

一种弯曲染色体自动矫直算法

doi:10.3969/j.issn.1674-7100.2018.06.010

吕丹丹 周平

浙江理工大学

信息学院

浙江 杭州 310018

摘要:染色体的非刚性使其在成像时出现不可预测的形状和大小,将其直接进行核型分析会造成较大误差,而对染色体进行矫直,可提高核型分析的准确性。为解决弯曲染色体矫直的问题,提出了一种有效的矫直算法。该算法使用细化的方法提取弯曲染色体的中轴并对其进行修正,通过记录与中轴上点相邻的8个像素点,获得拟合直线及其中垂线,再将中轴按照像素点拉伸,拉伸过程中根据重叠区域像素的相对空间位置,将重叠区域的像素搬运到矫直图像上得到矫直染色体。对比3种矫直算法可知本算法可以较好地实现弯曲染色体的矫直。

关键词:矫直算法;核型分析;中轴;像素搬运;直染色体

中图分类号: TP391.41

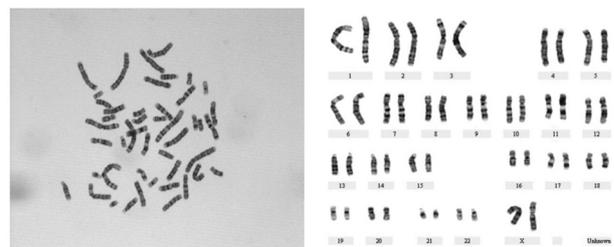
文献标志码: A

文章编号: 1674-7100(2018)06-0067-07

0 引言

染色体是细胞核中载有遗传信息的物质,正常人都拥有23对染色体,22对为常染色体,1对为性染色体(男性为XY染色体,女性为XX染色体),任何染色体数目或结构上的改变都会出现因遗传缺陷而导致的疾病。细胞遗传学专家在细胞分裂中期(细胞分裂分为前期、中期、后期和末期4个阶段),可通过光学显微镜拍摄染色体图像^[1]。细胞分裂中期的染色体浓缩变粗,显示出该物种所特有的数目和形态,此时专业操作员通过视觉分析图像,对染色体进行排列与配对,制作出核型图,再根据专业知识识别诊断染色体的异常情况。这个过程很繁琐且耗时,而且人为因素也可能影响判断的准确性,因此实现自动化的核型分析系统成为必然趋势。现有的自动核型分析系统大部分是人机交互型,存在处理效果不佳的缺点,开发完全自动化的核型分析系统仍然需求迫切。

核型分析系统主要包括分割和分类两个阶段,常根据染色体的几何特征和条带特征进行分析^[2]。由于染色体的非刚性特性,致使它们在成像时出现了不可预测的形状^[3],例如在图1中出现的不同弯曲程度的染色体。为了更好地提取染色体并研究其结构特征,提高分类准确率,揭示结构异常,需将弯曲的染色体进行矫直处理。



