

文章编号: 1671-7104(2020)01-0007-06

基于瓶颈检测和分水岭算法的重叠宫颈细胞图像分割方法

【作者】 段鹏^{1,2}, 程文播², 钱庆², 章强², 杨任兵², 潘宇骏²

1 中国科学技术大学, 合肥市, 230026

2 中国科学院苏州生物医学工程技术研究所, 苏州市, 215163

【摘要】 针对宫颈细胞图像存在大面积重叠的问题, 该研究提出一种基于瓶颈检测和分水岭算法的图像分割方法。首先使用多边形近似得到细胞轮廓上的所有特征点, 再通过瓶颈检测和椭圆拟合得到正确的分裂点对, 从而确定重合区域的大致范围, 对该区域的梯度图像使用分水岭算法得到内部的边界信息, 最后与外轮廓进行叠加, 得到重叠细胞的分割结果。实验结果表明: 该算法能够从重叠的宫颈细胞图像中分割出单个细胞轮廓, 具有很好的准确度和完整性。与医生手工标记的分割结果接近, 与现有其它算法相比, 分割效果更好。

【关键词】 宫颈细胞分割; 瓶颈检测; 重叠细胞; 分水岭算法

【中图分类号】 TP391.4

【文献标志码】 A

doi: 10.3969/j.issn.1671-7104.2020.01.002

Overlapping Cervical Cell Image Segmentation Based on Bottleneck Detection and Watershed Algorithm

【Writers】 DUAN Peng^{1,2}, CHENG Wenbo², QIAN Qing², ZHANG Qiang², YANG Renbing², PAN Yujun²

1 University of Science and Technology of China, Hefei, 230026

2 Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Science, Suzhou, 215163

【Abstract】 This study proposes an image segmentation method based on bottleneck detection and watershed algorithm to solve the problem of overlapping cervical cell image. First, we use polygon approximation to get all feature points on the cell contour and then use bottleneck detection and ellipse fitting to obtain the correct split point pairs. Therefore, the approximate range of the overlapping region was determined. The watershed algorithm was used to obtain the internal boundary information for the gradient image of the region. Finally, the segmentation results of the overlapped cells were obtained by superimposing with the outer contour. The experimental results show that this algorithm can segment the contour of a single cell from the overlapping cervical cell images with good accuracy and integrity. The segmentation result is close to that of doctors' manual marking, and the segmentation result is better than other existing algorithms.

【Key words】 cervical cell division, bottleneck detection, adhesion cells, watershed algorithm

0 引言

国家癌症中心2018年统计的数据表明, 在中国, 每年被检测出的宫颈癌患者达到13万, 因宫颈癌去世的女性人数约5.2万人。宫颈癌是危害

收稿日期: 2019-05-05

基金项目: 中国科学院科研仪器设备研制项目(YJKYYQ20170067, YJKYYQ20170068); 吉林省与中国科学院科技合作产业化专项资金(2018SYHZ0007); 中国科学院科技服务网络计划(KFJ-STC-SCYD-007)

作者简介: 段鹏, E-mail: duanpeng@mail.ustc.edu.cn

通信作者: 程文播, E-mail: chengwenbo@sibet.ac.cn

女性健康和生命的恶性肿瘤之一。宫颈癌的筛查工作对广大妇女的生命健康有着非凡的作用^[1]。目前经常被使用的宫颈癌的筛查方法是宫颈细胞学检测^[2]。考虑到宫颈癌普查的工作量极大, 通过计算机图像处理技术辅助宫颈癌的普查具有重要的现实意义^[3]。

