

文章编号:1002-2082(2019)01-0073-06

高污染基因芯片图像的网格划分

芦碧波,王 培,刘利群,王永茂,张自豪,林忠华

(河南理工大学 计算机科学与技术学院,河南 焦作 545000)

摘 要:对基因芯片图像进行网格划分是基因芯片图像处理的基础,针对高污染基因芯片图像中高亮污渍对网格划分造成的干扰,将高亮目标分为靶点、污渍块和污渍点 3 种类型,根据其不同特征分别进行处理,提出一种新的网格划分算法。利用全局分割确定高亮目标所在位置,根据污渍点的特征,使用图像腐蚀技术对其进行剔除;根据污渍块的特征,对其先进行膨胀处理,然后对其进行剔除,可消除污渍块及其边缘痕迹。使用自协方差对没有污渍的图像进行网格划分,实验表明,对于高污染基因芯片图像,该算法可以得到理想的网格划分结果,靶点检测平均准确率可以达到 94.73%。

关键词:基因芯片图像;网格划分;污渍块;污渍点

中图分类号:TN202;TP317.4

文献标志码:A

DOI:10.5768/JAO201940.0102002

Gridding dividing of high polluted gene chip image

LU Bibo, WANG Pei, LIU Liqun, WANG Yongmao, ZHANG Zihao, LIN Zhonghua

(School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454000, China)

Abstract: The gridding is a basic step in the gene chip image processing. To suppress the influence of the highlight stains which could interfere with the gridding of the high pollution gene chip images, the highlight objects were divided into three categories: spots, stain blocks and stain points. Based on its characteristics, every category was processed respectively and a new gridding algorithm was proposed. Firstly, the highlight objects were located by utilizing a global thresholding segmentation. Secondly, the strain spots were smoothed via image corrosion technology. The strain blocks were removed using a dilation technology to reduce the trace of edges. Finally, the autocovariance was utilized for gridding the noise-free gene chip images. Experimental results show that the proposed algorithm can obtain an ideal gridding result for the high pollution gene chip images and the spot detection average accuracy can reach 94.73%.

Key words: gene chip image; gridding dividing; stain block; stain point

引言

基因芯片在疾病诊断、基因表达研究、基因匹配、基因组研究、基因发现等领域得到了广泛应用^[1]。生物基因芯片运用微电子和计算机技术可以快速处理多个样品^[2]。使用基因芯片提取生物样品中的信息,一般包括基因芯片制备和基因芯

片信息提取 2 个主要步骤。受到人为因素和不可预料因素的干扰,基因芯片在制备过程中会受到污染,导致扫描获取的基因芯片图像质量下降,为后续基因芯片信息提取带来困难。基因芯片图像中常见的降质现象表现为图像中噪声过多、图像模糊、图像中包含污渍等现象,其中,图像中大面

收稿日期:2018-07-12; 修回日期:2018-08-28

基金项目:国家自然科学基金项目(U1404103);河南省教育厅科学技术研究重点项目(14A520029,15A520070)

作者简介:芦碧波(1978—),男,博士,副教授,主要从事数字图像处理中的图像去噪、图像分割、图像融合、色调映射等方面的研究。E-mail: lubibojz@gmail.com

积的污渍块造成图像中靶点信息提取困难。图像存在较大面积的污渍块对图像的干扰最为严重,也是最难以处理的问题。

基因芯片信息的提取^[3]要经过基因芯片制备、基因芯片扫描、基因芯片图像生成、基因芯片图像处理等流程。为准确提取基因芯片中的重要信息,需要经过图像增强等预处理^[4-6]、网格划分^[7-9]、图像分割^[10]、信息提取等步骤。研究人员提出了多种方法对基因芯片图像进行网格划分^[11-12]。针对规则网格,Helmy提出一种基于阈值的网格划分算法^[13],Bariamis提出了基于边际最大化的网格划分算法^[14],Zacharia利用遗传算法对网格进行划分^[15],Rueda给出一种无参多阈值方法划分网格^[16]。对于无污染或受污染程度比较轻微的基因芯片图像,上述算法可以得到较为理想的网格划分结果。但对于包含高亮污渍的基因芯片图像,由于上述算法未能考虑到污渍的高亮信息对整个基因芯片整体亮度的干扰,因此很难正确进行网格划分。