a) 显微镜下染色体图

b) 手动核型分析图

图1 中期G带染色体图与手动核型分析图

Fig. 1 The mid-G band image under the microscope and Karyotype analysis

收稿日期: 2018-09-11

作者简介: 吕丹丹(1991-),女,山东临沂人,浙江理工大学硕士生,主要研究方向为计算机视觉与模式识别,
E-mail: 1414665215@qq.com

通信作者: 周平(1960-),男,浙江杭州人,浙江理工大学教授,主要从事计算机视觉方面的教学与研究,
E-mail: 1500064071@qq.com

许多学者针对自动矫直弯曲染色体的问题做了研究。D. Somasundaram 等^[4]使用 Stentiford 细化算法 (stentiford thinning algorithm, STM) 获取染色体中轴^[5], 再使用投影矫直算法 (projective straightening algorithm, PSA) 矫正中轴, 最后将染色体图像投影到行矩阵上, 得到矫直的染色体图像。M. J. Roshtkhari 等^[6]分析和计算在各种旋转角度下获得的染色体二值图像的水平 and 垂直投影矢量角, 通过最小化旋转分数确定染色体弯曲轴和弯曲中心, 分离染色体臂, 对齐连接得到矫直的染色体。S. Khan^[7]等使用种子点作为主要预测, 以预测点与边界点的距离、边界上指定点处的梯度作为辅助预测, 使用辅助预测修正主要预测得到中轴, 进一步使用几何校正弯曲的染色体。A. Tanvi^[8]等利用线性插值方法对染色体进行了重建。S. Jahani 等^[9]利用传统细化算法提取中轴, 对其两端按照端点附近 5 个像素的斜率方向延伸 30 个像素, 从而得到修正后的内侧轴, 最后将染色体图像进行投影得到直染色体。这些研究虽然能将弯曲的染色体自动矫直, 但存在矫直效果不佳的缺点, 且大多数仅局限于略微弯曲的染色体, 对于高度弯曲或多重弯曲的染色体上述算法还需优化^[10]。

针对上述问题, 本文主要提出了一种适用于任意弯曲程度的染色体的全自动矫直算法。该算法主要通过最大类间差法对弯曲染色体图像进行图像二值化, 若出现孔洞则使用漫水填充法进行填充, 再使用改进的细化算法获得二值图像的单像素骨架, 去除处理过程中出现的分支 (毛刺) 后, 使用最小二乘法进行直线拟合, 沿着其方向延长缺失的像素得到完整的中轴线, 再将中轴上所有的像素点进行映射形成垂直方向上的直线段 (拉直的中轴), 再在二值图像中获取到中轴线及其所有垂线段坐标信息, 并在原始灰度图像中获取相应的灰度值, 再将原始图像中轴线上所有垂线上的像素点的灰度值, 以拉直后的中轴线为参考线进行映射, 最后生成垂直方向的染色体, 即矫直染色体。通过该算法, 以期实现自动矫直任意弯曲程度的染色体, 提高自动核型分析的准确性。

1 算法介绍

1.1 算法流程

本文提出的弯曲染色体矫直算法主要由 2 部分组成。1) 完整弯曲染色体中轴的提取; 2) 弯曲染色体

的矫直。算法流程如图 2 所示。

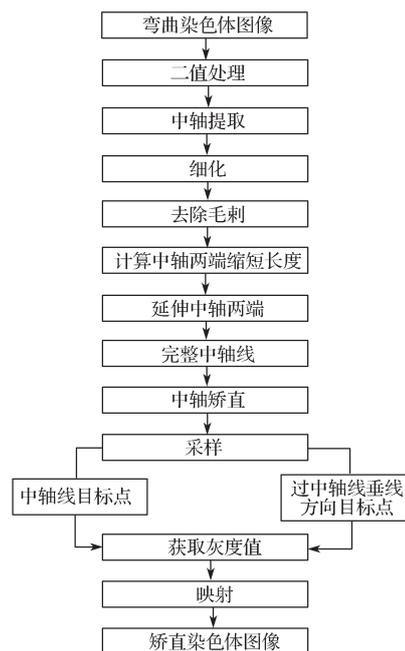


图 2 弯曲染色体矫直算法流程图

Fig. 2 Algorithm flow of curved chromosome straightening

1.2 图像二值化

二值图像上每一个像素只有两种可能的取值或灰度等级状态, 相比原始的灰度图像更容易识别, 而且具有内存要求小、执行时间快等优点^[11-13]。因此, 制作染色体的二值图像是将弯曲染色体拉直所必须进行的步骤。通过对灰度图像进行阈值处理, 将图像分成背景和前景两部分, 从而获得二值图像。而有些染色体中期图像经常受到灰度反差的影响, 导致使用最大类间算法进行二值处理时会出现内部封闭孔洞的现象, 本文采用漫水填充的方法来填充封闭的孔洞。步骤为: 1) 用白色填充整个背景区域, 得到的封闭孔洞全为黑色; 2) 反转得到白色的孔洞; 3) 与原二值图像相加得到填充后的图像, 此时内部的封闭孔洞被有效填充。孔洞填充过程如图 3 所示。



