

# 基于 PCR 凝胶成像的信息自动提取

朱虹 许朝晖 周健

(西安理工大学信息与自动化工程学院, 西安 710048)

**摘要** 在 PCR 凝胶成像及定量分析系统中,为了正确检测出 PCR 产物,以便于对被检测者的病理状况作出准确的分析和处理,介绍了一种多聚酶链反应(PCR)成像信息的自动提取方法。该方法首先用高斯平滑滤波和灰度信息进行水平方向的投影,以抵抗光照不均的影响,得到 PCR 扩增时所在的泳道;之后对每个泳道信息进行高斯平滑滤波和灰度的垂直投影,以抵抗 PCR 条带信息的强弱不均的影响和移动模糊的影响,准确地抽取出一条带区域;最后,通过局部校正来提高所圈定条带区域的精确度。实验结果表明,此方法可正确检测出 PCR 产物,对后面的分析处理具有十分重要的实际意义。

**关键词** 多聚酶链成像 光照不均 移动模糊 灰度投影 高斯平滑滤波器

**中图分类号**: TP391.4 Q503-39 Q-334 **文献标识码**: A **文章编号**: 1006-8961(2002)12-1285-06

## Auto-extraction Techniques of Information Based on PCR Gelation Image

ZHU Hong, XU Zhao-hui, ZHOU Jian

(School of Automation and Information Engineering, Xi'an University of Technology, Xi'an 710048)

**Abstract** In the quantitative analysis of PCR gelation images system, in order to detect PCR correctly and make more accurate diagnosis of pathological situation of the person detected, a method to extract information automatically from PCR images is presented in this paper. Firstly, Gaussian filter and horizontal projection of PCR images' gray information are used to resist the effect of non-uniform distribution of brightness and gain the position of the pervasive PCR. Here, a key step is to select a adaptable parameter value for the Gaussian filter. Then, the method uses smooth filtering and vertical projection to reduce the influence caused by the non-uniform and moving illegibility of the PCR bands, and then extract PCR band areas accurately. Finally, it makes use of local rectification to improve the precision of the enclosed band. In practice, this method is applied to the PCR image processing and promising results are obtained. The experiment result indicates that with the PCR image been detected accurately, subsequent diagnosis and analysis can achieve fairly efficiency, so the method is of very significant practicality.

**Keywords** PCR gelation image, Non-uniform distribution of brightness, Moving illegibility, Gray projection, Gaussian filter

## 0 引言

基因,作为基本的遗传单位,决定着一个人的人体生理特征和一些行为特征<sup>[1~3]</sup>。只要人体基因中的30亿分之一出了问题,就会导致疾病。因此,基因

图象的自动诊断技术给医学的准确诊断提供了一种可视化的手段。

多聚酶链反应(PCR)技术现已成为分子生物学研究的重要工具,它通过一对互补于靶序列两侧的核苷酸引物,特异地扩增出特定的DNA片段,可测出几个分子的核酸,由引物决定的对靶序列的高度

基金项目:陕西省自然科学基金专项项目(00JK264)

收稿日期:2002-01-08; 改回日期:2002-05-08

特异性又可避免样品中其他核酸分子产生错误信号.对于 PCR 凝胶成像及定量分析系统来说,只有先对图象进行分析提取,正确检测出 PCR 产物,才可以对被检测者的病理状况作出准确的分析和处理<sup>[4,5]</sup>.

在凝胶成像的胶片中,PCR 产物的形状呈条带状,在各自的泳道中顺序排列.如果能自动地、准确地将条带圈定,则可用于后续的定量分析.然而,硬件设备采集的基因样本图片随机性太大,给图象处理带来很大的难度.例如,因为采集样本时的光照不均匀造成局部画面的过亮或过暗;采集样本时,角度的随机性造成了基因泳道水平方向以及垂直方向的偏差;样本图象拍摄时,PCR 扩增带来的移动模糊;还有由于光照不均和采样角度随机性共同造成的一幅图象中的个别基因不清,难以识别等等.针对这些问题,提出了一种比较好的、可以抵抗信息强弱不均和移动模糊的方法.

## 1 PCR 条带信息的自动提取算法

### 1.1 基于灰度投影及高斯平滑阈值处理的信息增强算法

如图 1 所示,在 PCR 凝胶图象中,存在着光照不均(PCR 条带有强也有弱)以及拖尾(PCR 条带间滑动的痕迹)等现象,这就给准确提取条带信息带来了一定的困难.当然,图象的区域分割作为自动目标物识别的关键技术,吸引了众多学者进行了广泛的研究,并取得了许多的成果<sup>[6,7]</sup>.一般情况下,对于边界不清晰的图象,图象分割算法都比较复杂<sup>[8,9]</sup>.因此,针对 PCR 凝胶图象的特点,提出了一种简单的,但却可以准确进行条带分割的算法.

