

## 原子力显微镜观测生物大分子图像的一种处理方法\*

季超 张凌云<sup>†</sup> 窦硕星 王鹏业<sup>‡</sup>

(中国科学院物理研究所软物质重点实验室, 北京 100190)

(2011年1月21日收到; 2011年2月16日收到修改稿)

针对具有复杂结构的生物大分子图像, 本文改进了已有的半自动化方法, 设计了一种新的算法去消除背景噪声, 目的在于能够从生物大分子图像高效准确地提取量化信息. 通过利用这种新的处理程序对生物大分子 DNA 原子力显微镜图像进行了研究, 不仅可以对线性构像的 DNA 图像进行了处理, 而且对于具有分支构像复杂 DNA 图像也能够处理. 本文所提出的数字图像处理方法具有普适性, 它也可以被应用到其他成像领域.

**关键词:** 原子力显微镜, 数字图像处理, DNA

**PACS:** 87. 64. Dz, 87. 57. N -, 87. 14. gk

## 1. 引言

在近 10 年来, 越来越多的物理学技术和理论加入到生物领域的研究, 研究过程中, 使用量化的研究手段就变得越来越重要<sup>[1,2]</sup>, 其中对于生物体中最重要的遗传因子 DNA 的研究也更多, 更深入, 相应的量化信息采集与分析也就愈加有用和重要<sup>[3,4]</sup>. 在 DNA 分子大小的分析上, 电泳分析是使用最为广泛的方法. 但是电泳在实验过程中有诸多的限制, 在通常的电泳分析实验中, 往往要经过超过 2 h 的实验时间, 而且最为致命的是电泳实验分析所使用的 DNA 量必须足够多, 才能得到比较准确的结果. 另外一个方法就是借助于光学显微镜<sup>[5,6]</sup>, 在 DNA 分子上进行荧光染色, 使用光学显微镜观测荧光染料处理后的 DNA, 从而计算得到 DNA 长度信息, 这种方法能够提供高分辨率、低样品量、高产能, 但是对于小于 600 bp 的 DNA 片段无法提供准确的、清晰的信息. 同时我们应该注意到这两种常用的方法都需要对 DNA 分子进行荧光染色, 而染色后的 DNA 结构会发生变化, 使所得到的 DNA 片段量化信息的准确性、可靠性大大降低.

DNA 大分子片段精确量化的信息, 对于理解

DNA 分子的生物学和物理学特性有至关重要的意义<sup>[7,8]</sup>, 正因如此, 原子力显微镜的使用对生物大分子的定量研究带来革命性的变化<sup>[9,10]</sup>. 原子力显微镜是借助于尖锐而且富有弹性的探针, 通过探针针尖与 DNA 分子的相互作用, 使得探针悬臂产生扭曲, 将改变激光在光电二极管反射位置, 光信号转化为电信号, 通过计算机获得表面形貌图像<sup>[11]</sup>. 基于这个原理, 原子力显微镜不仅可以对固定在云母表面的 DNA 分子, 而且可以对在生物缓冲液中的 DNA 分子能够实施动态的观测<sup>[12]</sup>. 另外一方面, 原子力显微镜具有高信噪比的优势, 能够在纵向得到表面上 DNA 生物分子 0.1 nm 高度分辨率的信息和横向 1—10 nm 宽度分辨率的信息, 并且能够直接观测 DNA 片段的整体形貌图像.

综上所述, 得益于原子力显微镜观测技术, 我们能够清晰的观测到 DNA 片段分子的形貌特征, 接下来我们面临的挑战就是如何从 DNA 形貌构像图中获取准确的信息. 近些年来, 利用原子力显微镜图像开发出半自动和自动化算法. 半自动化的算法一种是需要操作人员在 AFM 图像上, 沿着 DNA 的轮廓, 在 DNA 片段上标记不连续的点, 然后将选择的点连接成线, 计算线段的长度<sup>[13]</sup>; 另外一种是需要给定在 DNA 片段末端的初始点, 之后程序才会运

\* 国家自然科学基金(批准号:10834014)和国家重点基础研究发展计划(973 计划)(批准号:2009CB930704)资助的课题.

