



(10) 授权公告号 CN 112470166 B

(45) 授权公告日 2024. 08. 23

(21) 申请号 201980000985.9

(22) 申请日 2019.06.18

(65) 同一申请的已公布的文献号
申请公布号 CN 112470166 A

(43) 申请公布日 2021.03.09

(85) PCT国际申请进入国家阶段日
2019.07.03(86) PCT国际申请的申请数据
PCT/CN2019/091761 2019.06.18(87) PCT国际申请的公布数据
W02020/252665 ZH 2020.12.24(73) 专利权人 香港理工大学
地址 中国香港九龙

(72) 发明人 宋有义 秦璟

(74) 专利代理机构 北京世峰知识产权代理有限公司 11713

专利代理师 卓霖 许向彤

(51) Int.Cl.
G06V 20/69 (2022.01)
G06V 10/40 (2022.01)(56) 对比文件
Youyi Song et al..Segmentation of
Overlapping Cytoplasm in Cervical Smear
Images via Adaptive Shape Priors
Extracted From Contour Fragments.IEEE
TRANSACTIONS ON MEDICAL IMAGING.2019,第38
卷(第12期),第2849-2862页.

审查员 刘玉莹

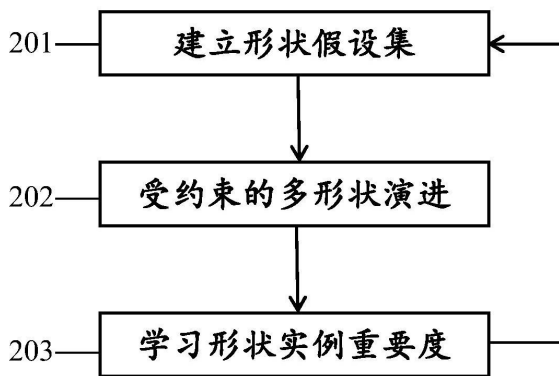
权利要求书4页 说明书10页 附图5页

(54) 发明名称

用于分割医学图像中的重叠细胞质的方法
及系统

(57) 摘要

一种分割医学图像中的重叠细胞质的方法,其包括:建立细胞质形状假设集(201);以及从建立的细胞质形状假设集中为每个细胞质选择形状假设,来进行受约束的多形状演进(202),从而分割医学图像中的重叠细胞质,其中,受约束的多形状演进(202)包括:对由多个重叠细胞质构成的团块区域进行分割以提供团块证据(301);用于评估选择的形状假设的质量的形状对齐(302);以及为每个细胞质确定更好的形状假设的形状演进(303)。还涉及一种分割医学图像中的重叠细胞质的系统。系统和方法能够更精确、更有效地基于先验形状进行重叠细胞质的分割。



1. 一种分割医学图像中的重叠细胞质的方法,其包括:

建立细胞质形状假设集;以及

从建立的细胞质形状假设集中为每个细胞质选择形状假设,来执行受约束的多形状演进,从而分割医学图像中的重叠细胞质,

其中,所述受约束的多形状演进包括以下步骤:

团块区域分割,其对由多个重叠细胞质构成的所述医学图像中的团块区域进行分割以提供团块证据;

形状对齐,其用于评估选择的形状假设的质量;以及

形状演进,其为每个细胞质确定更好的形状假设,

其中,所述形状对齐包括填充从细胞质形状假设集中选择的形状假设以获得二值图像,并利用下述公式来获得所述二值图像与对应的细胞质对齐所需要旋转的角度和缩放的大小:

$$\operatorname{argmax}(B_i \cap B_c), \text{ s.t. } B_i \subset B_c$$

其中, B_c 表示分割后的团块区域的图像; B_i 表示对齐的结果图像, B_i 是与 B_c 具有相同尺寸的二值图像,且需要 B_i 在 B_c 的内部。

2. 根据权利要求1所述的方法,还包括为所建立的细胞质形状假设集更新细胞质形状实例的学习形状实例重要度的步骤。

3. 根据权利要求1所述的方法,其中,所述形状对齐和所述形状演进是迭代进行的,形状对齐的输出作为形状演进的输入,而形状演进的输出作为形状对齐的输入,所述形状对齐评估是否需要开始新的形状演进,一旦确定不需要新的形状演进,则形状对齐后的当前的形状假设就是该重叠细胞质的分割结果。

4. 根据权利要求1所述的方法,其中,由以下公式建立所述细胞质形状假设集:

$$\mathcal{H}_s = \{s_i : s_i = \mu + Mx_i, i \in \mathbb{N}\}$$

其中, μ 表示收集的形状实例的平均形状, Mx_i 表示收集的形状的协方差矩阵的特征向量的线性组合,其中,矩阵M的每一列是一个特征向量, x_i 是该线性组合的权重向量, s_i 表示标号为i的形状假设。

5. 根据权利要求1所述的方法,其中,所述形状演进包括:

- 设定目标函数,所述目标函数为:

$$\mathbb{E}(B_c, \mathbf{x}^k) = \sum_{(x,y) \in \Omega_B} (B_g(x,y) - B_c(x,y))^2$$

其中, \mathbf{x} 表示集合 $\{\mathbf{x}_i\}_{i=1}^N$, \mathbf{x}^k 表示第k次演进的 \mathbf{x} ;N表示所述团块区域中细胞质的数量;

$B_g = \bigcup_{i=1}^N B_i$,是由对齐结果 $\{B_i\}_{i=1}^N$ 生成的二值图像; (x,y) 表示图像中像素的坐标;以及

- 确定能够使 \mathbb{E} 具有比 \mathbf{x}^k 更低的值的 \mathbf{x}^{k+1} 。

6. 根据权利要求5所述的方法,其中,所述确定能够使 \mathbb{E} 具有比 \mathbf{x}^k 更低的值的 \mathbf{x}^{k+1} 包括:

- 对于从所述目标函数得到的矩阵p,使用泰勒定理得到公式:

$$\mathbb{E}(\mathbf{x}^k + \mathbf{p}) = \mathbb{E}(\mathbf{x}^k) + \nabla \mathbb{E}(\mathbf{x}^k)^T \mathbf{p} + \frac{1}{2} \mathbf{p}^T \nabla^2 \mathbb{E}(\mathbf{x}^k) \mathbf{p}$$