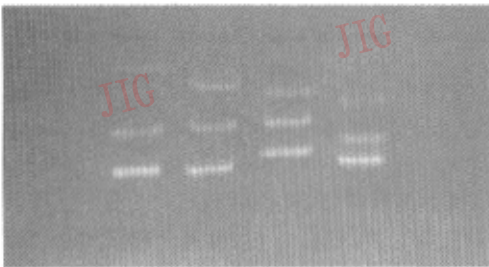


图 1 PCR 凝胶图象

PCR 凝胶图象在各自泳道中的扩增特点是在胶片中呈水平排列的较亮的条带状.因为条带的亮

度信息的分布在图象中是不均匀的,所以不是直接对原图进行二值化处理,而是通过对原图进行高斯平滑以及进行灰度投影来完成条带信息的提取.首先,对原图进行水平方向上的灰度投影.图 2 是图 1 在水平方向上的投影曲线.从这条曲线可以明显看出 4 个泳道在曲线上呈 4 个脉冲峰.同时也可以看出画面的左侧比右侧暗(肉眼直接从原图中不容易看出).如果直接以全局阈值进行分割,则会出现泳道过宽或泳道过窄的现象.

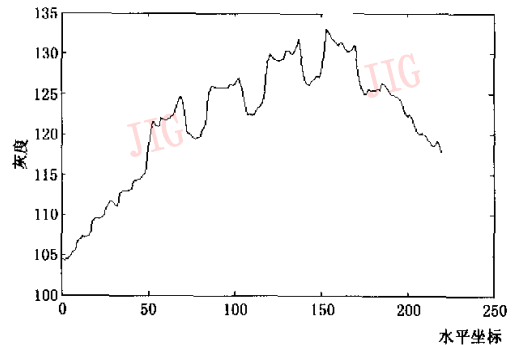


图 2 水平方向的灰度投影曲线

为此,对原图进行高斯平滑滤波.设原图为  $f(x, y)$ ,经过平滑滤波之后的图象为  $h(x, y)$ ,则

$$h(x, y) = g(x, y) * f(x, y)$$

$$g(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{\sigma^2}}$$

其中,“\*”为卷积运算, $\sigma$ 为平滑作用区域的宽度,是决定分割泳道和 PCR 条带信息准确度的参数.

### 1.2 平滑参数 $\sigma$ 的选取方法

参数  $\sigma$  采用迭代优化的方法来确定,具体步骤如下:

- (1) 设定  $\sigma$  的初始值为一个较小的值.
- (2) 构造相应的高斯函数对原图象进行平滑处理.
- (3) 分别做原图象和平滑图象的灰度投影(对泳道作水平投影,对条带作垂直投影).
- (4) 用平滑图象的投影曲线作为原图象投影曲线的阈值,得到一个二值脉冲曲线(其值只能为 1 或 0).
- (5) 求出所有值为 1 的脉冲宽度的均值和方差,如果方差太大,则  $\sigma$  参数不合适,令  $\sigma' = \sigma + \Delta\sigma$ ,返到第 2 步,否则,就得到了合适的脉冲宽度.
- (6) 在原图象中以此脉冲宽度提取相应的泳道和条带信息.

图 3 给出了选择不同参数  $\sigma$  的效果示意图.