本文针对基因芯片图像中存在的高亮污渍问题,在分析其亮度、形状等特征的基础上对其进行分类和处理,并给出一种有效的网格划分算法。试验证明,本文所使用的方法可以有效划分包含大块污渍的基因芯片图像。

1 消除基因芯片图像污染

1.1 高亮物体分类

图1为一幅典型的高污染基因芯片图像,与黑色的背景相比,靶点和污渍的颜色鲜艳,亮度值较高。一般而言,靶点虽亮度不一,但大部分为较为规则的圆形,且面积大小相近。而污渍则形状不规则,不同污渍的面积也相差较大。在一张受污渍污染的基因芯片图像中,污渍分布也不规则,而靶点则呈规则的行列分布。

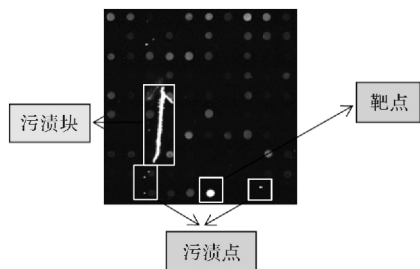


图1 高污染基因芯片图像中的高亮目标

Fig.1 Highlight objects in high polluted gene chip image

根据上述分析,将基因芯片图像中的高亮物

体分为3种类型:靶点、污渍点和污渍块。其中污渍点的特征为面积较小的规则物体,而污渍块为面积较大的不规则物体。为消除污渍点和污渍块对网格划分的影响,需要对它们进行精确定位并加以剔除。根据上述分类方法,从亮度分析出发给出一个污渍定位方法,具体流程见图2。

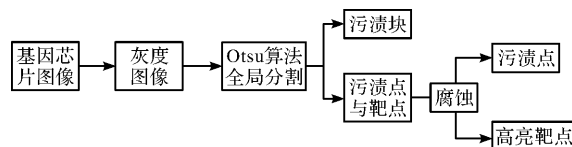


图2 污渍定位流程图

Fig.2 Flow chart of locating stains

1.2 污渍定位

图3为使用Otsu算法对图1中基因芯片图像对应的灰度图像进行分割得到的二值图像^[17],白色表示由Otsu全局阈值分离得到的高亮目标,包括污渍块、污渍点和高亮度靶点3类物体。若剔除图3中的全部目标,则会错误地删除高亮靶点,因此需要进行细致地分析以区分3种高亮物体。考虑到污渍块的形状,本文对白色区域计算物体的偏心率确定污渍块,偏心率较大的目标即为污渍块,结果如图4(a)所示,其中偏心率具体计算公式为



图3 Otsu算法分割图像结果

ig.3 Segmentation result by Otsu algorithm

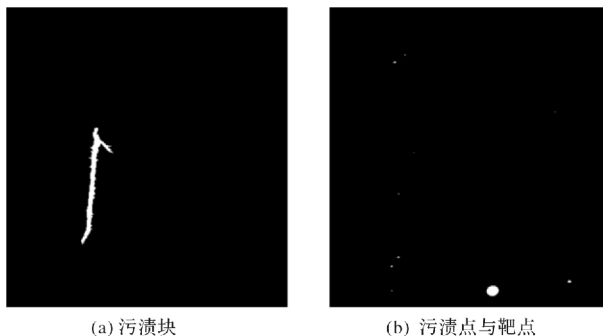


图4 高亮目标分类

Fig.4 Highlight target classification

$$e = \frac{\sqrt{2/[(A+B) - \sqrt{(A+B)^2 + 4C^2}]}}{\sqrt{2/[(A+B) + \sqrt{(A+B)^2 + 4C^2}]}} \quad (1)$$