宫颈细胞的形态、大小等特征是用来判断宫颈疾病的指标之一^[4]。宫颈细胞图像中, 细胞质之间往往会发生大面积重叠, 从重叠的细胞区域

研究 与 论 著

分割出单独的宫颈细胞区域是一项极具挑战性的任务^[5]，染色过程中引起的光照差异和染色浓度等问题都会导致细胞不同成分之间的颜色对比度降低。其它一些例如空气干燥、过多的血液、细菌、炎症的问题也会使得重叠细胞区域的识别更具挑战性^[6]。此外，宫颈细胞还存在本身的形状相对比较复杂，重叠程度较高，细胞质的边界信息较弱等问题。因此如何在宫颈细胞的筛查过程中，提高宫颈细胞图像的分割效果，将为宫颈癌疾病的治疗提供了最有效的帮助。因此本次实验的目的在于讨论重叠宫颈细胞图像的分割算法。

1 方法

1.1 预处理

本研究使用最大类间方差算法对重叠宫颈细胞图像进行预处理，获取宫颈细胞的轮廓信息。最大类间方差算法简称Otsu算法，是一种自动适应确定阈值的方法。算法实现假设图像的所有像素一定可以根据阈值被准确分成细胞区域和背景区域两部分。在此基础上，若需要得到更准确的分割结果，那么就要尽可能使得细胞区域与背景区域这两部分的灰度值区分度越大，而方差正好是用来描述灰度值分布是否均匀的方法，因此通过计算某个阈值处，两个区域的类间方差，当类间方差值达到最大时，此时的阈值就是能够使图像最准确分割的阈值点。

假设一个图像 $I(x, y)$ ，它的像素点区域为 $M \times N$ ，将阈值记为 T ，假定此时细胞区域包含的所有像素的总和占所有像素的比重是 w_0 ，像素的平均值是 μ_0 ，同理，假定背景区域包含的所有像素的总和占所有像素的比重为 w_1 ，像素的平均值是 μ_1 。整个图像灰度值的平均值设为 μ ，方差设为 g 。此外，我们将灰度值低于阈值 T 的像素总和设为 N_0 ，灰度值高于阈值 T 的像素总和设为 N_1 ，有如下的关系：

$$\begin{aligned}
 (1) \quad w_0 &= N_0 / (M \times N) \\
 (2) \quad w_1 &= N_1 / (M \times N) \\
 (3) \quad N_0 + N_1 &= M \times N
 \end{aligned}$$

$$w_0 + w_1 = 1 \tag{4}$$

$$\mu = w_0 \times \mu_0 + w_1 \times \mu_1 \tag{5}$$

$$g = w_0(\mu_0 - \mu)^2 + w_1(\mu_1 - \mu)^2 \tag{6}$$

将式(5)代入式(6)，得到：

$$g = w_0 w_1 (\mu_0 - \mu_1)^2 \tag{7}$$

将阈值从最小依次调整到最大，每次设定阈值后，求出类间方差 g ，当 g 的值最大时，此时预先设定阈值 T 就是使图像完美分割的最佳预置点，使用该阈值可以将细胞区域和背景区域分割出来。

1.2 确定分裂点

重叠细胞在重合的地方往往都会出现凹陷的现象，两个细胞之间会出现凹点而且是成对出现，因此找到成对的凹点是进行重叠细胞正确分割的关键点^[7]。我们将这个凹点对称之为使重叠细胞分割的分裂点。大多数传统的凹点检测方法是基于凸图形的定义提出的一种简便的凹点检测方法，原理简单，但需要大量冗余的计算，耗时较长。考虑到想让算法尽可能快速高效地处理图像，本研究结合多边形近似、瓶颈检测和椭圆拟合进行分裂点检测。

1.2.1 多边形近似

多边形近似^[8]的目的是在某一曲折连续的线段上进行判断，根据多边形的特点检测出特征点作为该线段上近似于多边形的顶点，从而找出可能存在重叠细胞的特征点区域。一个理想的多边形近似的算法应当以尽可能少的计算复杂度，却尽可能地标记出线段的所有形状特点。本研究采用的多边形近似的方法如下：

假设 $C = \{p_1, p_2, \dots, p_i, \dots, p_n\}$ 代表曲线的原始轮廓，式中， $p_i = (x_i, y_i)$ 代表曲线上的每一个点， n 代表原始轮廓上的点数。