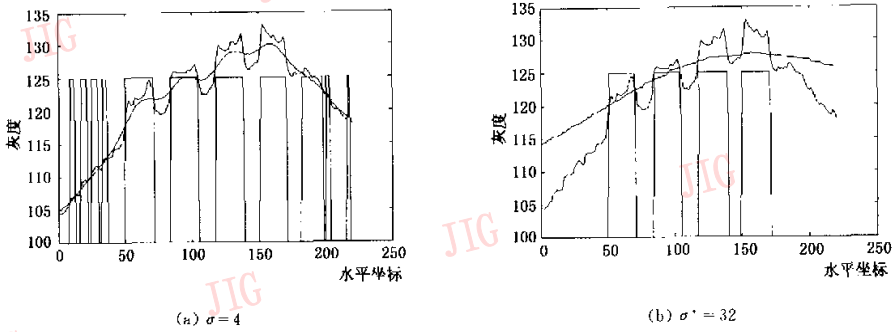


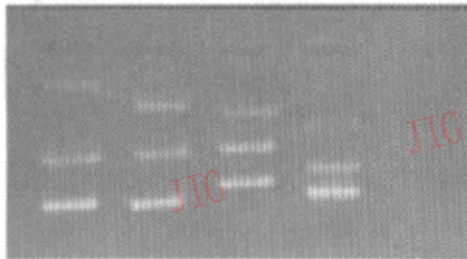
图 3 选择参数  $\sigma$  的效果示意图

## 2 实验结果与分析

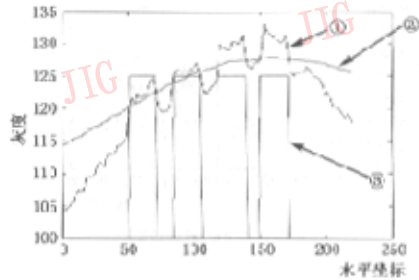
### 2.1 泳道的提取

泳道的提取解决的主要是光照不均的问题,用上节所介绍的方法可以准确地将泳道分离出来.如图 4(a)所示,原图中的 PCR 条带较背景更亮,但是画面存在光照不均,由图 4(b)中的水平投影曲线①可知,画面左侧比右侧暗.虽然肉眼不一定有察觉,

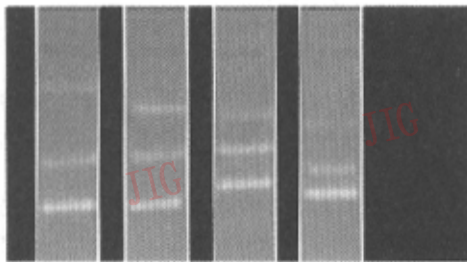
但是用全局阈值方法进行处理时,就可能会出现泳道的误提取(如图 4(d)).因此,在这里先将原图象进行高斯模糊滤波,这个模糊图象仍旧可以反映原图象画面的大致光照趋势,所以将其作为阈值选取的依据时,对光照不均具有自适应能力;之后分别对其进行水平方向的投影,得到图 4(b)中的曲线②;将曲线②作为曲线①的阈值线,则可得到曲线③;曲线③所呈现的脉冲所表示的就是泳道宽度.由此,完成 PCR 泳道的提取,结果如图 4(c)所示.



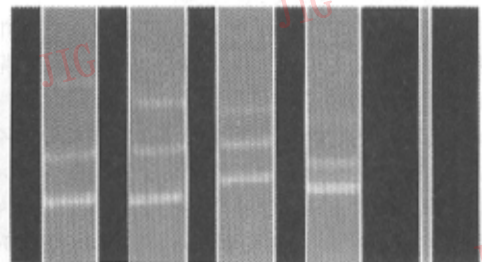
(a) 原图



(b) 水平方向的投影



(c) 本方法提取出的泳道



(d) 全局阈值方法处理的最好结果

图 4 自动提取 PCR 所在的泳道

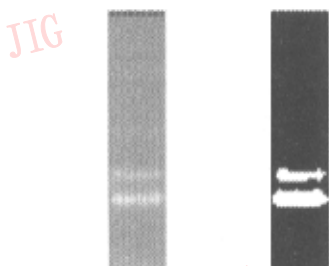
### 2.2 PCR 条带的自动圈定

2.2.1 同一泳道中 PCR 条带信息不同强弱的处理  
PCR 条带所在的泳道被确定出来后,还要从各

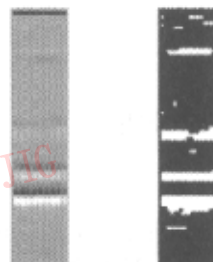
自的泳道中,将条带信息自动识别出来,以便对其进行定量分析.但是,如前所述,由于光照环境等因素的影响,会出现在同一个泳道中的 PCR 条带信息

的强度不同.如图4(c)所示,在最右边的泳道中的最上面的一个条带信息很弱,如果用全局阈值方法对其进行二值化处理的话,就会出现如图5所示的结果,图5(a)二值图象中只有两个信息较强的条带区域得到增强(白色部分),亮度信息较弱的条带信息则无法得到.考虑到条带是以水平方向排列在泳道中的事实,对条带中的灰度信息进行水平方向锐化处理,然后再对锐化后的图象进行二值化处理,处理结果如图5(b)所示,依此二值图象信息,会误将

