedge (y \in R_k) \wedge \min\{L(x, y), Arc(x, y)\} > T_a\} \quad (10)$$

其中,  $D_k$  为第  $k$  条染色体的凸点集合,  $S_r$  为位于两个凸点间的候选分割点集,  $S$  为最终确立的分割点集。  $Arc < \cdot >$  为从左至右顺序计算两点间的边界弧长。  $T_p$  和  $T_a$  分别为不同的弧长阈值。

(2)这类粘连引起染色体长度的增加,作为配对染色体,如果其长度与剩下染色体中最长的那条有明显区别,那么就能判断处它是有染色体的端部粘连形成。至于在何处分割,鉴于此时染色体仅仅是粘连,而并非交叠,所以可能会出现在粘连区域的灰度值要比染色体平均灰度值小。所以在候选分割点中选取两个点的间距不超过染色体宽度的 2 倍,且两点连线上大部分的点的灰度值小于平均灰度值。

#### 情况 2 首尾粘连

首尾粘连的染色体如图 7(a)所示,其细化后得到的中轴同样没有分叉。如图 7(b)所示。它的处理区别与端部粘连染色体分割过程在于,要先找到间距最短的两个候选分割点,每个分割点在边界上相同方向(同是顺时针方向或同是逆时针方向)的一定阈值距离  $T_s$  内,有凸点存在,如图 7(c)的圆圈标记所示,才进一步认定该两个候选分割点满足分割条件。另外与端部染色体粘连情况相似,它也需要进行长度和弧长阈值判断。



(a)首尾相连染色体原图 (b)轮廓中轴图 (c)轮廓图  
图7 首尾相连染色体

## 3 染色体交叠分割

对于最常出现的“十字型”交叠,由于染色体在交叠处同样有凹角点存在,分割点的数目比粘连分割要多,但是寻找候选分割点的方法以及分割路径与重度粘连分割一致,要运用候选分割点间的距离阈值和弧长阈值进行筛选判断。与粘连分割不同的是,交叠分割还必须要利用每两个相邻的候选分割点之间都必须存在一个凸点的这么一个特征,这种特征也是由染色体的长条状决定的。

## 4 结论

应用上述方法,对 50 幅染色体图像,2 300 条左右的染色体进行分割实验,结果表明,对于轻度粘连的染色体分割,成功率达 90.3%,对于一般重度粘连和交叠的染色体成功率达 88.6%,对于判断中轴无分叉情况下的重度粘连,成功率可达到 75%以上。

所以结果表明,所归类的算法提高了染色体分割自动化的程度和正确率。同时有效处理了端部或首尾相粘连的情况,避免了一些遗漏分割发生,为后面的染色体特征提取和识别做出很好的准备。

(下转 187 页)