其中, ∇ 和 ∇^2 分别表示梯度和海森矩阵计算; γ 是间隔 (0, 1) 之间的一个标量, 从而获得以 \mathbf{x}^k 为圆心、以 $\|\mathbf{p}\|_2$ 为半径的区域内的 \mathbb{E} 的整体区域最小值;

- 回到 \mathbf{x}^k 处近似 \mathbb{E} : $m_k(\mathbf{p}) = \mathbb{E}(\mathbf{x}^k) + \nabla \mathbb{E}(\mathbf{x}^k)^T \mathbf{p} + \frac{1}{2} \mathbf{p}^T \nabla^2 \mathbb{E}(\mathbf{x}^k) \mathbf{p}$ 其中, m_k 表示 \mathbf{x}^k 处的 \mathbb{E} ;

- 由 m_k 的最小值来近似 \mathbb{E} 的最小值, 并通过信赖域方法来解公式:

$$\mathbf{p}^* = \underset{\mathbf{p} \in \Omega_p}{\operatorname{argmin}} m_k(\mathbf{p}), \text{ s.t. } \|\mathbf{p}\|_2 \leq \Delta_k; \text{ 以及}$$

- 得到第 k 次演进的输出结果: $\mathbf{x}^{k+1} = \mathbf{x}^k + \mathbf{p}^*$ 。

7. 根据权利要求2所述的方法, 其中, 所述学习形状实例重要度的步骤包括:

- 随机选择一组形状实例, 根据以下公式计算所述一组形状实例的平均形状:

$$\mu = \frac{1}{W} \sum_{i=1}^{K_s} w_i s_i$$

其中, K_s 是选择的形状实例的数量; w_i 是每个形状实例 s_i 的重要度; W 是所有 w_i 的和; 并且,

- 根据得到的 μ 计算协方差矩阵:

$$\mathbf{M}_c = \frac{1}{K_s} \sum_{i=1}^{K_s} (s_i - \mu)(s_i - \mu)^T$$

其中, 矩阵 \mathbf{M}_c 的第一 t 个特征向量构成矩阵 $\mathbf{M} = (\mathbf{e}_1 \ \mathbf{e}_2 \ \cdots \ \mathbf{e}_t)$, 其对应的特征值为 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_t$ 。

8. 根据权利要求7所述的方法, 其中, 如果由所述受约束的多形状演进得到的分割结果大于预定阈值, 则由所述学习形状实例重要度的步骤重新计算, 以更新所述形状假设集; 直到所述分割结果不能再减小或达到了所述预定阈值, 才停止所述更新。

9. 一种用于分割医学图像中的重叠细胞质的系统, 其包括:

用于建立细胞质形状假设集的形状假设集模块; 以及

多形状演进模块, 其用于从建立的细胞质形状假设集中为每个细胞质选择形状假设, 来进行受约束的多形状演进, 从而分割医学图像中的重叠细胞质,

其中, 所述多形状演进模块被配置为执行以下步骤:

团块区域分割, 其对由多个重叠细胞质构成的所述医学图像中的团块区域进行分割以提供团块证据;

形状对齐, 其用于评估选择的形状假设的质量; 以及

形状演进, 其为每个细胞质确定更好的形状假设,

其中, 所述形状对齐包括填充从细胞质形状假设集中选择的形状假设以获得二值图像, 并利用下述公式来获得所述二值图像与对应的细胞质对齐所需要旋转的角度和缩放的大小:

$$\operatorname{argmax}(B_i \cap B_c), \text{ s.t. } B_i \subset B_c$$

其中, B_c 表示分割后的团块区域的图像; B_i 表示对齐的结果图像, B_i 是与 B_c 具有相同尺寸的二值图像, 且需要 B_i 在 B_c 的内部。

10. 根据权利要求9所述的系统, 还包括为所述建立的细胞质形状假设集更新细胞质形状实例的学习形状实例重要度模块。

11. 根据权利要求9所述的系统, 其中, 所述形状对齐和所述形状演进是迭代进行的, 形状对齐的输出作为形状演进的输入, 而形状演进的输出作为形状对齐的输入, 所述形状对齐评估是否需要开始新的形状演进, 一旦确定不需要新的形状演进, 则形状对齐后的当前的形状假设就是该重叠细胞质的分割结果。

12. 根据权利要求9所述的系统, 其中, 由以下公式建立所述细胞质形状假设集:

$$\mathcal{H}_s = \{s_i : s_i = \mu + \mathbf{M}\mathbf{x}_i, i \in \mathbb{N}\}$$

其中, μ 表示收集的形状实例的平均形状, $\mathbf{M}\mathbf{x}_i$ 表示收集的形状的协方差矩阵的特征向量的线性组合, 其中, 矩阵 \mathbf{M} 的每一列是一个特征向量, \mathbf{x}_i 是该线性组合的权重向量, s_i 表示标号为 i 的形状假设。

13. 根据权利要求9所述的系统, 其中, 所述形状演进包括:

- 设定目标函数, 所述目标函数为:

$$\mathbb{E}(B_c, \mathbf{x}^k) = \sum_{(x,y) \in \Omega_B} (B_g(x,y) - B_c(x,y))^2$$

其中, \mathbf{x} 表示集合 $\{\mathbf{x}_i\}_{i=1}^N$, \mathbf{x}^k 表示第 k 次演进的 \mathbf{x} ; N 表示所述团块区域中细胞质的数量;

$B_g = \bigcup_{i=1}^N B_i$, 是由对齐结果 $\{B_i\}_{i=1}^N$ 生成的二值图像; (x, y) 表示图像中像素的坐标; 以及

- 确定能够使 \mathbb{E} 具有比 \mathbf{x}^k 更低的值的 \mathbf{x}^{k+1} 。

14. 根据权利要求13所述的系统, 其中, 所述确定能够使 \mathbb{E} 具有比 \mathbf{x}^k 更低的值的 \mathbf{x}^{k+1} 包括:

- 对于从所述目标函数得到的矩阵 \mathbf{p} , 使用泰勒定理得到公式:

$$\mathbb{E}(\mathbf{x}^k + \mathbf{p}) = \mathbb{E}(\mathbf{x}^k) + \nabla \mathbb{E}(\mathbf{x}^k)^T \mathbf{p} + \frac{1}{2} \mathbf{p}^T \nabla^2 \mathbb{E}(\mathbf{x}^k) \mathbf{p}$$