式中: A 为绕 X 轴的转动惯量; B 为绕 Y 轴的转动惯量; C 为惯性积。

与靶点相比,污渍点的面积较小,因此使用形态学中的腐蚀和膨胀技术检测污渍点的位置。使用 W 表示二值图像中的白色目标, S 表示结构元素,则 $W+S$ 表示用 S 对 W 进行膨胀处理, $W-S$ 表示用 S 对 W 进行腐蚀处理。 $W \circ S$ 、 $W \bullet S$ 分别表示用 S 对 W 进行开运算和闭运算,其关系式如下:

$$\begin{aligned} W \circ S &= (W - S) + S \\ W \bullet S &= (W + S) - S \end{aligned} \quad (2)$$

式中使用的开运算结构元素 $S = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$; 闭运算结构元素 $S = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \end{bmatrix}$ 。

使用图像腐蚀技术,可消除的目标即为基因芯片图像中的污渍点,而高亮靶点目标不会消失,但会在腐蚀过程中产生面积缩小,因此需要使用图像膨胀技术处理腐蚀后的高亮靶点,即可确定高亮靶点的原始位置。图 5(a)为靶点,图 5(b)为污渍点和污渍块。

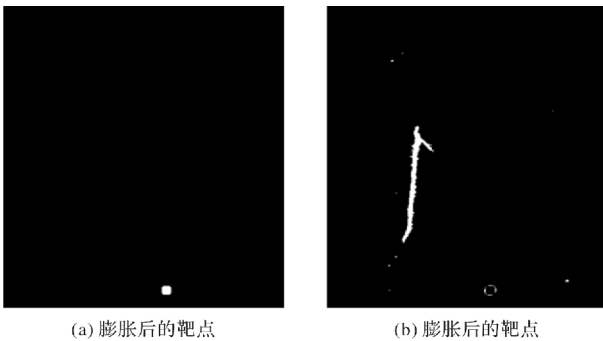


图 5 污渍定位

Fig. 5 Stains location

1.3 污渍痕迹消除

污渍定位后需要消除污渍,以减少污渍对网格划分的影响。但由于污渍所在区域与背景所在区域间的边界并非理想的阶梯型边界,而是存在亮度缓慢变化的过渡区域。过渡区域的亮度虽然低于污渍的亮度,但仍然远高于背景亮度,仍然会对网格正确划分造成影响,因此需要进一步处理以消除过渡区域的干扰。为此,利用下面公式计算消除污渍痕迹:

$$\begin{aligned} c_{xy}(m) &= \begin{cases} \sum_{n=0}^{N-|m|-1} \left(x(n+m) - \frac{1}{N} \sum_{i=1}^{N-1} x_i \right) \times \left(y_n^* - \frac{1}{N} \sum_{i=1}^{N-1} y_i^* \right) & m \geq 0 \\ c_{yx}^*(-m) & m < 0 \end{cases} \\ c(m) &= c_{xy}(m-N) \quad m = 1, 2, \dots, 2N-1 \end{aligned} \quad (4)$$

$$I_{fin} = [1 - (B + S)] * I_{gray} \quad (3)$$

式中: B 为污渍二值图像; S 为进行膨胀处理的开运算结构元素; I_{gray} 为基因芯片图像的灰度图像; I_{fin} 为无污渍的基因芯片灰度图像。对二值图像进行膨胀是为了适度扩大污渍区域,然后将膨胀前后对应的灰度之间进行差值运算。计算结果如图 6(c)所示。从图中可以看出,污渍周围过渡区域已经被很好地消除掉。