<sup>†</sup> 通讯联系人. E-mail: lyzhang@aphy. iphy. ac. cn

<sup>‡</sup> 通讯联系人. E-mail: pywang@aphy. iphy. ac. cn

行<sup>[14]</sup>. 半自动的算法使用起来非常费时费力, 而且容易出现人为错误. 自动化算法被越来越多的应用, 并且发展出多种计算方法<sup>[15,16]</sup>, 但是这些自动化算法仍有其不足之处, 它们不能提供矢量的角度计算. 结合我们的工作和实验需求, 采用能够进行矢量运算的半自动化算法<sup>[14]</sup>, 并对其进行了优化和改进, 不仅实现了自动化工作的要求, 而且得到了更多的构像信息. 我们方法的优势在于:

1) 对 AFM 图像优化了背景噪声处理算法. 根据实际的 AFM 图像的特点, 采用了比较简单而且非常高效的优化方法, 并且设计方便的修复程序, 修复因为消除背景噪声带来的信息丢失. 在已有半自动算法的基础上, 使其处理方法自动化.

2) 对具有更加复杂的线性 DNA 图像进行计算. 在研究 DNA 分子与蛋白、多价金属离子的实验中, DNA 片段往往不再仅仅是线性构像. 对于具有复杂结构 DNA 片段, 我们设计了新的处理方法, 成功处理了具有分支构像的 DNA 片段.

在研究过程中, 我们根据算法设计的程序解决了半自动算法所带来的人为偏差, 但是在程序处理完成后, 我们生成 AFM 处理后的图像, 分析人员可以进行检查, 兼顾了 AFM 图像处理的效率和准确性. 我们利用算法编写了 Matlab 程序成功解决了实验研究过程中的问题, 从 AFM 图像中提取了完整的 DNA 构像信息, 帮助我们更好的理解 DNA 分子.

## 2. 原子力显微镜数字图像优化处理

在原子力显微镜观测过程中, 扫描管在扫描时会受到仪器震动或者空气声音的影响, 形成背景噪声, 为了能够清晰准确的从 AFM 图像中得到量化的数据, 我们需要设计程序, 对原始 AFM 图像进行预处理. 首先, 程序把原始的彩色图像(图 1(a)) 转化为灰度图像(图 1(b)), 方便图像的处理效率. 对于背景噪声的处理, 我们使用竖向邻域阈值滤镜处理, 计算 AFM 图像的每竖向一定范围内像素灰度平均值, 对相应竖向像素上的每一点与邻域的平均值进行比较, 若像素灰度值小于平均值, 相应位置像素灰度值则置零; 若像素灰度值大于平均值, 相应位置像素灰度值则置为灰度值与平均值之差. 按照这种方法将图像上所有竖向像素处理后, 形成优化后的图像. 在此图像基础上进行全区域阈值处理,

高于阈值的像素灰度值置为“1”, 低于阈值的像素灰度值置为“0”, 我们获得了二值黑白图像(图 1(c)). 在获得的二值图像中, 因消除背景噪声和进行阈值处理的操作, 会出现图像中的 DNA 部分片段误当背景噪声删除掉, 因此我们需要对二值图像进行优化, 以便获得更加方便准确的图像. 在我们的图像处理程序中使用了卷积的方法进行优化, 根据所用 AFM 图像的尺寸和 DNA 片段的相对宽度, 设计相应的矩阵滤镜, 滤镜中的元素呈正态分布, 对图像进行卷积处理, 最后得到所需要的图像(图 1(d)). 在三维图像比较中, 我们可以看到图 2(a) 是原始图像(图 1(b)) 三维图, 图 2(b) 是经过优化后的图像(图 1(d)) 三维图. 在图像上, 背景噪声大部分已经被消除, 同时 DNA 的形貌图得到很好的保留.

## 3. 线性构像 DNA 图像处理

### 3.1. DNA 片段初始点

在优化后的图像图 1(d) 上, 以图像左上角像素为起始点, 沿横向方向顺序搜索. 我们设定固定的阈值, 规定为圆心点. 当所搜索到的图像像素灰度值高过阈值, 则以两倍步长(2l) 为半径, 开始沿着逆时针方向寻找相同阈值的像素位置. 若完成其半圆搜索后, 没有找到相应的像素位置, 则说明此位置不是我们所希望找到的 DNA 片段, 我们设计程序将此位置以及邻域全部置“0”, 并且继续寻找下一个满足条件的像素位置. 若完成半圆搜索后, 找到了符合阈值的像素位置, 确定为邻域点, 则开始下一步确定初始点位置的处理(图 3(a)). 规定圆心点 A 和邻域点 B 的方向为切线方向  $\vec{AB}$ , 切线方向的垂线方向为法线方向. 首先, 在过邻域点 B 的法线方向上, 以一定长度的范围内, 计算法线方向上位置信息和相对应的灰度值的平均值, 得到法线方向上灰度值最大的位置, 规定为新的邻域点 B'. 通过新邻域点 B', 将切线方向上的圆心点平移, 固定为新的圆心点 A'(图 3(b)).