其中, ∇ 和 ∇^2 分别表示梯度和海森矩阵计算; γ 是间隔 $(0, 1)$ 之间的一个标量, 从而获得以 \mathbf{x}^k 为圆心、以 $\|\mathbf{p}\|_2$ 为半径的区域内的 \mathbb{E} 的整体区域最小值;

- 回到 \mathbf{x}^k 处近似 \mathbb{E} : $m_k(\mathbf{p}) = \mathbb{E}(\mathbf{x}^k) + \nabla \mathbb{E}(\mathbf{x}^k)^T \mathbf{p} + \frac{1}{2} \mathbf{p}^T \nabla^2 \mathbb{E}(\mathbf{x}^k) \mathbf{p}$ 其中, m_k 表示 \mathbf{x}^k 处的 \mathbb{E} ;

- 由 m_k 的最小值来近似 \mathbb{E} 的最小值, 并通过信赖域方法来解公式:

$$\mathbf{p}^* = \operatorname{argmin}_{\mathbf{p} \in \Omega_p} m_k(\mathbf{p}), \text{ s.t. } \|\mathbf{p}\|_2 \leq \Delta_k; \text{ 以及}$$

- 得到第 k 次演进的输出结果: $\mathbf{x}^{k+1} = \mathbf{x}^k + \mathbf{p}^*$ 。

15. 根据权利要求10所述的系统, 其中, 所述学习形状实例重要度模块被配置为:

-随机选择一组形状实例,根据以下公式计算所述一组形状实例的平均形状:

$$\mu = \frac{1}{W} \sum_{i=1}^{K_s} w_i s_i$$

其中, K_s 是选择的形状实例的数量; w_i 是每个形状实例 s_i 的重要度; W 是所有 w_i 的和;并且,

-根据得到的 μ 计算协方差矩阵:

$$\mathbf{M}_c = \frac{1}{K_s} \sum_{i=1}^{K_s} (s_i - \mu)(s_i - \mu)^T$$

其中,矩阵 \mathbf{M}_c 的第一 t 个特征向量构成矩阵 $\mathbf{M} = (e_1 \ e_2 \ \cdots \ e_t)$,其对应的特征值为 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_t$ 。

16.根据权利要求15所述的系统,其中,如果由所述受约束的多形状演进模块得到的分割结果大于预定阈值,则由所述学习形状实例重要度模块重新计算,以更新所述形状假设集;直到所述分割结果不能再减小或达到了所述预定阈值,才停止所述更新。

用于分割医学图像中的重叠细胞质的方法及系统

技术领域

[0001] 本发明涉及一种用于分割医学图像中的重叠细胞质的方法及系统,更具体地,涉及一种用于宫颈癌筛查的医学图像中的重叠细胞质的分割方法及系统。

背景技术

[0002] 宫颈癌在女性的恶性肿瘤中死亡率居第四位,也是发病率排名第四的癌症,而高质量的宫颈癌筛查能够极大地减少宫颈癌的发病率和死亡率。具体而言,宫颈癌筛查是指在显微镜下检查从宫颈处取样并放置在玻片上的每个宫颈细胞的异常程度,从而评估是否存在宫颈癌细胞。

[0003] 对于发展宫颈癌自动筛查系统来说,对宫颈图像中的细胞的重叠的细胞质进行分割是其中很关键的任务,因为为了检查细胞的异常程度,提取细胞水平的特征(例如,细胞的形状、尺寸、和细胞质与细胞核的面积比)在临床上是很重要的,然而,由于在重叠区域中强度(或颜色)信息通常是混乱的,甚至会让人产生误解,因此,这种强度(或颜色)信息的不足导致这项任务是非常挑战的。

[0004] 传统的分割医学图像中的重叠细胞质的方法是利用(leveraging)团块中的各细胞的细胞质之间的强度信息或者结合空间信息来实现的。普遍通过扩展典型的分割模型来实现这个目的,例如采用阈值分割法、分水岭分割法和图像切割法。这些方法在理论上假设了强度信息足以识别被遮挡的边界部分。然而,这种假设是有缺陷的,实际上,重叠区域的强度信息通常是混乱的,甚至会让人误解。

[0005] 为了消除强度信息不足带来的问题,基于先验形状(shape prior-based)的方法显现出了很好的分割表现,其将额外的形状信息插入到分割方法中。而这些方法要么通过简单的形状推测(例如细胞质具有椭圆形或星形的形状)、要么从预先收集的有限的形状集合中匹配形状实例,来建模先验形状。然后,利用强度信息,同时需要分割的结果与建模的先验形状尽可能地相似,来分割重叠的细胞质。通常采用主动轮廓模型或水平集模型来实现,其中,将先验形状设计成能量函数中的正则项,假设该能量函数的最小(或最大)值由产生最佳分割结果的分割函数获得。

[0006] 尽管现有技术已改进了分割的精确度,但是这些现存的先验形状方法仍然具有三个主要的缺点。首先,这些方法都是通过有限的形状假设(例如,形状推测或收集的形状实例)来建模先验形状,由于这些具体的形状假设不能很好地还原细胞质被遮挡的边界部分,导致这些方法建模的先验形状通常不足以查明细胞质被遮挡的边界部分。其次,这些方法仅仅使用了局部先验形状(即,先验信息仅仅是单个的细胞质形状)来演进细胞质的形状,而没有考虑所有细胞质和团块之间的形状关系,这导致这些方法的分割结果通常与团块证据不一致,被分割的细胞质边界与团块的理想边界有偏差。再次,尽管这些方法要求最终形状与建模的先验形状尽可能地相似,但是这些方法没有对最终形状施加形状约束,实际上,这些方法试图通过平衡参数在强度证据和局部先验形状之间找到合适的妥协,因此,当强度证据与局部先验形状相矛盾时,这些方法会产生令人难以置信的分割结果。

[0007] 此外,对于无限形状假设集的建立,现有技术中(例如,T.F.Cootes,C.J.Taylor, and et al.D.H.Cooper.“Active shape models-their training and application”, Computer Vision and Image Understanding,61(1):38-59,1995)公开了相关内容。虽然采用该方法能够通过选择无限的数值来得到无限大的形状假设集,但是其主要缺点在于:很难收集一组形状实例,使得能确保不可见的细胞质形状被建立的形状假设很好地还原。现有技术中,建立形状假设集依赖于如何收集到好的形状实例,通常,不得不采用手动选择具有代表性的形状实例,这是一种实验性并具有误差的方式,也是一种劳动密集型方式,当需要收集无限多的实例用来近似复杂的形状时,这种方法是不可行的。