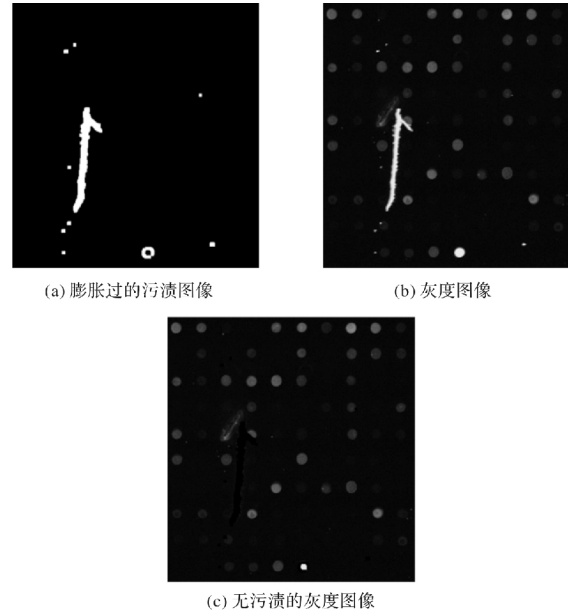


图 6 污渍去除

Fig. 6 Stains removing

2 网格划分算法

由于基因芯片图像中噪声较大,且靶点亮度差异较大,无论是全局阈值分割方法或其他的全局分割方法,会经常出现低亮度靶点被遗漏的情况。为提高分割的精度,一般的处理方式是在图像分割前对基因芯片图像进行网格划分,使每个靶点坐落在特定的区域网格中,然后对相应网格内进行局部分割,因此网格划分则是基因芯片图像分割的前提^[18]。消除基因芯片图中污渍点和污渍块后,可以得到没有污渍的灰度基因芯片图像,根据该灰度图像可以进行网格的划分。

对输入灰度基因芯片图像 I ,首先计算图像列均值 x ,如图 7(a)所示。计算列均值 x 的自协方差:

式中： $*$ 表示复共轭； $c(m)$ 为输出值，包含 $2n-1$ 个元素。自协方差描述信号中2个时刻取值起伏变化的相关程度，求得图像列均值的自协方差 $c(m)$ 。然后计算自协方差 $c(m)$ 的向前差分 s_1 和向后差分 s_2 ，取 $\{s_1>0, s_2<0\}$ 的坐标点集合maxima为自协方差的峰值位置，如图7(b)所示。根据坐标点集合maxima求标准间距，对集合maxima作差分，取其中间值，然后对中间值进行四舍五入运算得到标准间距estPeriod，根据标准间距使用顶帽过滤法过滤 x 得到 x_2 ，使用Otsu算法进行阈值分割，然后计算中心点位置，得到竖直方向网格划分位置坐标 $xGrid$ ，结果如图7(c)所示。同理可得到水平方向网格位置坐标 $yGrid$ ，合并2个方向的结果即可完成基因芯片图像的网格划分，叠加在彩色基因芯片的效果见图8。

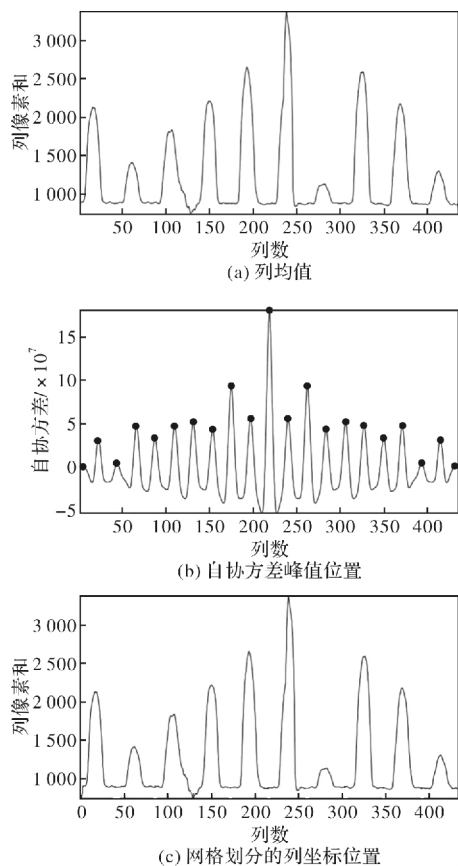


图7 竖直方向的网格划分

Fig. 7 Gridding in vertical direction

由于图像受到噪声污染，图像中有较多的噪声点，这些噪声点影响到网格划分的结果，所以，需要将图像进行去噪，本文使用常见的中值滤波对图6(c)进行滤波，然后使用本文的网格划分算法对滤波后的图像进行网格划分处理，得出的网