第一步：随机选择两个轮廓点 $p_i (i=1)$ 和 $p_j (j=i+2)$ ，如图1所示：

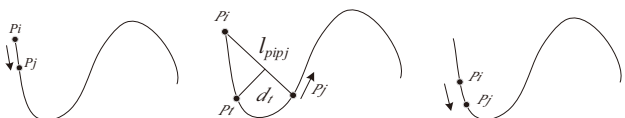


图1 多边形近似的例证
Fig.1 Example of polygon approximation

第二步：计算 p_i 和 p_j 之间每一个轮廓点到 I_{p_i, p_j} 的距离，并和预设阈值 T_d 进行比较，如图1所示。如果存在点 p_i ，其对应的距离 d_i 满足 $d_i = T_d$ ，那么 p_i 被认为是轮廓的分裂点，并且 p_i 和 p_j 从 p_i 和 p_{i+2} 开始继续，否则， p_i 保持不变，而 p_j 变成 p_{j+1} 继续。

第三步：重复第二步，直到遍历了整个轮廓的所有点。

在多边形近似中，阈值 T_d 用于控制精度，阈值设置越小越好。但是相应带来的缺点就是：当阈值设置过小，那么曲线上得到的特征点就过多。因此，在后续步骤中的计算时间也会变的越长。所以，阈值 T_d 出于效率和精度两方面的考虑。

该多边形近似的效果如图2所示，图左为某一细胞的轮廓信息，图右是使用上述步骤得到的所有特征点标注。

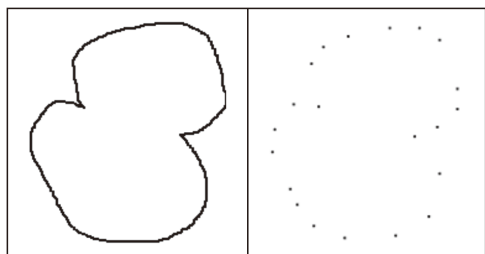


图2 检测特征点
Fig.2 Feature detection

多边形近似处理后得到的所有特征点，都有可能是两个细胞重合部分的分裂点。因此，需要从所有特征点中找到正确的分裂点，才能正确地分割重叠细胞。

1.2.2 瓶颈检测

瓶颈检测^[9]的主要目的是从上述多边形近似找到的所有特征点中，找到两个重叠细胞重合的凹点，即能够使重叠细胞被正确分割的分裂点，正确识别重叠细胞的分裂点是实现重叠细胞正确分割的重要步骤。

瓶颈检测的方法大多数都是基于凹面和形状算法，本文瓶颈检测算法如下所示：对于通过多边形近似得到的所有特征点，通过计算每对特征点之间的瓶颈率来判断该特征点对是否为正确

的分裂点对。假设两个特征点 p_a 和 p_b ，瓶颈率 e 的计算方法是：

$$e(p_a, p_b) = \frac{dist(p_a, p_b)}{\min\{length(p_a, p_b), length(p_b, p_a)\}} \quad (8)$$

式中， $dist(p_a, p_b)$ 为 p_a 和 p_b 两点间的欧几里得距离， $length(p_a, p_b)$ 为点 p_a 到 p_b 顺时针方向的边界长度。 $length(p_a, p_b)$ 和 $length(p_b, p_a)$ 分别表示 p_a 和 p_b 两点之间的上边缘和下边缘的长度。若 p_a 和 p_b 两个特征点计算出的瓶颈率小于阈值 T_e ，那么这两个特征点就被认定为是一组候选分裂点对。对于标准圆形单元，瓶颈率的最小值为 $2/\pi$ 。但是，考虑到宫颈细胞图像可能存在不规则性或椭圆形，阈值 T_e 应该适当小于 $2/\pi$ ，因此本文的阈值设为 $1/2$ 。如果没有检测到候选分裂点对，那表明该区域只包含一个细胞单元。