发明内容

[0008] 因此,本发明要解决的一个技术问题是:提供一种能够更精确、更有效地基于先验形状的分割医学图像中的重叠细胞质的方法及系统。

[0009] 在一个实施方式中,本发明涉及一种分割医学图像中的重叠细胞质的方法,其包括:建立细胞质形状假设集;以及从建立的细胞质形状假设集中为每个细胞质选择形状假设,来进行受约束的多形状演进,从而分割医学图像中的重叠细胞质,其中,所述受约束的多形状演进包括:对由多个重叠细胞质构成的团块区域进行分割以提供团块证据;用于评估选择的形状假设的质量的形状对齐;以及为每个细胞质确定更好的形状假设的形状演进。

[0010] 所述分割医学图像中的重叠细胞质的方法还包括为所述建立的细胞质形状假设集更新细胞质形状实例的学习形状实例重要度的步骤。

[0011] 优选地,所述形状对齐和所述形状演进是迭代进行的,形状对齐的输出作为形状演进的输入,而形状演进的输出作为形状对齐的输入,所述形状对齐评估是否需要开始新的形状演进,一旦确定不需要新的形状演进,则形状对齐后的当前的形状假设就是该重叠细胞质的分割结果。

[0012] 优选地,由以下公式建立所述细胞质形状假设集:

$$[0013] \quad \mathcal{H}_s = \{s_i : s_i = \mu + Mx_i, i \in N\}$$

[0014] 其中, μ 表示收集的形状实例的平均形状, Mx_i 表示收集的形状的协方差矩阵的特征向量的线性组合,其中,矩阵M的每一列是一个特征向量, x_i 是该线性组合的权重向量, s_i 表示标号为i的形状假设。

[0015] 优选地,所述形状对齐包括填充从细胞质形状假设集中选择的形状假设以获得二值图像,并利用下述公式来获得所述二值图像与对应的细胞质对齐所需要旋转的角度和缩放的大小:

$$[0016] \quad \operatorname{argmax}(B_i \cap B_c), \text{ s.t. } B_i \subset B_c$$

[0017] 其中, B_c 表示分割后的团块区域的图像; B_i 表示对齐的结果图像, B_i 是与 B_c 具有相同尺寸的二值图像,且需要 B_i 在 B_c 的内部。

[0018] 优选地,所述形状演进包括:设定目标函数,所述目标函数为:

$$[0019] \quad \mathbb{E}(B_c, x^k) = \sum_{(x,y) \in \Omega_B} (B_g(x,y) - B_c(x,y))^2$$

[0020] 其中, x 表示集合 $\{x_i\}_{i=1}^N$, x^k 表示第 k 次演进的 x ; N 表示所述团块区域中细胞质的数量; $B_g = \bigcup_{i=1}^N B_i$, 是由对齐结果 $\{B_i\}_{i=1}^N$ 生成的二值图像; (x, y) 表示图像中像素的坐标; 以及确定能够使 \mathbb{E} 具有比 x^k 更低的值的 x^{k+1} 。

[0021] 优选地, 所述确定能够使 \mathbb{E} 具有比 x^k 更低的值的 x^{k+1} 包括: 对于从所述目标函数得到的矩阵 p , 使用泰勒定理得到公式:

$$\mathbb{E}(x^k + p) = \mathbb{E}(x^k) + \nabla \mathbb{E}(x^k)^T p + \frac{1}{2} p^T \nabla^2 \mathbb{E}(x^k) p$$

[0022] 其中, ∇ 和 ∇^2 分别表示梯度和海森矩阵计算; γ 是间隔 $(0, 1)$ 之间的一个标量, 从而获得以 x^k 为圆心、以 $\|p\|_2$ 为半径的区域内的 \mathbb{E} 的整体区域最小值; 回到 x^k 处近似

$$\mathbb{E}: m_k(p) = \mathbb{E}(x^k) + \nabla \mathbb{E}(x^k)^T p + \frac{1}{2} p^T \nabla^2 \mathbb{E}(x^k) p$$

[0023] 其中, m_k 表示 x^k 处的 \mathbb{E} ; 由 m_k 的最小值来近似 \mathbb{E} 的最小值, 并通过信赖域方法来解公式:

[0024] $p^* = \underset{p \in \Omega_p}{\operatorname{argmin}} m_k(p), \text{ s.t. } \|p\|_2 \leq \Delta_k$, 以及得到第 k 次演进的输出结果: $x^{k+1} = x^k + p^*$ 。

[0025] 优选地, 所述学习形状实例重要度的步骤包括: 随机选择一组形状实例, 根据以下公式计算所述一组形状实例的平均形状:

$$\mu = \frac{1}{W} \sum_{i=1}^{K_s} w_i s_i$$

[0027] 其中, K_s 是选择的形状实例的数量; w_i 是每个形状实例 s_i 的重要度; W 是所有 w_i 的和; 并且, 根据得到的 μ 计算协方差矩阵:

$$M_c = \frac{1}{K_s} \sum_{i=1}^{K_s} (s_i - \mu)(s_i - \mu)^T$$

[0029] 其中, 矩阵 M_c 的第一 t 个特征向量构成矩阵 $M = (e_1 \ e_2 \ \cdots \ e_t)$, 其对应的特征值为 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_t$ 。

[0030] 优选地, 如果由所述受约束的多形状演进得到的分割结果大于预定阈值, 则由所述学习形状实例重要度的步骤重新计算, 以更新所述形状假设集; 直到所述分割结果不能再减小或达到了所述预定阈值, 才停止所述更新。

[0031] 在另一方面, 本发明还涉及一种用于分割医学图像中的重叠细胞质的系统, 其包括: 用于建立细胞质形状假设集的形状假设集模块; 以及多形状演进模块, 其用于从建立的细胞质形状假设集中为每个细胞质选择形状假设, 来进行受约束的多形状演进, 从而分割医学图像中的重叠细胞质, 其中, 所述多形状演进模块被配置为: 对由多个重叠细胞质构成的团块区域进行分割以提供团块证据; 用于评估选择的形状假设的质量的形状对齐; 以及为每个细胞质确定更好的形状假设的形状演进。