图 3 孔洞填充效果图

Fig. 3 Hole filling effect diagram

1.3 中轴提取

图像的中轴也称为骨架, 是描述图像几何及拓扑性质的重要特征之一。在染色体矫直中, 准确找到弯曲染色体的中轴是一个非常关键的因素。本文先对二值图像进行细化处理, 再去除骨架中出现的毛刺, 最后通过计算得到中轴两端近似缩短长度进行两端延伸操作, 最终较好地完成了弯曲染色体完整中轴线的提取, 这为其矫直处理创造了良好条件。

1.3.1 细化

细化算法^[14-17]是在二值图像基础之上进行“层层剥夺”的方法, 即从边缘向内部一层一层删除符合条件的点, 直到线条剩下一个像素为止。这里, 我们将像素值为1的值设为目标点, 像素值为0的值设为背景点, 定义一个像素点为 P_1 , 它的八邻域点 $P_2 \rightarrow P_9$ 位置如图3所示, 边界点即为其八邻域中至少有一个背景点的目标点。

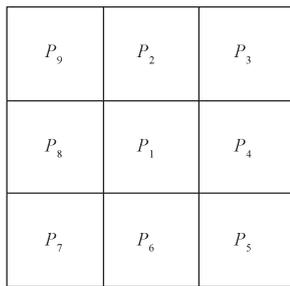


图3 P_1 的八邻域

Fig. 3 Eight neighborhoods of P_1

细化算法的具体步骤如下:

1) 扫描目标点, 如果满足 $2 \leq N(P_1) \leq 6$ ($N(P_1)$ 表示 P_1 的八邻域中目标点的个数), $S(P_1)=1(S(P_1)$ 表示 $P_2 \rightarrow P_9$ 这些点中值出现 0 到 1 变化的次数), $P_2 \times P_4 \times P_6=0$ 、 $P_4 \times P_6 \times P_8=0$ 或 $P_2 \times P_4 \times P_8=0$ 、 $P_2 \times P_6 \times P_8=0$, 则将 P_1 标记为删除;

2) 重复以上步骤, 直到没有像素被标记为删除, 则获得二值图像的细化中轴。

细化效果如图4所示。



图4 细化效果图

Fig. 4 Thinning renderings

1.3.2 去除毛刺

由细化效果图可知, 图中出现了中轴线分支偏移与毛刺问题, 这干扰了染色体特征的提取。因此, 本课题组将毛刺去除。去除毛刺的算法描述如下:

1) 在只有细化中轴线的二值图中, 从图像左上角开始遍历获得每一个目标像素点, 即中轴线的每一个点 P ;

2) 使用启发式搜索方法对 P 搜索下一个目标点, 如果搜索到的目标点个数不低于 1, 表明此点 P 为分支点;

3) 沿着 P 点分支方向继续搜索并对其目标点进行计数操作, 如果目标点的计数值低于毛刺的门限值 (本文设置门限值为 5), 就去除此毛刺。

去除毛刺后与染色体原始图像叠加的效果图如图5所示。由图可知, 中轴线保留了最初的完整形状。

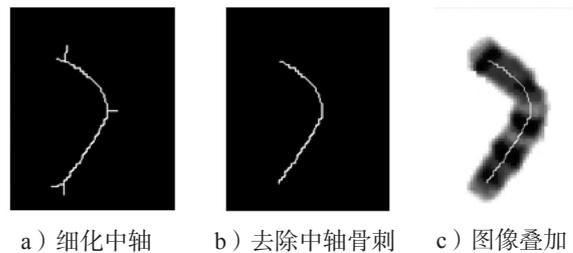


图5 毛刺去除对比图

Fig. 5 Burr removal comparison chart

1.3.3 中轴修正

通过以上处理所得到的中轴线只是染色体中间的一部分, 其两端会有一些像素点缺失。中轴修正主要是为了解决中轴提取过程中造成的纵向缩短问题, 本文通过计算中轴的实际缩短长度, 再根据中轴两端的最佳拟合直线方向对其进行延长, 从而获得完整的中轴线。