格划分结果如图8所示。从图中可以看到，网格划分的结果完全没有受到污渍块与污渍点的影响，每个靶点都坐落在某个网格之中，没有出现错误的网格划分。

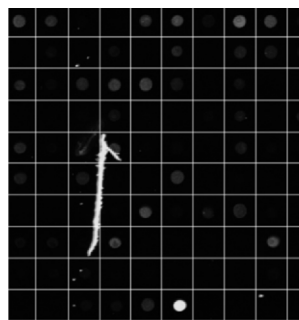


图8 网格划分结果

Fig. 8 Gridding result

3 实验结果分析

本文对近百幅受污染的基因芯片图像进行了试验。试验环境为MATLAB2014a, CPU为Intel Core i5, 内存4 GByte。

在图9中，图9(a)~图9(c)为输入基因芯片图像。图9(a)中包含一个面积较大的狭长污渍块，且其亮度最高，会对整个图像的网格划分产生严重的负面影响；在图9(b)与图9(c)中不仅包含面积较大的狭长污渍块，还包含数量众多的污渍点，污渍点与污渍块的亮度较高、像素值较大。图9(d)~图9(f)为文献[19]对图9(a)~9(c)进行划分的结果。由于原基因芯片中污渍块亮度较大，图9(b)和9(c)中污渍点数量较多，文献[19]使用图像亮度信息进行网格划分，容易受到污渍点的干扰，所以网格划分结果中有多余网格线，使得靶点没有坐落在有效的区域之中。图9(g)~图9(i)为使用文献[20]对图9(a)~9(c)进行划分的结果。文献[20]基于图像信号阈值进行网格划分，由于图9(a)~9(c)中存在大面积污渍块和污渍点，在划分过程中，被误认为污渍块和污渍点为靶点信号，从而不进行网格线划分，所以在网格划分结果图9(g)~9(i)中，缺失多条网格线，致使很多靶点坐落在同一个大型网格之中，由于污渍点的干扰，多个污渍点被误判为靶点，导致在网格划分结果中出现虚假网格线，上述分析表明，该算法在网格划分过程中会同时产生欠分割和过分割现象。图9(k)~图9(m)为本文算法对图9(a)~图9(b)的网

格划分结果。使用本文算法剔除图 9(a)~9(c)中的污渍块与污渍点后,对其灰度图像进行网格划分,可得结果如图 9(k)~图 9(m)。在图 9(k)~图 9(m)中,没有出现错误的网格划分情况,每个靶点都坐落在特定网格中,利于后续的靶点检测和分割。

为了评价网格划分结果,使用文献[19]中的

准确率进行结果评估,准确率计算公式如下:

$$A = (N_{\text{correct spots}} / N_{\text{total spots}}) \times 100\% \quad (5)$$

式中: $N_{\text{correct spots}}$ 为正确分离出的靶点数目; $N_{\text{total spots}}$ 为每幅图像中全部靶点数目。根据公式(5),对 98 幅受污染基因芯片图像进行测试,平均准确率有较大幅度提升,达到了 94.73%。