[0032] 优选地,所述用于分割医学图像中的重叠细胞质的系统还包括为所述建立的细胞质形状假设集更新细胞质形状实例的学习形状实例重要度模块。

[0033] 优选地,所述学习形状实例重要度模块被配置为:随机选择一组形状实例,根据以下公式计算所述一组形状实例的平均形状:

$$[0034] \quad \mu = \frac{1}{W} \sum_{i=1}^{K_s} w_i s_i$$

[0035] 其中, K_s 是选择的形状实例的数量; w_i 是每个形状实例 s_i 的重要度; W 是所有 w_i 的和;并且,根据得到的 μ 计算协方差矩阵:

$$[0036] \quad M_c = \frac{1}{K_s} \sum_{i=1}^{K_s} (s_i - \mu)(s_i - \mu)^T$$

[0037] 其中,矩阵 M_c 的第一 t 个特征向量构成矩阵 $M = (e_1 \ e_2 \ \cdots \ e_t)$,其对应的特征值为 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_t$ 。

[0038] 优选地,如果由所述受约束的多形状演进模块得到的分割结果大于预定阈值,则由所述学习形状实例重要度模块重新计算,以更新所述形状假设集;直到所述分割结果不能再减小或达到了所述预定阈值,才停止所述更新。本发明的方法和系统通过无限大的形状假设集来建模先验形状,同时将局部先验形状和整体先验形状与强度信息合并用于演进,并且将每次演进中的结果形状约束在形状假设集中。相比于现有的分割医学图像中的重叠细胞质的方法和系统,本发明的方法和系统能够更好的查明被遮挡的边界部分,由此更好的分割重叠的细胞质,从而为医学诊断提供了更加精确的形状特征。本发明建立的无限大的形状假设集能够更好的描述细胞质所有可能的形状,由此,更有效率地分割具有不同形状的重叠细胞质;本发明受约束的多形状演进算法通过考虑所有细胞质和整个团块的形状关系,来将局部先验形状和整体先验形状与强度信息合并用于演进,从而获得更多的信息用于分割;本发明在形状统计计算中利用了每个形状实例的重要度,从而是不可见的形状都能由形状假设集中的形状假设来很好地近似;本发明的多形状演进步骤中植入学习步骤,能够更加有效地获得有用的信息。因此,相比于现有技术,本发明的分割医学图像中的重叠细胞质的方法和系统能更有效地获得更精确的结果。

附图说明

[0039] 通过附图以及下面的描述,可以更好地理解本发明的技术方案,其中:

[0040] 图1a-1d示出了根据本发明的方法分割具有重叠的细胞质的团块的示意图;

[0041] 图2示出了本发明的方法的流程图;

[0042] 图3示出了本发明的受约束的多形状演进步骤的流程图;

[0043] 图4示出了本发明的重叠细胞质分割系统的示意图;以及

[0044] 图5示出了根据本发明的方法和现有技术的方法得到的分割结果的比较图像

具体实施方式

[0045] 本发明通过演进由建模的局部先验形状和整体先验形状引导的细胞质形状、同时演进细胞质的相互形状约束,来分割团块中的所有重叠细胞质,使得本发明的基于先验形

状的分割医学图像中的重叠细胞质的方法能够精确而高效地获得细胞质的分割结果,从而提高了宫颈癌筛查的精确性和效率。具体而言,通过利用统计的形状信息建模局部先验形状,从而建立具有无限大的细胞质的形状假设的形状假设集;在多形状演进步骤中,除了考虑局部先验形状,本发明通过演进细胞质的形状、继而使用算法使分割结果与团块证据一致,从而获得整体先验形状;此外,在多形状演进步骤中,要求演进过程得到的最终形状在形状假设集中,从而减少了现有技术中难以置信的分割结果;而为了使建立的形状假设更好地还原任何不可见的细胞质形状,本发明还在形状统计计算中增加了学习形状实例重要度的步骤。

[0046] 本发明采用以下新的算法和步骤来实现新的分割医学图像中的重叠细胞质的团块的方法。图1a-1d示出了采用本发明的方法分割具有重叠细胞质的团块的示意图,。其中,图1a示出了输入的团块图像;图1b示出了最初被分割的团块区域;图1c示出了在随机演进中的偏差区域11和由所有对齐后的形状覆盖的区域12;图1d示出了采用本发明的方法得到的分割结果。图2示出了本发明的方法的流程图,该方法包括三个步骤:建立形状假设集步骤201、受约束的多形状演进步骤202,以及学习形状实例重要度步骤203。

[0047] 以下内容对于上述建立形状假设集步骤201、受约束的多形状演进步骤202以及学习形状实例重要度步骤203三个步骤进行详细的描述。

[0048] 建立形状假设集

[0049] 首先,对细胞的细胞质的形状进行参数化。采用向量 s 形式的边界点来描述每个细胞的细胞质的形状。第 k 项 s 存储极坐标系中角度值等于 k 的边界点的距离值,该极坐标系的原点位于细胞核的质心。需要注意的是,每个细胞都由细胞质和细胞核组成,本发明选择细胞核的质心而不是细胞质的质心作为极坐标系的原点是出于可实施性的考虑,因为当各细胞的细胞质重叠时(见图1a),相比于检测细胞质的质心,检测细胞核的质心会容易地多。

[0050] 其次,根据现有技术已有的建立无限形状假设集的方法,利用细胞质统计形状信息建立形状假设集。在本发明中,将形状假设集表示为以下的公式(1):

$$[0051] \quad \mathcal{H}_s = \{s_i : s_i = \mu + Mx_i, i \in \mathbb{N}\} \quad (1)$$

[0052] 其中 μ 表示收集的形状实例的平均形状, Mx_i 表示收集的形状的协方差矩阵的特征向量的线性组合(其中,矩阵 M 的每一列是一个特征向量, x_i 是该线性组合的权重向量), s_i 表示标号为 i 的形状假设。