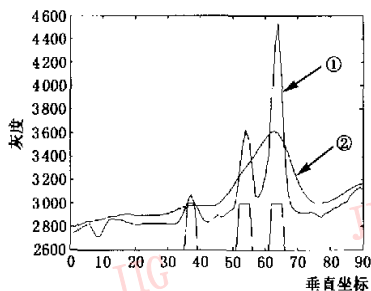
非条带区域当作条带区域抽取出来.因此,在这里再一次用高斯平滑作为确定阈值的依据这种处理方法.如图5(c)所示,对每一个泳道的信息作垂直方向上的投影(图5(c)中的曲线①),同时,对经过高斯平滑滤波之后图象的泳道信息也作垂直方向上的投影(图5(c)中的曲线②).将曲线②作为曲线①的阈值,就可以正确地圈定这个泳道中的3个条带.依照这个方法得到该图象的最终结果如图5(d)所示.



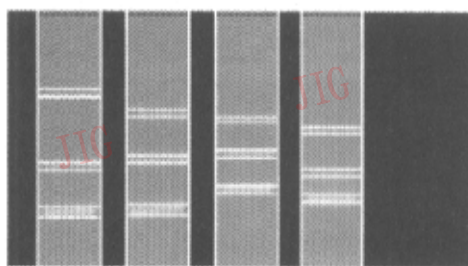
(a) 单个泳道原图及二值图象



(b) 经水平锐化后的单个泳道及二值图象



(c) 单个泳道的垂直投影



(d) 图4(a)基于投影的条带圈定结果

图5 PCR 条带信息的自动圈定

### 2.2.2 条带移动模糊干扰的抑制

由于所得到的图象是在 PCR 扩增过程中所拍摄的,所以,难免在得到的图象中存在拖尾现象(见下页图6(a)).如果不作特殊处理,在第2、第3、第4泳道中很容易产生条带的误抽取.同样,高斯平滑阈值方法可以对其移动模糊干扰进行抑制.图6(b)为第3泳道的原图及其高斯平滑后的图象.图6(c)为图6(b)在垂直方向上的投影曲线.图6(d)为PCR条带的抽取结果.

## 3 实验结果与分析

利用前面所论述的方法,对120幅样本图象进

行了处理,PCR条带的自动圈定准确率为96%(见下页图7).但是,由于PCR在扩增过程中存在一定的轻度倾斜和平移,自动圈定的条带面积忽略了扩增过程中的平移和倾斜.如果情形严重的话,会给后面的定量分析带来一定的误差,因此,给出了人机交互界面,对其进行微调.

通过人机界面圈出平移或倾斜比较严重的局部区域进行自动的校正处理.如图8(a)所示,被虚线圈出的条带存在微小的倾斜,经自动判断,倾斜角度为3°,未经倾斜校正之前,圈定的PCR条带区域的高度为8个像素(见下页图8(a)),经过倾斜调整之后,圈定的PCR条带区域的高度为6个像素(见下页图8(b)),由此,使其面积的统计更加精确.