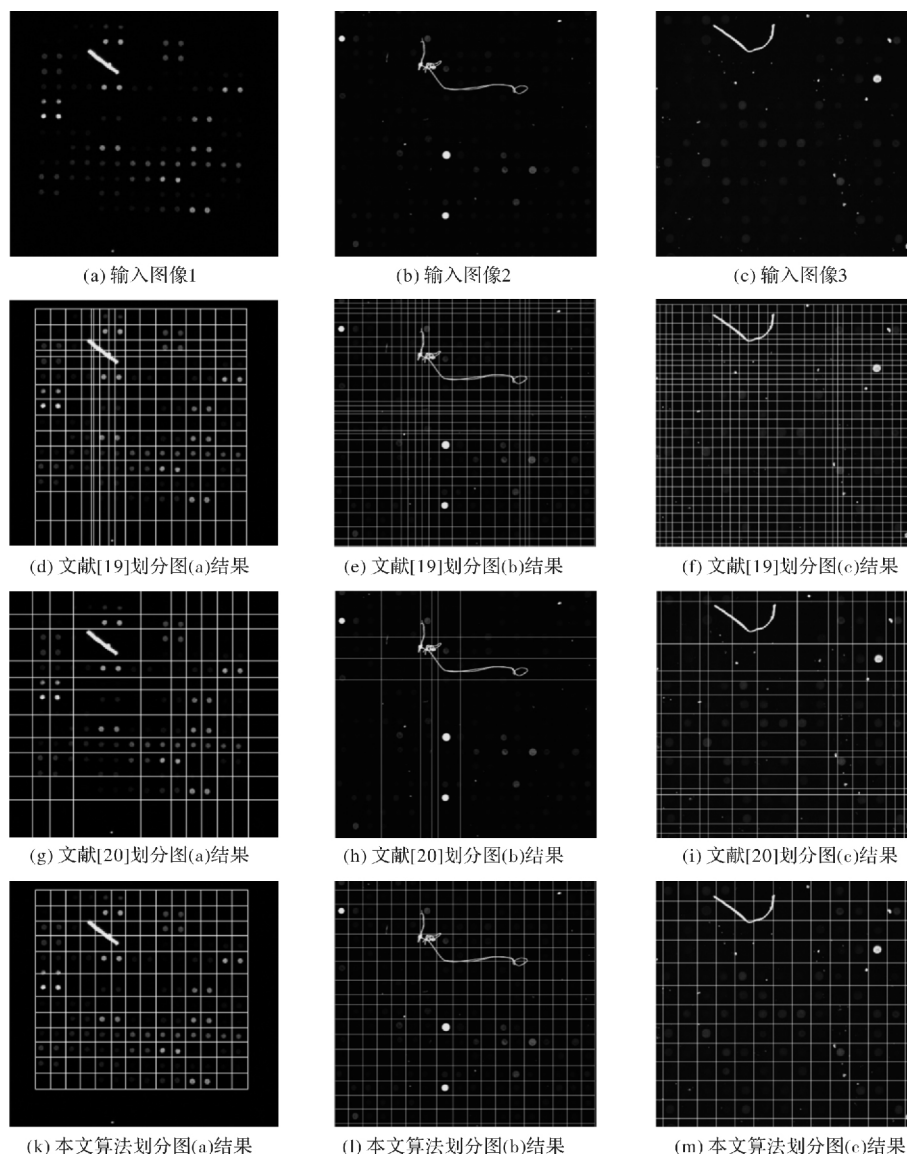


图 9 实验结果

Fig. 9 Experimental results

4 结论

网格划分是基因芯片图像中信息提取的关键步骤。前人所提网格划分算法中考虑了噪声干扰,但未能考虑大面积污渍的干扰,因此无法得到正确的网格划分结果,影响了后续的信息提取。本文针对高污染图像的高亮目标,根据目标的形状、大小等特征,将其分为污渍点、污渍块、靶点三类,对全局分割算法之后的结果使用偏心率、膨

胀、腐蚀进行处理,然后根据自协方差进行网格划分。对 98 幅受污染基因芯片图像进行测试,靶点检测准确率可以达到 94.73%。

参考文献:

- [1] FODOR S P, READ J L, PIRRUNG M C, et al. Light-directed, spatially addressable parallel chemical synthesis[J]. Science, 1991, 251(4995): 767-773.