### 3.2. DNA 片段的连接点

$$X_z = \frac{\sum_{n=0}^m Z(x) \cdot x(n)}{\sum_{n=0}^m Z(x)}, \quad (1)$$

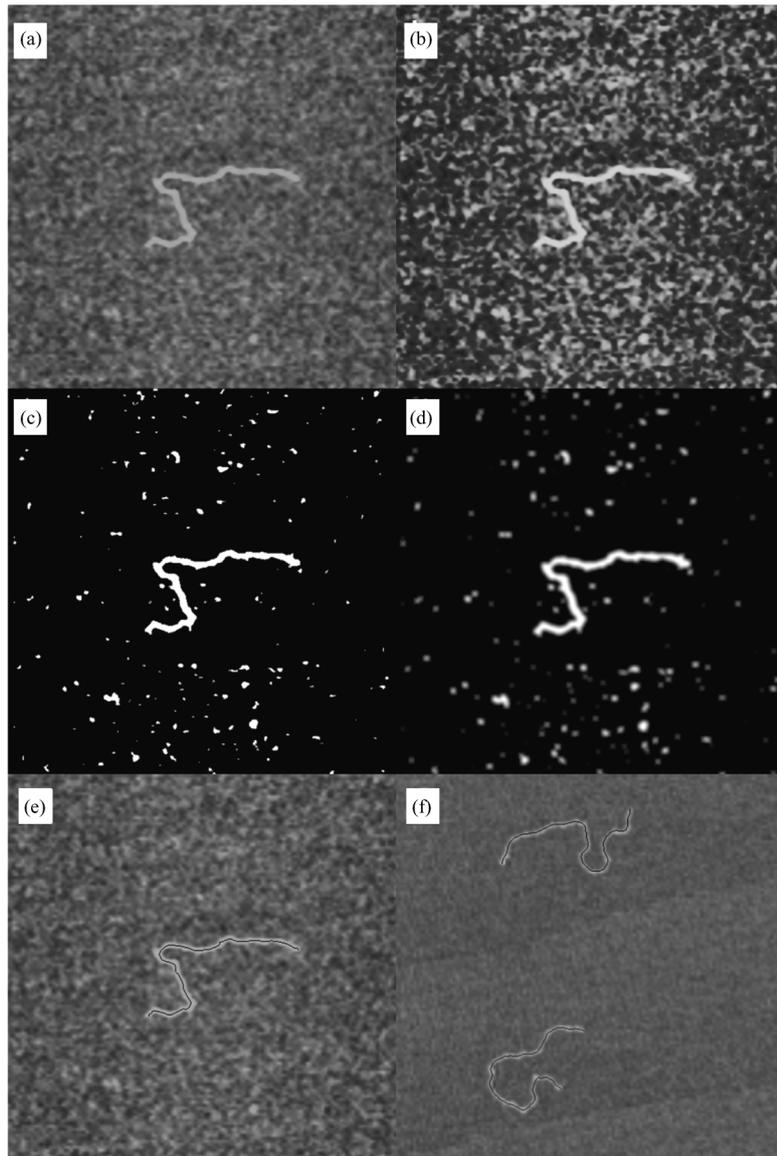


图1 线性 DNA 数字图像处理程序对 AFM 图像处理结果

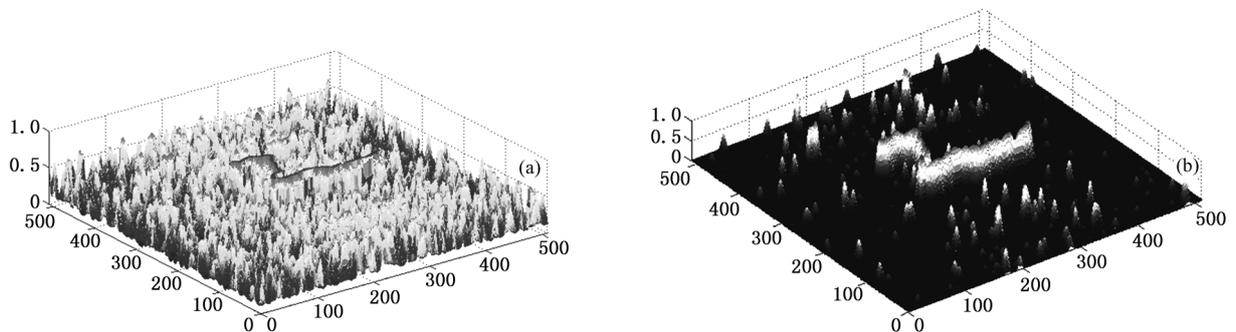


图2 (a)原子力显微镜观测得到生物大分子 DNA 的原始数字图像;(b)使用背景噪音处理程序优化后的数字图像

式中,  $\mathbf{x}(n)$  表示法线方向上一定范围内的位置;  
 $Z(\mathbf{x})$  表示在相应位置的灰度值;  $X_s$  表示在一定范

围内的灰度最大值位置.