中轴提取过程中使用的方法是一种类似腐蚀的操作。因此, 假设染色体图像在各个方向上的缩短距离都是相等的, 通过计算各个方向上缩短的平均长度作为中轴提取过程中的纵向缩短长度。长度计算公式如下:

$$D = \frac{\sum_{i=0}^N d_i}{2N} \quad (1)$$

式中: N 为中轴上像素个数; d_i 为中轴第 i 个像素处缩短总距离。

长度计算示意图如图6所示。

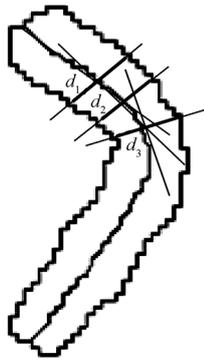


图6 长度计算示意图

Fig. 6 Schematic diagram of length calculation

中轴提取的最终结果如图7所示。



a) 二值图像

b) 完整中轴

图7 中轴提取效果

Fig. 7 Axis extraction

1.4 中轴矫直

在二值图像中沿弯曲中轴线依次遍历每一个白色像素点获取坐标位置并进行计数(即弯曲染色体的长度),然后在原始灰度图像中进行采样得到像素值,最后在新建的空白图像水平方向的中间位置,从上往下依次复制像素值,从而得到矫直后的中轴线。弯曲中轴线被矫直示意图如图8所示。图中标注的194个像素点是指完整中轴线的像素点数,即为染色体的实际长度。

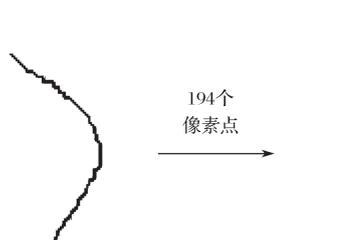


图8 中轴矫直

Fig. 8 Axis straightening

1.5 染色体矫直

将染色体中轴的像素叠加放在一条直线上完成染色体拉伸过程,根据相对位置关系,补全染色体图像其余部分。补全过程中逐个访问中轴上的点,

并将与其相邻的8个点放在一起计算最佳拟合直线。这里的相邻并不意味着访问点是9个点的中心,当访问点的位置小于4时,直接选取中轴开端的9个点即可。访问到中轴结尾时道理相同,只有在中间部分时访问点才是9个点的中心。9个连续的散点选取后,按照公式(2)进行最佳拟合直线,计算如下^[18-19]:

$$a = \frac{\sum xy - \frac{1}{N} \sum x \sum y}{\sum x^2 - \frac{1}{N} (\sum x)^2}, \quad (2)$$

$$b = \bar{y} - a\bar{x}. \quad (3)$$

式中, N 为散点个数。

可得 N 个散点的最佳拟合直线方程为:

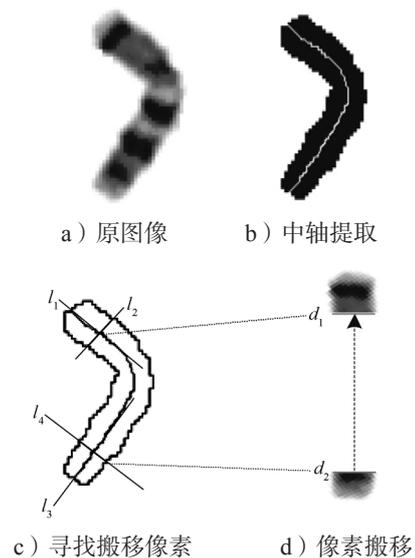
$$y_i = ax + b_0.$$

假设拟合直线的垂线方程为:

$$y = -\frac{1}{a}x + c, \quad (4)$$

式中 c 为常数, c 值的求解只需将当前访问点带入式(4)即可。

根据最佳拟合直线方程,可得到8条连续中轴散点的直线,再由拟合直线的垂线方程得到垂线,进而通过平移的方式获得每个中轴散点的垂线,最后将染色体和垂线重合的部分通过像素搬移的方式搬运到矫直图像中,即可得到矫直的染色体,整个矫直的过程示意图如图9所示。图中 l_1 、 l_3 为两条最佳拟合直线, l_2 、 l_4 为拟合直线上过散点的垂线, d_1 、 d_2 为与染色体重叠部分的垂线段。