- [2] YOU Xiaojuan, LU Zhen. The gene chip technology and its application in microbiological detection[J]. Journal of Xi'an University: Natural Science Edition, 2017 (02): 94-96, 116.
 有小娟, 鲁珍. 基因芯片技术及其在微生物检测中的应用[J]. 西安文理学院学报: 自然科学版, 2017 (02): 94-96, 116.
- [3] SHAO Guifang, LI Gang, LI Tiejun, et al. Research on noise processing algorithm for micro-array image highlighting[J]. Journal of Xiamen University: Natural Science Edition, 2009 (03): 330-333.
 邵桂芳, 李刚, 李铁军, 等. 基因芯片图像高亮噪声处理算法研究[J]. 厦门大学学报: 自然科学版, 2009 (03): 330-333.
- [4] LI Qingzhong, LIU Qing. Adaptive enhancement algorithm for low illumination images based on wavelet transform [J]. Chinese Academy of lasers, 2015 (02): 280-286.
 李庆忠, 刘清. 基于小波变换的低照度图像自适应增强算法[J]. 中国激光, 2015(02): 280-286.
- [5] LI Bing, SU Juan, WANG Yanzhao. Infrared image enhancement combined with PCNN segmentation and fuzzy set theory [J]. Acta Optica Sinica, 2016, (09): 1-12.
 李冰, 苏娟, 王延钊. 结合 PCNN 分割和模糊集理论的红外图像增强[J]. 光学学报, 2016(09): 1-12.
- [6] YANG Jianqing, LUO Jijun, XU Jun, et al. Based on the atmospheric environment shelter of the visible image enhancement of [J]. Laser & Optoelectronics Progress, 2017, 54(10): 239-245.
 杨建清, 罗积军, 徐军, 等. 基于大气环境实验方舱的可见光图像增强研究[J]. 激光与光电子学进展, 2017, 54(10): 239-245.
- [7] MOENA D Q. Microarray image gridding by using self-organizing maps[C]. 2011 5th International Conference on Bioinformatics and Biomedical Engineering, 2011: 1-4.
- [8] FOUAD A A, LABIB F Z M, SHARAWY A A R, et al. Developing a new methodology for de-noising and gridding cDNA microarray images [C]. US: IEEE, 2012: 142-145.
- [9] LI Tiejun, SUN Yue, SHAO Guifang, et al. An active analysis method for microarray image gridding [J]. Journal of Xiamen University: Natural Science, 2016, 55(6): 888-894.
 李铁军, 孙跃, 邵桂芳, 等. 基于主动分析的基因芯片图像网格定位[J]. 厦门大学学报: 自然版, 2016, 55(6): 888-894.
- [10] HARIKIRAN J, PHANENDRA M, SWAMY N N, et al. K-means with bi-dimensional empirical mode decomposition for segmentation of microarray image [J]. International Journal of Computer Science Issues, 2012, 9(5): 427-433.
- [11] DATTA A, KONG W K, YOW K C. A fully automatic method for gridding bright field images of bead-based microarrays[J]. IEEE Journal of Biomedical & Health Informatics, 2016, 20(4): 1148-1159.
- [12] AHMAD M M, JAMBEK A B, MASHOR M Y B. A study on microarray image gridding techniques for DNA analysis[C]. US: IEEE, 2015: 171-175.
- [13] HELMY A K, EL-TAWEEL G S. Regular gridding and segmentation for microarray images[J]. Computers & Electrical Engineering, 2013, 39 (7): 2173-2182.
- [14] BARIAMIS D, IAKOVIDIS D K, MAROULIS D. M3G: Maximum margin microarray gridding [J]. BMC Bioinformatics, 2010, 11(1): 49.
- [15] ZACHARIA E, MAROULIS D. An original genetic approach to the fully automatic gridding of microarray images[J]. IEEE Transactions on Medical Imaging, 2008, 27(6): 805-13.
- [16] RUEDA L, REZAEIAN I. A fully automatic gridding method for cDNA microarray images[J]. BMC Bioinformatics, 2011, 12(1): 1-17.
- [17] CHEN Z, TU Y. Improved image segmentation algorithm based on OTSU algorithm[J]. International Journal of Advancements in Computing Technology, 2012, 4(15): 206-215.
- [18] ZHANG Y, WU S. Statistics-adaptive method for cDNA microarray images gridding[C]. US: IEEE, 2012: 380-383.
- [19] MUKHTAR M, JAMBEK A B, MASHOR M Y B. Image gridding algorithm for DNA microarray analyzer[C]. US: IEEE, 2017: 452-457.
- [20] ZACHARIA E, MAROULIS D. A spot modeling evolutionary algorithm for segmenting microarray images[M]. Germany: InTech, 2012.