规定新邻域点  $B'$  为初始点, 通过初始点  $B'$ ,

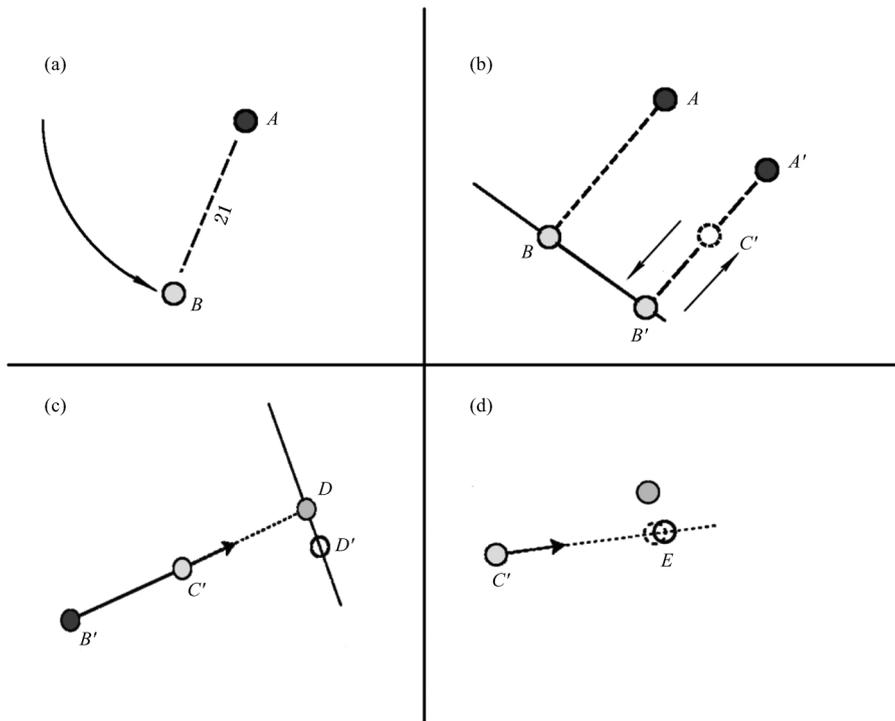


图3 线性 DNA 数字图像处理方法示意图

沿着切线方向上  $\overrightarrow{B'A'}$ , 在距离初始点  $B'$  步长  $l$  的位置, 确定另外一个初始点  $C$ . 通过初始点  $C$ , 沿着垂直  $\overrightarrow{B'A'}$  的法线方向  $x$  上一定范围内, 通过 (1) 式得到灰度最大值位置, 确定为新的初始点  $C'$ . 由初始点  $B'$  和初始点  $C'$  确定新的切向方向  $\overrightarrow{B'C'}$ , 重复以上步骤, 寻找新的初始点. 经过三次重复之后, 确定最终的初始点  $B'$  和初始点  $C'$  (图 3(b)).