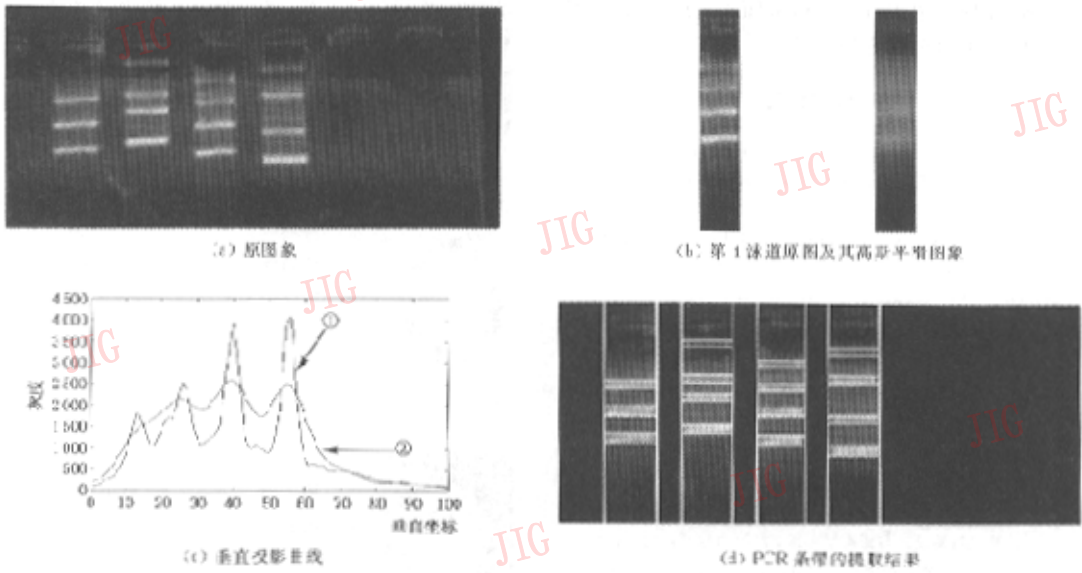


图 6 拖尾阴影影响的提取效果

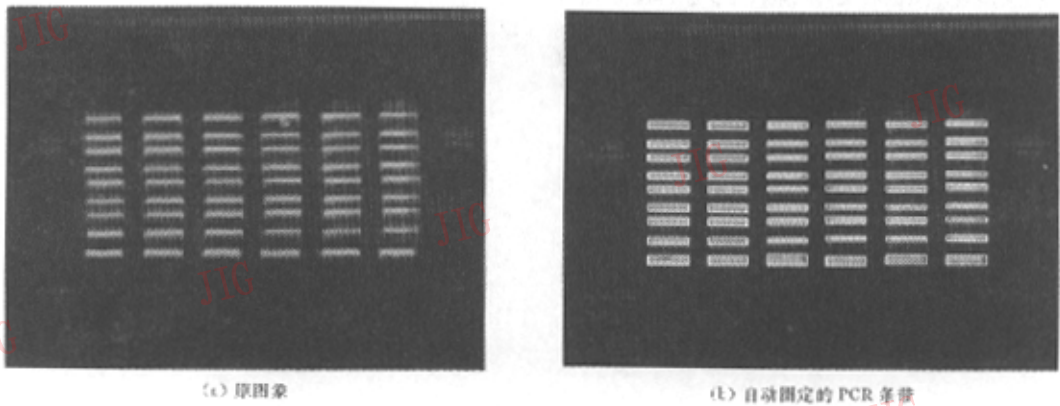


图 7 自动圈定的 PCR 条带信息



图 8 局部校正处理效果

### 4 结 语

本文通过综合使用高斯平滑技术和灰度投影技术,充分利用了 PCR 凝胶成像信息的条带特征和泳

道特征,自动提取出其所含信息。大量的实验结果表明,该方法可以抵抗光照不均,信息强弱不均以及移动模糊等的影响,达到了较高的准确率(96%),而且,为了解决 PCR 在扩散过程中的平移和倾斜的问题,在所设计的系统中,可以通过人机对话,圈定某

个局域,在局域中进行自动检测倾斜角度,并进行校正处理,即可达到更好的效果。

### 参 考 文 献

- 1 Richard Mansfield. Visual basic 5 the comprehensive guide[M]. North Carolina USA:Ventana Communications Group Inc,1997:10~1101.
- 2 Evangelos Petroustos. Mastering visual basic 6[M]. USA: SYBEX Inc,1998:65~394.
- 3 Kyrpides NC. Genomes online database(GOLD 1.0):a monitor of complete and ongoing genome projects world-wide[J]. Bioinformatics, 1999,15(9):773.
- 4 徐秀林,方响亮,果忠祥等. 多媒体分子生物学图像分析系统[J]. 生物医学工程学杂志,2000,17(4):447~450.
- 5 黄淳,吕元,林祥通. 免疫PCR技术的现状和展望[J]. 上海医学检验杂志,1997,12(2):109~110.
- 6 赵雪松,陈淑珍. 综合全局二值化与边缘检测的图像分割方法[J]. 计算机辅助设计与图形学学报, 2001,13(2):118~121.
- 7 付忠良. 基于图象差距度量的阈值选取方法[J]. 计算机研究与发展,2001,38(5):565~567.
- 8 聂生东,陈瑛,顾顺德等. 磁共振脑部图象快速模糊聚类分割算法的研究[J]. 中国生物医学工程学报, 2001,20(2):104~109.
- 9 桑农,张天序. 基于Hopfield神经网络的FLIR图象分割[J]. 自动化学报,2001,27(3):303~309.



朱虹 1963年生,1999年获日本福井大学信息工程学科博士,现为西安理工大学信息与自动化工程学院教授。主要从事图象处理与应用、图象自动识别技术、图象压缩编码以及模式识别等方面的研究。



许朝晖 1975年生,硕士研究生。主要从事医学图象处理等方面的研究。



周健 1975年生,硕士研究生。主要从事图象处理、信息处理等方面的研究。