[0053] 通过选取不同值的 x_i 代入公式(1)中,可以获得不同的形状假设 s_i ,由于可以选择无穷多的 x_i ,因此,可以建立无穷大的形状假设集,但是依靠公式(1)建立的形状假设集很难收集到能够很好地还原不可见的细胞质形状的形状实例,而本发明通过在形状统计计算中运用以下将描述的学习形状实例重要度的步骤,克服了这一缺陷。

[0054] 受约束的多形状演进

[0055] 图3是本发明的受约束的多形状演进的流程图。本发明的受约束的多形状演进步骤是基于团块证据、同时从建立的形状假设集中为每个细胞质选择形状假设,来分割重叠的细胞质。其包括三个步骤:用于提供团块证据的团块区域分割(步骤301);评估当前形状假设的质量的形状对齐(步骤302);以及为每个细胞质寻找更好的形状假设的形状演进(步骤303)。其中形状对齐(步骤302)和形状演进(步骤303)是迭代进行的,形状演进(步骤303)将形状对齐的结果作为输入,而形状演进(步骤303)的输出作为形状对齐的输入,从而检测

是否需要开始新一轮的形状演进。一旦确定不需要新的形状演进,则经形状对齐的当前的形状假设就被视为该细胞质的分割结果(见图1d)。

[0056] 在团块区域分割(步骤301)中,本发明采用了多尺寸卷积神经网络(CNN)来分割细胞质和细胞核区域(见图1b)。该多尺寸卷积神经网络将图像中的每个像素分成三组:细胞核部分、细胞质部分和背景部分。其包含三个平行CNN,每个CNN在输入块(input patch)的分辨率方面具有不同尺寸。然后将这三个CNN的输出进行融合,以助于以不同的尺寸捕获更情境化(contextual)的信息。最后,采用马可夫随机场来进一步优化分割结果。

[0057] 在形状对齐(步骤302)中,对于形状假设 s_i ,由于其仅仅是存储边界点信息的矢量,因此,需要通过填充 s_i 描述的轮廓内部的区域,来获得 s_i 相应的二值图像(即,二进制图像),该轮廓内部的像素被标为1,轮廓外部的像素被标为0。如上所述, s_i 被赋值为形状演进步骤的输出,但从形状假设集中收集的实例的平均形状作为每个细胞质的初始 s_i 。此外,由于本发明可以通过以下将描述的演进形状假设来绕开非刚性变换,所以本发明将形状对齐限定为刚性对齐。

[0058] 具体而言,对于每个 s_i ,首先将 s_i 被填充的区域的质心与图像中的细胞核的质心对齐。然后,通过以下公式(2)得到用于对齐的缩放因子(r_i)和旋转系数 θ_i 。

$$[0059] \quad \operatorname{argmax}(B_i \cap B_c), \text{ s.t. } B_i \subset B_c \quad (2)$$

[0060] 其中, B_c 表示分割后的团块区域的图像; B_i 表示对齐的结果,其通过将 s_i 被填充的区域旋转角度 θ_i 并使用数字 r_i 进行缩放后得到, B_i 是与 B_c 具有相同尺寸的二值图像;其中 r_i 与 θ_i 的值由格点搜索来确定。

[0061] 需要对齐的结果 B_i 在 B_c 的内部,如果没有这种约束,则实际得到的形状假设是与整个团块区域对齐的,而不是与细胞质本身对齐的。

[0062] 在形状演进(步骤303)中,对于对齐的结果 B_i ,本发明的形状演进算法能够找到比 s_i 更合适的细胞质的形状假设。首先,需要限定如以下公式(3)所示的目标函数:

$$[0063] \quad \mathbb{E}(B_c, \mathbf{x}^k) = \sum_{(x,y) \in \Omega_B} (B_g(x,y) - B_c(x,y))^2 \quad (3)$$

[0064] 其中, \mathbf{x} 用来表示集合 $\{\mathbf{x}_i\}_{i=1}^N$,用 \mathbf{x}^k 来表示第k次演进的 \mathbf{x} ,如公式(1)所述,由 \mathbf{x}_i 来确定 s_i ;N表示团块中细胞质的数量; $B_g = \bigcup_{i=1}^N B_i$,其是由对齐结果 $\{B_i\}_{i=1}^N$ 生成的二值图像;(x,y)表示图像中像素的坐标。

[0065] 由此可见,公式(3)表示的目标函数实际上是检测已分割的团块区域与由对齐结果组成的团块区域之间关于像素的差。在理想状态下,如果非常精确地分割所有的细胞质, \mathbb{E} 等于0。如上所述,这种理想状态是很难实现的,由此,在本发明的方法中设计了目标函数,其主要原因是充分利用团块的边界信息,同时最小化重叠区域中不充分的强度信息的影响。

[0066] 因此,本发明通过以下公式(4)-(6)来找到能够使 \mathbb{E} 具有比 \mathbf{x}^k 更低的值的 \mathbf{x}^{k+1} 。其中,对于任何从公式(3)得到的矩阵p,在此使用泰勒定理得到以下公式(4):

$$[0067] \quad \mathbb{E}(\mathbf{x}^k + \mathbf{p}) = \mathbb{E}(\mathbf{x}^k) + \nabla \mathbb{E}(\mathbf{x}^k)^T \mathbf{p} + \frac{1}{2} \mathbf{p}^T \nabla^2 \mathbb{E}(\mathbf{x}^k + \gamma \mathbf{p}) \mathbf{p} \quad (4)$$

[0068] 其中, ∇ 和 ∇^2 分别表示梯度和海森矩阵 (Hessian) 计算; γ 是间隔 (0,1) 之间的某个标量。上述公式 (4) 表示可使用仅仅关于函数值、 \mathbf{x}^k 处的一阶导数和二阶导数的信息来近似 \mathbf{x}^k 附近的 \mathbb{E} , 从而获得以 \mathbf{x}^k 为圆心、以 $\|\mathbf{p}\|_2$ 为半径的区域内的 \mathbb{E} 的整体区域最小值。

[0069] 从理论上而言, 该整体区域最小值是能够在第 k 次演进时使用的最佳 \mathbf{x}^{k+1} 。但是, 由于标量 γ 的值是未知的, 不能直接分析得到 \mathbb{E} 。因此, 回到 \mathbf{x}^k 处来近似 \mathbb{E} , 由以下公式 (5) 中的 m_k 来表示 \mathbf{x}^k 处的 \mathbb{E} :

$$[0070] \quad m_k(\mathbf{p}) = \mathbb{E}(\mathbf{x}^k) + \nabla \mathbb{E}(\mathbf{x}^k)^T \mathbf{p} + \frac{1}{2} \mathbf{p}^T \nabla^2 \mathbb{E}(\mathbf{x}^k) \mathbf{p} \quad (5)$$