首先按照图 3(c) 表示, 以初始点  $\overrightarrow{B'C'}$  方向, 步长  $l$  为距离, 沿着初始点的切线方向  $\overrightarrow{B'C'}$ , 确定新的位置点  $D$ , 通过位置点  $D$ , 沿着垂直切线方向  $\overrightarrow{B'C'}$  法线方向, 寻找到法线方向上灰度值最大的位置点, 规定为位置点  $D'$ . 由位置点  $C'$ ,  $D'$  为新切向方向  $\overrightarrow{C'D'}$ , 重复利用 (1) 式计算, 寻找法线方向上灰度值最大的位置. 经过三次重复处理后, 最终确定新的位置点  $E$  (图 3(d)). 将位置点  $C'$  规定为新的初始点  $B'$ , 新的位置点  $E$  规定为初始点  $C'$ , 重复以上步骤.

按照图 3(c), (d) 所示, 连续按照以上程序的处理方法, 直至整条 DNA 片段被固定步长  $l$  的位置点所覆盖. 沿初始点  $\overrightarrow{C'B'}$  的方向重复以上步骤, 将整条 DNA 片段上所有的位置信息记录下来并储存在一个数组中.

### 3.3. 数字图像的数据计算

$$l_{i,i-1} = \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2}, \quad (2)$$

$$\cos(\alpha) = \frac{X_a \cdot X_b + Y_a \cdot Y_b}{\sqrt{X_a^2 + Y_a^2} \cdot \sqrt{X_b^2 + Y_b^2}}, \quad (3)$$

其中  $X_a = x_i - x_{i-1}$ ;  $Y_a = y_i - y_{i-1}$ ;  $X_b = x_{i+1} - x_i$ ;  $Y_b = y_{i+1} - y_i$ . 把所有位置点线段的距离利用 (2) 式计算, 然后加和得到 DNA 片段的轮廓长度, 计算末端位置的距离得到 DNA 片段的首末端距, 切线方向的角度可以通过 (3) 式计算得到.

### 3.4. 数字图像的后续处理

在彩色图像 (图 1(a)) 上用有颜色的线段把所有位置点顺序连接起来, 输出到另外一个图像中彩色图像中 (图 1(e)) 以便后期人工检查. 在灰度图像 (图 1(d)) 上, 把所有标记的位置点邻域置“0”, 防止程序重复寻找 DNA 片段. 完成所有以上处理之后, 从 DNA 片段的初始点继续搜索合适的灰度值位置, 直至这个图像搜索完成.

在另一个 AFM 图像 (图 1(f)) 中, 可能会出现多条 DNA 片段构像, 我们在程序中使用了不同的颜色来描述 DNA 片段, 这样可以在最后的阶段, 方便人工检查数据的可靠性. 这个方法一方面可以得到

DNA 片段的长度信息,更主要的是同时能够得到 DNA 弯曲角度的信息.

#### 4. 分支构像 DNA 图像处理

在利用 AFM 研究生物大分子的工作中,经常需要对蛋白与 DNA、小分子与 DNA 相互作用进行研究<sup>[17,18]</sup>,这种状况下,需要对局部出现分支结构的 DNA 进行处理.我们在原先处理的基础上改进了方法,使其能够处理更加复杂的问题.

##### 4.1. 数字图像的预处理

在之前优化的图像上(图 1(d)),使用阈值处理,把图像转化为二值图像(图 4(a)),然后将图像上的 DNA 片段进行骨骼化处理,使其转化为连续的线状构像(图 4(b)),使用滤镜对 DNA 片段上的骨骼化分支进行修剪,将图像上 DNA 片段的线状分支末端进行删除(图 4(c)).在图像修剪合适后,停止图像的优化处理,可以进行下一步处理.