该算法可以找到一些候选分裂点对，但当单细胞图像本身呈现外凸时，也有可能检测到分裂点对，而此时的分裂点对是错误的分裂点对。图3就是一个未正确检测到分裂点对的例子，右图的细胞轮廓为一个单细胞，此时点与点之间的瓶颈率为0.30，满足 $e(p_a, p_b) < T_e$ ，算法判断它是存在分裂点对，出现了明显的误判断。这就说明利用瓶颈检测检测出来的分裂点对不一定是正确的分裂点对。

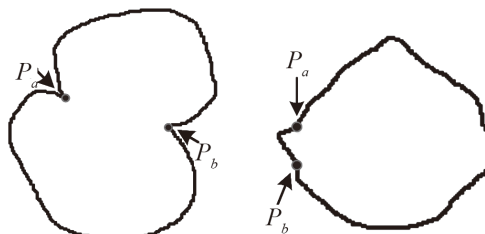


图3 瓶颈检测的例子
Fig.3 Examples of bottleneck detection

1.2.3 椭圆拟合

椭圆拟合^[10]主要有聚类和最小二乘拟合两大类方法。相比聚类方法，最小二乘法的椭圆拟合技术拥有耗时更少、鲁棒性更好、精度更高等优势。鉴于此，本研究采用一种被广泛应用的最小二乘拟合方法直接作为椭圆拟合的工具。

首先,把所有候选分裂点对根据瓶颈率从低到高排序。从第一对候选分裂点开始,将重叠细胞区域的轮廓划分为 L_1 和 L_2 两段。然后,利用多边形近似得到的每个线段上的特征点拟合椭圆,表示为:

$$ax^2 + bxy + cy^2 + dx + ey + f = 0 \quad (9)$$

通常,宫颈细胞在图像中的区域范围差异应该不大。因此,如果两个拟合椭圆的面积差较大时,那么该分裂点对可能是不正确的。用 S_1 和 S_2 分别表示根据 L_1 和 L_2 上的特征点拟合的两个椭圆的面积, r 是由 $r = \max(S_1/S_2, S_2/S_1)$ 计算出的面积比的最大值,如果满足 $r > T_r$,则认为此时对应的候选分裂点不正确。其中 T_r 是一个预设阈值,另外,还利用平均代数距离(MAD)作为另一种度量,进一步确定候选分裂点对是否正确。令 x_i 代表线段 L 上的特征点 (x_i, y_i) , $i=1,2,\dots,s$,在线段 L 上所有特征点与拟合椭圆之间的MAD计算公式^[11]如下:

$$MAD(\hat{a}; L) = \sum_{i=1}^s |F(\hat{a}; x_i)| / s \quad (10)$$

式中, s 代表线段 L 上所有特征点的数目, $x_i = [x_i^2, x_i y_i, y_i^2, x_i y_i]^T$, $\hat{a} = [a, b, c, d, e, f]^T$, $F(\hat{a}; x_i)$ 代表点 (x_i, y_i) 到椭圆 $ax^2 + bxy + cy^2 + dx + ey + f = 0$ 的距离,根据 $F(\hat{a}; x_i) = \hat{a} \cdot x_i = ax_i^2 + bx_i y_i + cy_i^2 + dx_i + ey_i + f$ 计算得出。

分别计算线段 L_1 和 L_2 的 MAD_1 和 MAD_2 ,并将它们的最小值 MAD_{min} 与预先设定的阈值 T_{mad} 进行比较,如果 $MAD_{min} < T_{mad}$,表示至少有一个分段 L_1 或者 L_2 的拟合误差要小得多,此时可以认为对应的候选分裂点对是正确的。对于偏心率较大的单细胞,瓶颈检测可能会误检出一些分裂点对。此时在这种情况下,这两个条件的拟合误差都比较小,因此不能用 $MAD_{min} < T_{mad}$ 消除不正确的分裂点对。为了解决这一问题,增加了另一项限制为: $MAD_c < T_{mad}$,其中, MAD_c 表示从整个轮廓计算出的平均代数距离。从单个细胞中计算出的 MAD_c 值通常会比较小,因为它们近似于椭圆形状。单个细胞可以通过阈值 T_{mad} 来对比它们的拟合误差 MAD_c ,从而去除不正确的分裂点对。只有当 MAD_{min} 和 MAD_c 分别满足上述的