[0071] 当 $\|\mathbf{p}\|_2$ 很小时是非常准确的, 近似误差为 $O(\|\mathbf{p}\|_2^3)$, 然后由 m_k 的最小值来近似 \mathbb{E} 的最小值, 如公式 (6) 所示:

$$[0072] \quad \mathbf{p}^* = \underset{\mathbf{p} \in \Omega_{\mathbf{p}}}{\operatorname{argmin}} m_k(\mathbf{p}), \text{ s.t. } \|\mathbf{p}\|_2 \leq \Delta_k \quad (6)$$

[0073] 可通过现有的信赖域方法 (可参见 J.Nocedal 和 S.J.Wright. "Numerical Optimization". Springer, 2006) 来解公式 (6), 最终, 第 k 次演进的输出结果在公式 (7) 中示出:

$$[0074] \quad \mathbf{x}^{k+1} = \mathbf{x}^k + \mathbf{p}^* \quad (7)$$

[0075] 一旦得到 \mathbf{x}^{k+1} , 就将新的形状假设与图像对齐, 然后开始新一轮的演进计算, 直到 \mathbb{E} 不能再减小或达到了预定阈值 \mathbb{E}^* 。对于细胞质 i 而言, 对齐后的最终形状假设 \mathbf{s}_i 为最终的分割结果。

[0076] 学习形状实例重要度

[0077] 为了使根据上述公式 (1) 计算的形状假设集 \mathcal{H}_s 中的形状假设能够更好地还原任何不可见的细胞质形状, 本发明还在形状统计计算中采用了学习形状实例重要度的步骤 203。本发明的学习形状实例重要度的方法能够解决现有技术中需要手动收集形状实例所带来的一系列问题。

[0078] 具体而言, 先随机地选择一组 K 个输入-输出对, 见以下公式 (8):

$$[0079] \quad \mathcal{D} = \{(B_c^j, \{\mathbf{s}_i^j\}_{i=1}^{N_j})\}_{j=1}^K \quad (8)$$

[0080] 其中, B_c^j 是训练图像 j 中团块区域被分割的图像; \mathbf{s}_i^j 表示图像 j 中细胞质 i 的形状向量; N_j 是图像 j 中细胞质的数量。从上述公式 (8) 得到 \mathcal{D} , 选择一小组形状实例 $\{\mathbf{s}_i\}_{i=1}^{K_s}$ 。每个实例 \mathbf{s}_i 最初具有一个重要度, 由 w_i 表示。然后, 利用以下公式 (9) 计算平均形状:

$$[0081] \quad \mu = \frac{1}{W} \sum_{i=1}^{K_s} w_i \mathbf{s}_i \quad (9)$$

[0082] 其中, w 是所有 w_i 的和。然后, 利用以下公式 (10) 计算协方差矩阵:

$$[0083] \quad \mathbf{M}_c = \frac{1}{K_s} \sum_{i=1}^{K_s} (s_i - \mu)(s_i - \mu)^T \quad (10)$$

[0084] 其中, 矩阵 \mathbf{M}_c 的第一 t 个特征向量构成矩阵 $\mathbf{M} = (e_1 \ e_2 \ \cdots \ e_t)$, 其对应的特征值为 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_t$ 。

[0085] 学习形状实例重要度的步骤203与上述的受约束的多形状演进步骤202和建立形状假设集步骤201的关联 (见图2) 在于: 在每个训练图像 B_c^j 中, 根据建立形状假设集步骤中初始的 μ 和 \mathbf{M} , 运行所述受约束的多形状演进步骤, 随后检测最终的 \mathbb{E}_j 。如果 \mathbb{E}_j 大于预定阈值 \mathbb{E}^* , 则当前的 μ 和 \mathbf{M} 不够理想, 因此需要根据学习形状实例重要度的步骤中的公式 (9) 和 (10) 重新计算以增加 $\{s_i^j\}_{i=1}^{N_j}$, 从而更新形状假设集中的 μ 和 \mathbf{M} 。在更新过程中, 每个形状实例重要度 w_i^j 是由格点搜索来确定的。继续上述增加 $\{s_i^j\}_{i=1}^{N_j}$ 的步骤, 直到 \mathbb{E}_j 开始变小。整个学习形状实例重要度的步骤都根据公式 (8) 在 \mathcal{D} 中重复进行, 直到所有最终的差值 $\sum_{j=1}^K \mathbb{E}_j$ 的和不再减小。

[0086] 如图4所示, 本发明还涉及一种重叠细胞质分割系统, 其包括: 用于建立形状假设集的形状假设集模块M1、用于进行受约束的多形状演进的演进模块M2和用于学习形状实例重要度的学习模块M3。首先, 形状假设集模块M1提供初始的 μ 和 \mathbf{M} , 然后在演进模块M2中运行受约束的多形状演进, 随后检测最终的 \mathbb{E}_j 。如果 \mathbb{E}_j 大于预定阈值 \mathbb{E}^* , 说明当前的 μ 和 \mathbf{M} 不够理想, 则需要在学习模块M3中根据上述公式 (9) 和 (10) 重新计算以增加 $\{s_i^j\}_{i=1}^{N_j}$, 从而更新形状假设集中的 μ 和 \mathbf{M} 。继续上述增加 $\{s_i^j\}_{i=1}^{N_j}$ 的步骤, 直到 \mathbb{E}_j 开始变小。

[0087] 实施例

[0088] 本实例基于两种典型的宫颈刮片数据集, 分别是巴氏染色数据集和伊红染色 (H&E stain) 数据集, 其中巴氏染色数据集的获得, 参见 (Z.Lu, G.Cameiro, A.Bradley等, "Evaluation of three algorithms for the segmentation of overlapping cervical cells", IEEE Journal of Biomedical and Health Informatics, 21(2):441-450, 2017), 这个数据集包括8个公众可获得的图像, 每个图像具有平均有3.3个细胞质的11个团块; 伊红染色数据集由伊红染色制备, 该数据集包括21个图像, 每个图像具有平均有6.1个细胞质的7个团块。