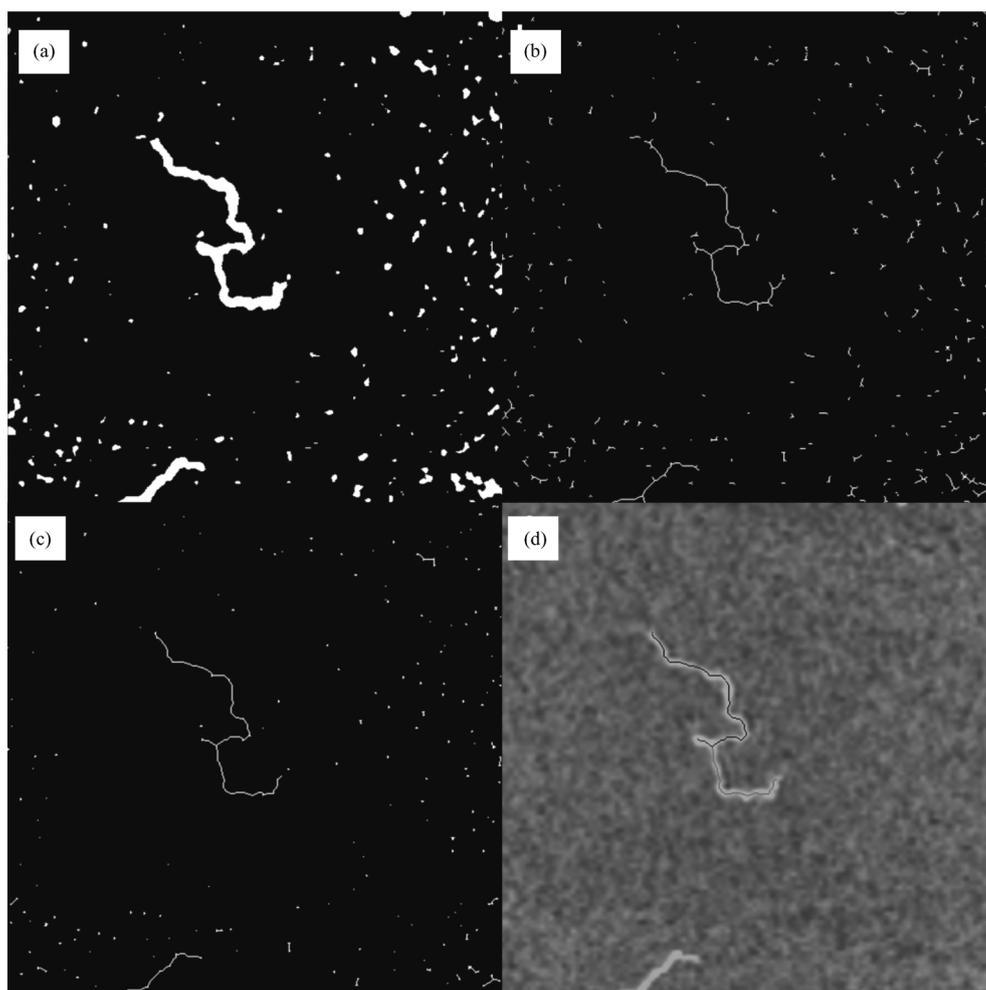


图4 分支构像 DNA 数字图像处理程序对 AFM 图像处理结果

##### 4.2. DNA 片段的初始点

由于我们所使用的图像是二值图像,所以图像中,DNA 片段的线状部分是“1”,背景部分是“0”.在图像的初始点处,沿着横向方向进行自动寻找亮度值是“1”的位置.程序找到合适的位置后,以这一位置点为中心  $A$ ,一定的像素点为步长(我们使用了 4 个像素点为步长),沿着逆时针方向旋转,寻找

亮度同样是“1”的位置,如果没有合适的位置,则说明程序找到的位置不是我们所需的位置,将找到的位置点  $A$  置“0”.如果找到符合条件的位置,则程序再以顺时针方向旋转,同样寻找亮度是“1”的位置,如果找到的位置和逆时针旋转找到的位置相同,则将圆心位置的点坐标储存为 DNA 片段初始点(图 5(b));如果找到的位置和逆时针旋转找到的位置不同(图 5(c)),则将顺时针找到的位置  $B'$  设

为新的圆心,以步长为半径,继续按照上面的步骤,判断逆时针寻找到的位置  $B$  和顺时针找到的位置

$B'$  是否相同,直至找到相同的位置,则圆心位置点的坐标存储为 DNA 片段初始点.