a) 原图像

b) 中轴提取

 l_1 l_2 d_1 l_3 l_4 d_2 l_1 l_2 d_1 l_3 l_4 d_2

c) 寻找搬移像素

d) 像素搬移

图9 染色体矫直过程

Fig. 9 chromosome straightening process

由图 9 可知, 染色体矫直就是先在二值图像中扫描搜索得到中轴线的起始点, 并沿中轴线方向跟踪得到每一个像素点, 再沿着此像素点的垂直方向获得所有目标点及其坐标位置, 然后在原始图像中, 从中轴线上紧密定位的点向两侧同步进行采样, 以拉直后的中轴线为参考线向两侧同步映射, 每一条垂直线段上的点都会被映射成为矩阵中累计的一行, 而中轴线上所有的垂直线段经采样映射后就得到了垂直方向上的染色体, 即为矫直后的染色体。最终效果如图 10 所示。

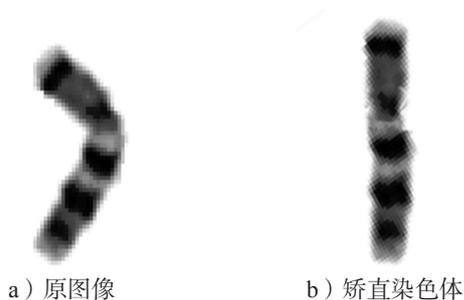


图 10 染色体矫直

Fig. 10 Chromosome straightening

2 实验与结果分析

2.1 数据集

本文所使用的染色体图像, 均由浙江省人民医院检验中心提供的。采用人外周血细胞进行培养后制片, 并使用 BX-51 型光学显微镜对细胞分裂中期的 550 条染色体 G 显带进行拍摄, 得到具有 764×575 像素的 400 张灰度图像, 再从中分割提取弯曲染色单体图像。

2.2 算法比较

2.2.1 SPV 算法

投影矢量矫直 (straightening via projection vector, SPV)^[20] 算法将旋转图像水平投影中主峰的相对幅度定义为旋转分数 S , 通过旋转分数确定染色体图像中的弯曲中心, 然后应用于人工矫直弯曲的染色体图像中。当弯曲中心应用于高度弯曲的染色体的实际图像时, SPV 算法可以矫正数据集内的大部分染色体图像。

2.2.2 SMAC 算法

内侧轴提取和众包矫正 (straightening via medial axis extraction and crowdsourcing, SMAC)^[21] 算法是在染色体的二值图像上, 通过细化程序提取染色体中轴 (medial axis, MA), 再比较端点的欧几里德距

离和 MA 的长度, 识别出弯曲的染色体。对于染色体矫直, 需考虑 MA 的斜率, 通过延伸两端来修改最初提取的中轴。输入原图像时, 在 MA 紧密定位的垂直线上进行采样, 将采样映射拉伸的染色体上, 产生垂直取向的直染色体。

2.2.3 实验结果对比

弯曲染色体的矫直对染色体分类具有重要的作用, 可提高核型分析的准确率。本文通过中轴提取、二值图像像素坐标获取、原图采样及双向映射提出了一种简单而有效的染色体矫直算法。为验证本文算法的有效性与泛化性, 采用 SPV 算法、SMAC 算法、本文算法对弯曲染色体图像进行测试。表 1 展示了 SPV 算法、SMAC 算法以及本文中提出的算法矫直弯曲染色体的效果图。

表 1 3 种矫直方法的对比

Table 1 Comparison of three straightening methods

	弯曲染色体序号				
	1	2	3	4	5
矫直算法					
SPV 算法					
SMAC 算法					
本文算法					

由表 1 可知, 在矫直 5 种不同弯曲程度的染色体时, 本文算法均优于 SPV 算法和 SMAC 算法。其原因主要为:

1) SPV 算法需要将弯曲中心应用到手动辅助标定的染色体中轴上, 若弯曲程度过大, 弯曲中心与辅助标定位置会出现较大差异, 因此 SPV 算法无法解决高度弯曲位置的矫直问题, 如 3 号染色体经 SPV 算法矫直后, 仍有不同程度的弯曲。

2) SMAC 算法是通过“剥离”的思想计算中轴, 但直接将中轴进行染色体矫直, 缺少了中轴毛刺去除的过程, 导致染色体中轴在有毛刺的位置出现了较大程度的变形。

3) 本文算法则在提取中轴后增加了毛刺去除的步骤,同时使用像素点及其8个邻接点进行直线拟合,以减少每一次像素搬移点的选取误差,从而可以有效地将高度弯曲的染色体进行矫直。

3 结语

针对不同弯曲程度染色体的矫直问题,本文提出了新的矫直算法。对比分析3种染色体矫直算法的效果图,表明本文算法可以有效地处理不同弯曲程度的染色体。但是本文算法在中轴提取两端缺失像素的修复模块时存在一定的误差,从而使拉直染色体再重建时也会存在一定的误差,这需要在未来的工作中进一步优化和完善,以期能提取更精确的带纹特征,提高核型分析准确率,实现自动化染色体异常诊断。

参考文献:

- [1] MOHAMMADI M R. Accurate Localization of Chromosome Centromere Based on Concave Points[J]. *Journal of Medical Signals & Sensors*, 2012, 2(2): 88-94.
- [2] ARORA T, DHIR R. An Automatic Human Chromosome Metaspread Image Selection Technique[J]. *Knowledge & Information Systems*, 2017, 52(3): 1-18.
- [3] JAHANI S, SETAREHDAN S K. Centromere and Length Detection in Artificially Straightened Highly Curved Human Chromosomes[J]. *International Journal of Biological Engineering*, 2012, 2(5): 56-61.
- [4] SOMASUNDARAM D, VIJAY KUMAR V R. Straightening of Highly Curved Human Chromosome for Cytogenetic Analysis[J]. *Measurement*, 2014, 47(1): 880-892.
- [5] JANSI S, SUBASHINI P. Optimal Thinning Algorithm for Detection of FCD in MRI Images[J]. *International Journal of Scientific & Engineering Research*, 2013, 2(9): 125-133.
- [6] ROSHTKHARI M J, SETAREHDAN S K. A Novel Algorithm for Straightening Highly Curved Images of Human Chromosome[J]. *Pattern Recognition Letters*, 2008, 29(9): 1208-1217.
- [7] KHAN S, DSOUZA A, SANCHES J, et al. Geometric Correction of Deformed Chromosomes for Automatic Karyotyping[C]//2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. San Diego: IEEE, 2012: 4438-4441.
- [8] TANVI A, RENU D, MANISH M. An Algorithm to Straighten the Bent Human Chromosomes[J]. *2017 Fourth International Conference on Image Information Processing(ICIIP)*, 2017(12): 522-527.
- [9] JAHANI S, SETAREHDAN S K. An Automatic Algorithm for Identification and Straightening Images of Curved Human Chromosomes[J]. *Biomed Eng Appl Basis Commun*, 2012, 24(6): 503-511.
- [10] SONKA M, HLAVAC V, BOYLE R. *Image Processing, Analysis, and Machine Vision*[M]. 2nd ed. New York: Springer Science+Business Media B.V., 1993: 442-444.
- [11] 何颖, 张星阳, 赵金龙. 基于区域生长分割算法在医学图像中的研究[J]. *数字技术与应用*, 2017(10): 115-117.
HE Ying, ZHANG Xingyang, ZHAO Jinlong. Research on Medical Image Based on Region Growing Segmentation Algorithm[J]. *Digital Technology and Application*, 2017(10): 115-117.
- [12] 顾秀华, 王持, 周平. 基于图像识别的薄板自动计数算法[J]. *包装学报*, 2017, 9(5): 56-61.
GU Xiuhua, WANG Chi, ZHOU Ping. Research on Automatic Counting of Sheet Number Based on Image Recognition[J]. *Packaging Journal*, 2017, 9(5): 56-61.
- [13] 卢尊兰. 基于图像分析的色偏检测算法[J]. *包装学报*, 2017, 9(1): 40-45.
HU Zunlan. Color Cast Detection Method Based on Image Analysis[J]. *Packaging Journal*, 2017, 9(1): 40-45.
- [14] ZHANG T Y, SUEN C Y. A Fast Parallel Algorithm for Thinning Digital Patterns[J]. *Communications of the ACM*, 1984, 27(3): 236-239.
- [15] 牟少敏, 杜海洋, 苏平, 等. 一种改进的快速并行细化算法[J]. *微电子学与计算机*, 2013, 30(1): 53-55.
MOU Shaomin, DU Haiyang, SU Ping, et al. A New Improved Fast Parallel Thinning Algorithm[J]. *Microelectronics & Computer*, 2013, 30(1): 53-55.
- [16] 葛杰, 曹晨晨, 李光. 基于机器视觉的图像形状特征提取方法研究进展[J]. *包装学报*, 2015, 7(1): 54-60.
GE Jie, CAO Chenchen, LI Guang. Research Progress in Shape Feature Extraction Methods Based on Machine Vision[J]. *Packaging Journal*, 2015, 7(1): 54-60.
- [17] 汪祖辉, 孙刘杰, 邵雪. 一种小波域的图像增强算法[J]. *包装学报*, 2017, 9(1): 21-27.
WANG Zuhui, SUN-LIU Jie, SHAO Xue. An Image Enhancement Algorithm Based on Wavelet Domain[J]. *Packaging Journal*, 2017, 9(1): 21-27.