限制条件时,才能将对应区域视为重叠细胞区域,并判断此时的候选分裂点对是正确的分裂点对。

图4中显示了错误分割的例子和使用椭圆拟合校正分裂点对,其中实点代表细胞图像轮廓上的特征点,虚线代表使用特征点拟合成的椭圆形状,其中特征点 p_a 和 p_b 表示候选分裂点对。在图4的左半部分,拟合椭圆对特征点的拟合不正确,拟合椭圆计算得到的 MAD_{min} 值为0.031,即该细胞轮廓对应的区域不是一个单细胞,认为此时的候选分裂点对不正确,应予以丢弃。作为对比,给出了一个用正确的分裂点对拟合椭圆的例子,其中 $MAD_c = 0.171$, $MAD_{min} = 0.009$ 。可以看出,特征点与拟合椭圆的分布非常吻合,拟合椭圆可以帮助找到正确的分裂点对,从而能够正确分割相应的区域。如果当前的候选分裂点对判断为不正确,则将丢弃该候选分裂点对,并选择下一对进行判断,直到找到一个满足条件的分裂点对为止。如果所有的分裂点对都不满足条件,则表示该重叠区域其实是单个细胞区域。



图4 椭圆拟合校正
Fig.4 Correction by ellipse fitting

1.3 重叠细胞重合区域的分割

根据1.2节的算法,可以得到重叠细胞的正确分裂点对,由于宫颈细胞大多数呈现圆形或者似圆形。因此,我们可以断定重叠细胞的重合区域应当包含于以两个分裂点连接的线段为直径的圆内,如图5所示。

因为在梯度图像中,细胞边缘处的灰度值通常较高,并且能够很好地保留下细胞图像。1.2节中得到了重叠细胞的正确分裂点对,根据分裂点对又可以缩小重叠细胞重合的区域,优化算法难度。本研究先求出两个重叠细胞的重合部分的梯度图像,再采用分水岭算法对该重合区域的梯

研究 与 论 著

度图像进行分割，这样可以大大缩小图像处理的范围，从而可以提高分割的性能。但是，直接用水分岭算法处理梯度图像往往会出现过分割的情况，因此，本研究预先使用阈值算法对梯度图像进行处理，去除掉微弱（梯度较小）的边界信息，并将图像的梯度图像归一化到区域[0, 1.0]之间，再对梯度图像进行分水岭算法，从而可以求出重叠细胞重合区域的分割边界，再与预处理获得的细胞前景区域轮廓进行叠加，从而最后得到完整的重叠细胞的分割效果。

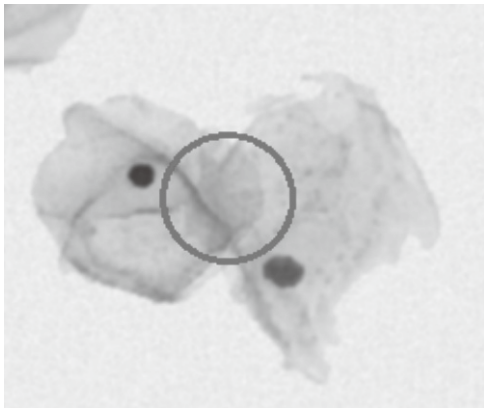


图5 重合区域的确定
Fig.5 Determination of coincidence region

由图6可以看出，该算法明显地去除了重叠细胞重合区域的微弱边界区域，同时为了解决过分割的问题，将重叠细胞的重合区域单独拿出来进行分割，得到重合区域的边界后，再叠加上细胞前景分割算法提取出的外细胞轮廓，构成了完整的重叠细胞的分割结果。