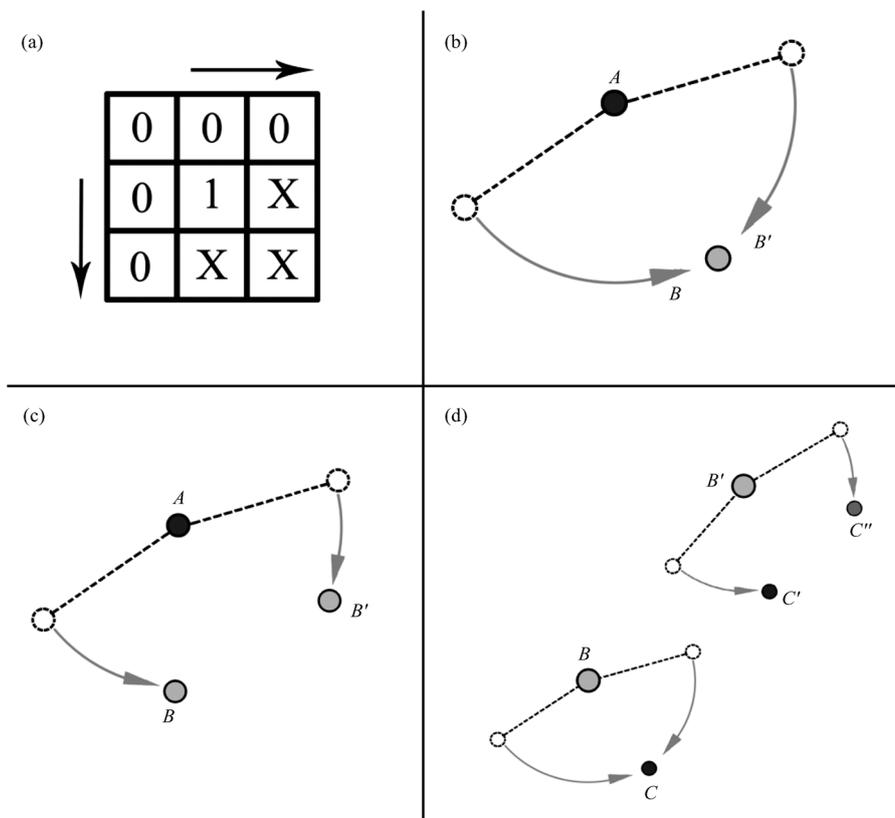


图 5 分支构像 DNA 数字图像处理示意图

### 4.3. DNA 片段分支结构的确定和处理

设片段初始点为圆心  $A$  开始,在圆心点邻域八个像素中(图 5(a)),分别沿着逆时针方向和顺时针方向需找亮度为“1”的位置,分别是  $B$  和  $B'$ ,若亮度位置  $B$  和  $B'$  坐标相同,则说明没有分支结构,将圆心置“0”,在彩色图像上进行标记,并且纪录此处圆心位置的坐标.以找到的亮度点为新的圆心点  $A$ ,重复以上操作;若找到的亮度位置坐标不同,则以步长为半径,重新以原圆心点  $A$  做逆时针和顺时针方向寻找亮度点,如果没有找到,则说明此处是图像上 DNA 片段的一个末端;如果找到的两个位置  $B$  和  $B'$  坐标位置相同(图 5(b)),则说明此处是 DNA 片段上的亮点,将此圆心  $A$  的邻域置“0”,并且以找到的亮度点位置为新的圆心点,继续上面的步骤;如果找到的两个位置  $B$  和  $B'$  坐标不同,则说明此处是 DNA 片段的分支位置(图 5(c)),将两个位置坐标和圆心坐标进行平均化,计算出其分支结构的中心坐标.对于分支结构的两个位置  $B$  和  $B'$  坐标,分别设置为初始点,按照以上处理方法进行处

理,并使用不同的颜色点在彩色图像中标出不同的分支片段(图 5(d)).

### 4.4. DNA 片段的数据处理

在所有分支结构搜索完毕之后,对所有找到的数组依次进行处理,根据实际的需要,设定不同的步长,利用(2)式计算 DNA 片段中各个分支的长度,同时用不同颜色的线段在彩色图像上标记各分支片段(图 2(d)),方便后期工作的检查.

## 5. 结 论

在本文中,我们提供了一套有效、省时省力而又准确的方法去处理由原子力显微镜所观测得到的生物大分子构像的图像,我们从数字图像中识别并提取了大量高精度的生物大分子信息.我们改进了以往的算法,使用了自制的背景噪声消除方法,实现了在大范围 AFM 扫描图像中提取多条 DNA 的功能,提高了 AFM 在该研究领域内的效率.我们发明了一种能够解决复杂构像 DNA 图像的方法,从而

可以进一步研究诸如 DNA 生物大分子与小分子相互作用,使得 AFM 在生物大分子研究领域的量化信息能力方面进一步提高. 此外,针对日益增多的数字图像获取手段,这种数字图像处理方法不仅仅能

够处理 AFM 数字图像,也可以用来处理其他实验手段获得的数字图像. 这种数字图像处理方法的应用以及今后的改进,能够节省人力和时间,提供准确的定量信息.