- [18] SAEED K, TABĘDZKI M, RYBNIK M, et al. K₃M: A Universal Algorithm for Image Skeletonization and a Review of Thinning Techniques[J]. International Journal of Applied Mathematics & Computer Science, 2010, 20(2): 317-335.
- [19] 韩庆瑶, 肖 强, 乐 英. 空间离散点最小二乘法分段直线拟合的研究 [J]. 工业仪表与自动化装置, 2012(4): 107-109.
HAN Qingyao, XIAO Qiang, YUE Ying. Study of Space Discrete Point's Piecewise Linear Fitting on Least Square Method[J]. Industrial Instrumentation & Automation, 2012(4): 107-109.
- [20] JAHANI S, SETAREHDAN S K. A Novel Method for Centromere and Length Detection in Microscopic Images of Human Chromosomes[C]//18th Iranian Conference on BioMedical Engineering. Tehran: IEEE, 2011: 274-277.
- [21] SWATI, GUPTA G, YADAV M, et al. Siamese Networks for Chromosome Classification[C]//IEEE International Conference on Computer Vision Workshop. Venice: IEEE, 2017: 72-81.

Research on an Automatic Straightening Algorithm for Curved Chromosomes

LÜ Dandan, ZHOU Ping

(School of Informatica Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: The non-rigidity of the chromosome enables it to have unpredictable shape and size at the time of imaging. Directly performing chromosome classification and pairing would cause large errors and low efficiency, while the operation of the straightened chromosome could not only improve the accuracy and efficiency of classification and pairing, but also be conducive to the diagnosis of chromosomal abnormalities. In order to effectively solve the problem of curved chromosome straightening, an effective chromosome straightening algorithm was proposed, which could be used to straighten chromosomes with different degrees of curvature. The algorithm mainly used the refinement method to extract the central axis of the curved chromosome and correct it. By recording the adjacent 8 pixel points of the point on the central axis, the overlapping area of the vertical line of the fitted line and the original image was obtained, and then the central axis was followed. The pixel was stretched, and the pixels of the overlapping area were moved to the corrected chromosome picture according to the relative spatial position of the pixels in the overlapping area during the stretching process. By comparing the renderings of the three straightening algorithms, the method proved better realization of the straightening of curved chromosomes.

Keywords: straightening algorithm; chromosome classification; medial axis; pixel shift; straight chromosome