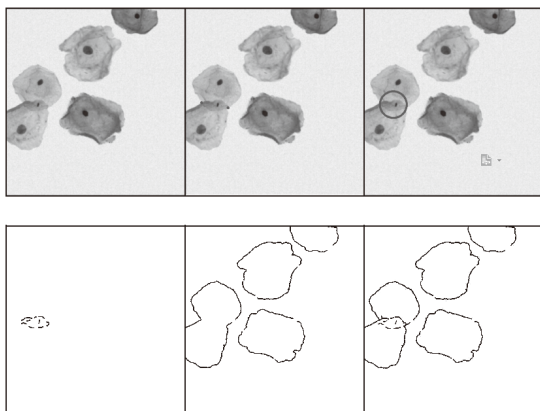


图6 分割过程
Fig.6 The segmentation process

综上所述，重叠细胞分割算法步骤如下：

输入：待处理图像f
输出：重叠细胞分割结果P
1 利用基于形态学闭包重建和隶属度滤波的分割算法得到细胞轮廓；
2 利用多边形近似求取轮廓上所有特征点；
3 利用瓶颈检测从特征点中找到候选的重叠细胞的分裂点对；
4 利用椭圆拟合从候选的分裂点对中找到正确的细胞分裂点对；
5 利用分水岭算法对分裂点对所确定的重合区域进行边界检测；
6 重合区域的边界与外轮廓叠加得到重

2 实验结果与分析

本实验使用华硕台式电脑（Intel i5-4590 3.3 GHz CPU处理器，8 GB内存，GTX1050Ti显卡），使用Matlab R2017a软件作为开发工具进行研究开发。

实验部分将50幅重叠的宫颈细胞显微图像进行分割，图像尺寸为250×250，为了说明算法的分割结果，将分割结果与医生手工分割结果进行比较，并与文献^[9]的分割结果进行对比。部分分割结果如图7所示，其中第一行为原始输入图像，第二行为本研究算法的分割结果，第三行为医生标注的重叠细胞的分割结果。

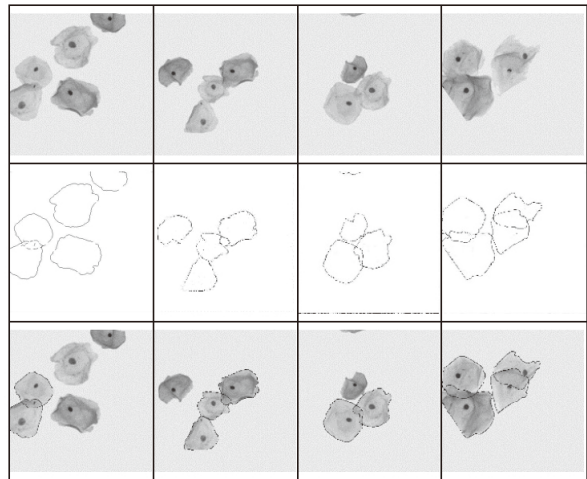


图7 重叠细胞分割结果
Fig.7 Overlapping cell segmentation results

从图7的分割结果可以得出，本研究算法能够将重叠的宫颈细胞图像有效分离开来，实现了高重叠细胞的显微图像分割，且分割效果接近于医生的手工分割结果。针对重叠区域轮廓较为明显的重叠细胞，本研究的分割算法较好，前三列的分割结果与医生手工分割结果接近，误差较小。最后一列的重叠区域颜色较深，轮廓较为模

糊,难以分割准确。但相对文献[9]已经有了一定的提高。本研究采用分割区域面积的误差比 w 来定量评价分割结果,算式为:

$$w = \frac{|area_s - area_r|}{area_r} \quad (11)$$

式中, $area_s$ 为本研究分割出的区域面积, $area_r$ 为医生手工分割的区域面积。

本研究选取目前国内外主流的宫颈细胞图像分割算法进行实验,并将他们的结果进行对比分析,如表1所示。

表1 本研究算法与其它算法的比较
Tab.1 Comparison of proposed algorithm and other algorithms

算法	误差比 w	运行时间/s
本研究算法	0.08	3.2
文献[9]	0.15	3.0
文献[12]	0.12	3.8
文献[13]	0.20	2.3

由表1可知,虽然本研究算法运行时间较长,牺牲了一定的运行效率。但本研究算法的分割结果的面积误差为0.08,是四种算法中效果最好的,说明与医生的手工分割结果较为接近。