- [1] Rivetti C, Guthold M, Bustamante C 1996 *J. Mol. Biol.* **264** 919
- [2] Wei Z Y, Zang L H, Li M, Fan W, Xu Y J 2005 *Acta Phys. Sin.* **54** 4955 (in Chinese) [魏志勇、臧黎慧、李明、范我、许玉杰 2005 物理学报 **54** 4955]
- [3] Rivetti C, Guthold M, Bustamante C 1999 *EMBO J.* **18** 4464
- [4] Rivetti C, Codeluppi S 2001 *Ultramicroscopy* **87** 55
- [5] Meng X, Benson K, Chada K, Huff E J, Schwartz D C 1995 *Nat. Genet.* **9** 432
- [6] Liu Y Y, Don S X, Wang P Y, Xie P, Wang W C 2005 *Acta Phys. Sin.* **54** 622 (in Chinese) [刘玉颖、窦硕星、王鹏业、谢平、王渭池 2005 物理学报 **54** 622]
- [7] Hamon L, Pastre D, Dupaigne P, Le Breton C, Le Cam E, Pietrement O 2007 *Nucleic Acids Res.* **35** e58
- [8] Ratcliff G C, Erie D A 2001 *J. Am. Chem. Soc.* **123** 5632
- [9] Kada G, Kienberger F, Hinterdorfer P 2008 *Nano Today* **3** 12
- [10] Hu H L, Zhang K, Wang Z X, Kong T, Hu Y, Wang X P 2007 *Acta Phys. Sin.* **56** 1674 (in Chinese) [胡海龙、张琨、王振兴、孔涛、胡颖、王晓平 2007 物理学报 **56** 1674]
- [11] Fan K Q, Jia J Y, Zhu Y M, Liu X Y 2007 *Acta Phys. Sin.* **56** 6345 (in Chinese) [樊康旗、贾建援、朱应敏、刘小院 2007 物理学报 **56** 6345]
- [12] Zuccheri G, Dame R T, Aquila M, Muzzalupo I, Samori B 1998 *Appl. Phys. A: Mater. Sci. Process.* **66** S585
- [13] Lysetska M, Knoll A, Boehringer D, Hey T, Krauss G, Krausch G 2002 *Nucleic Acids Res.* **30** 2686
- [14] Wiggins P A, Van der Heijden T, Moreno-Herrero F, Spakowitz A, Phillips R, Widom J, Dekker C, Nelson P C 2006 *Nat. Nanotechnol.* **1** 137
- [15] Ficarra E, Benini L, Macii E, Zuccheri G 2005 *IEEE Trans. Inf. Technol. Biomed.* **9** 508
- [16] Sanchez-Sevilla A, Thimonier J, Marilley M, Rocca-Serra J, Barbet J 2002 *Ultramicroscopy* **92** 151
- [17] Solis F J, Bash R, Yodh J, Lindsay S M, Lohr D 2004 *Biophys. J.* **87** 3372
- [18] Wang H, Bash R, Lindsay S M, Lohr D 2005 *Biophys. J.* **89** 3386

## A new method to deal with biomacromolecular image observed by atomic force microscopy\*

Ji Chao Zhang Ling-Yun<sup>†</sup> Dou Shuo-Xing Wang Peng-Ye<sup>‡</sup>

(Key Laboratory of Soft Matter Physics, Beijing National Laboratory for Condensed Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China)

(Received 21 January 2011; revised manuscript received 16 February 2011)

### Abstract

In this paper, we develop a new method to eliminate the background noise in the image of biomacromolecule with complex conformation in order to obtain highly effective and accurately quantitative information. By using this algorithm, we are able to deal with the atomic force microscopic images of both linear DNA and branch DNA. Furthermore, our method can be extended to the study of the images in the other field.

**Keywords:** atomic force microscopy (AFM), image analysis, DNA

**PACS:** 87.64.Dz, 87.57.N-, 87.14.gk

\* Project supported by the National Natural Science Foundation of China (Grant No. 10834014) and the National Basic Research Program of China (973 Program) (Grant No. 2009CB930704).

<sup>†</sup> Corresponding author. E-mail: lyzhang@aphy.iphy.ac.cn

<sup>‡</sup> Corresponding author. E-mail: pywang@aphy.iphy.ac.cn