3 讨论和结论

针对重叠的宫颈细胞图像重叠情况复杂、难以分割的问题。本研究提出了一种基于瓶颈检测和分水岭算法的分割算法。首先,通过细胞前景区域的分割算法获得整个重叠细胞的轮廓。其次,通过瓶颈检测和椭圆拟合来检测重叠细胞轮廓的分裂点,从分裂点中判断出正确分割图像的分裂点对。再次,根据分裂点进一步缩小重叠细胞重合的区域范围,降低处理难度。最终通过分水岭算法获得重叠区域的边界信息。将重叠区域的边界与外轮廓叠加起来,获得重叠细胞的分割结果。将该算法应用于重叠的宫颈细胞图像分割,能够分割出准确完整的单个宫颈细胞图像,且与医生手动标注的分割结果接近,与最新的其他基于瓶颈检测的分割算法相比,误差更小,更为准确。但是牺牲了一定的运行效率,且当重叠区域非常复杂时,仍然存在较大的分割误差,如何解决这两个问题有待进一步的研究改进。

参考文献

- [1] HUH W K, AULT K A, CHELMOW D, et al. Use of primary high-risk human papillomavirus testing for cervical cancer screening: Interim clinical guidance[J]. Gynecol Oncol, 2015, 136(2): 178-182.
- [2] CRAMER D W. The role of cervical cytology in the declining morbidity and mortality of cervical cancer[J]. Cancer, 2015, 34(6): 2018-2027.
- [3] TOTA J E, BENTLEY J, BLAKE J, et al. Introduction of molecular HPV testing as the primary technology in cervical cancer screening: Acting on evidence to change the current paradigm[J]. Prev Med, 2017, 98: 5-14.
- [4] ZHOU Y, HAO Y, LI Y, et al. Amplification and up-regulation of MIR30D was associated with disease progression of cervical squamous cell carcinomas[J]. Bmc Cancer, 2017, 17(1): 230.
- [5] TAREEF A, YANG S, CAI W, et al. Automatic segmentation of overlapping cervical smear cells based on local distinctive features and guided shape deformation[J]. Neurocomputing, 2017, 221: 94-107.
- [6] SANYAL P, GANGULI P, BARUI S, et al. Pilot study of an open-source image analysis software for automated screening of conventional cervical smears[J]. J Cytol, 2018, 35(2): 71-74.
- [7] 张珊珊. 宫颈LCT图像分割算法的研究与实现[D]. 广州: 华南理工大学, 2013.
- [8] RANNOU F, GREGOR J. Equilateral polygon approximation of closed contours [J]. Pattern Recogn, 2015, 29(7): 1105-1115.
- [9] 张丙奎. 基于瓶颈检测和Dijkstra算法的重叠细胞分割方法 [J]. 技术与市场, 2016, 23(12): 174-175.
- [10] LIAO M, ZHAO Y Q, LI X H, et al. Automatic segmentation for cell images based on bottleneck detection and ellipse fitting[J]. Neurocomputing, 2016, 173(P3): 615-622.
- [11] 张竣凯. 基于瓶颈点探测与椭圆拟合的细胞图像分割方法探究[J]. 电子测试, 2013(10): 112-114.
- [12] ZHAO M, WU A, SONG J, et al. Automatic screening of cervical cells using block image processing[J]. Biomed Eng Online, 2016, 15(1): 14.
- [13] TAREEF A, SONG Y, HUANG H, et al. Multi-pass fast watershed for accurate segmentation of overlapping cervical cells[J]. IEEE T Med Imaging, 2018, 37(9): 2044-2059.