[0089] 首先, 从巴氏染色数据集随机选取3个图像, 并从伊红染色数据集随机选取5个图像, 来构造训练集 \mathcal{D} , 剩下的图像构成测试集。训练集 \mathcal{D} 具有72个团块和324个细胞质实例, 其中的28个孤立的细胞质用于初始化小的形状实例集合 $\{s_i\}_{i=1}^{K_s}$ 。在测试集中具有包含907个细胞质的184个团块, 每个团块中的细胞质的数量是2至13 (平均值是4.93, 标准偏差为1.81)。

[0090] 在本发明描述的方法中, 需要设定两个参数: 用于终止多形状演进步骤的预定阈

值 \mathbb{E}^* ,以及用于计算矩阵 M 的特征向量 t 的数值。尽管通常小的 \mathbb{E}^* 有助于改进分割结果的精度,但越小的 \mathbb{E}^* 会使计算时间越长。本发明为了平衡分割结果的精度和计算时长,将预定阈值 \mathbb{E}^* 设定约为团块中像素数量的3%-7%之间,优选约为5%。对于特征向量 t 的数值,越大的 t 会使评估的形状 s_i 呈现出更多整体形状的细节,但是同样会使形状演进过程消耗更多的计算资源。在本发明中,根据现有技术中(参见T.F.Cootes,C.J.Taylor,D.H.Cooper等,“Active shape models-their training and application”,Computer Vision and Image Understanding,61(1):38-59,1995)已有的公式 $(\sum_{i=1}^t \lambda_i / \sum \lambda_i) > 0.995$ 来确定 t ,在本实例中 t 值设定为20。

[0091] 在本实施例中,将根据本发明的方法得到的结果与四个现有技术进行了比较(见以下的表1)。这四个现有技术分别是联合水平集函数法(参见Z.Lu,G.Carneiro和A.P.Bradley,“An improved joint optimization of multiple level set functions for the segmentation of overlapping cervical cells”,IEEE Transactions on Image Processing,24(4):1261-1272,2015.)、多细胞标记法(参见Y.Song,E.L.Tan,X.Jiang等,“Accurate cervical cell segmentation from overlapping clumps in pap smear images”,IEEE Transactions on Medical Imaging,36(1):288-300,2017)、多通道分水岭法(参见A.Tareef,Y.Song,H.Huang等,“Multi-pass fast watershed for accurate segmentation of overlapping cervical cells”,IEEE Transactions on Medical Imaging,2018.)以及轮廓分段法(参见Y.Song,J.Qin,B.Lei等,“Automated segmentation of overlapping cytoplasm in cervical smear images via contour fragments”,In Proceedings of the 32th AAAI Conference on Artificial Intelligence,168-175页.AAAI,2018.),在表1中分别使用LSF、MCL、MPW和CF来表示上述四种现有技术的分割结果。其中,LSF、MCL和CF属于现有的基于先验形状的方法,MPW属于基于分水岭方法的变型。

	巴氏染色数据集					伊红染色数据集				
	$(0, \frac{1}{4})$	$[\frac{1}{4}, \frac{2}{4})$	$[\frac{2}{4}, \frac{3}{4})$	$[\frac{3}{4}, 1)$	总体	$(0, \frac{1}{4})$	$[\frac{1}{4}, \frac{2}{4})$	$[\frac{2}{4}, \frac{3}{4})$	$[\frac{3}{4}, 1)$	总体
[0092] LSF	0.81	0.77	0.74	0.71	0.77	0.80	0.75	0.72	0.69	0.75
MCL	0.80	0.79	0.77	0.72	0.78	0.81	0.76	0.72	0.70	0.76
MPW	0.82	0.80	0.77	0.73	0.79	0.81	0.79	0.74	0.71	0.78
CF	0.84	0.82	0.79	0.77	0.81	0.83	0.81	0.79	0.75	0.80
本发明	0.85	0.83	0.82	0.81	0.83	0.84	0.83	0.82	0.80	0.82

[0093] 表1

[0094] 表1列出了多个重叠度条件下的、由不同方法得到的分割结果的定量比较。重叠度是为了衡量重叠的程度,其是由被遮挡的边界部分与细胞质的整个边界部分的长度比率来计算的。从表1的结果可以看出:本发明的方法得到了最好的分割结果,相比于其他方法,本发明的方法的精确度平均改进了大约5%。具体而言,当重叠度小于0.5(见表中的 $(0, \frac{1}{4})$ 、 $[\frac{1}{4}, \frac{2}{4})$ 栏的内容)时,本发明的方法的精确度平均改进了大约3%;当重叠度大于0.5(见表中的 $[\frac{2}{4}, \frac{3}{4})$ 、 $[\frac{3}{4}, 1)$ 栏的内容)时,精确度平均改进了大约8%。

[0095] 此外,图5示出了根据本发明的方法和现有技术的方法得到的分割结果的比较图像。其中,图5-(a)是原始输入的图像,图5-(b)至图5-(f)分别依次表示由LSF、MCL、MPW、CF

和本发明的方法获得的分割结果的图像,图5- (g) 是实际的分割图像。从图5可以更直观地看出,本发明的分割方法 (如图5- (f) 所示) 获得了最好的分割结果,对于其中部分细胞质实例,甚至与实际的分割结果 (如图5- (g) 所示) 完全一致。

[0096] 本发明的方法和系统克服了重叠区域中由于缺少强度信息而导致不能精确地分割细胞质的问题。相比于现有的基于先验形状的技术,本发明的方法和系统通过建立无限大的形状假设集、计算演进局部先验形状和整体先验形状、并对最终结果施加形状约束而提供了一种更加精确的重叠区域细胞质分割方法与系统。

[0097] 本发明的方法和系统不仅限于用于宫颈癌检测中,而且本领域技术人员能够进行适当的改进,使本发明的方法和系统应用于定量测量细胞水平特征的其他显微镜检查图像,例如病理学图像测量。

[0098] 尽管上面已经参考基于先验形状的检测重叠细胞质的分割方法及系统的具体实施例描述了本发明,但当然能想到,本领域普通技术人员可以推导出许多变型,因此,本领域普通技术人员容易想到的变型被认作本发明的一部分。本发明的范围在所附的权利要求书中限定。

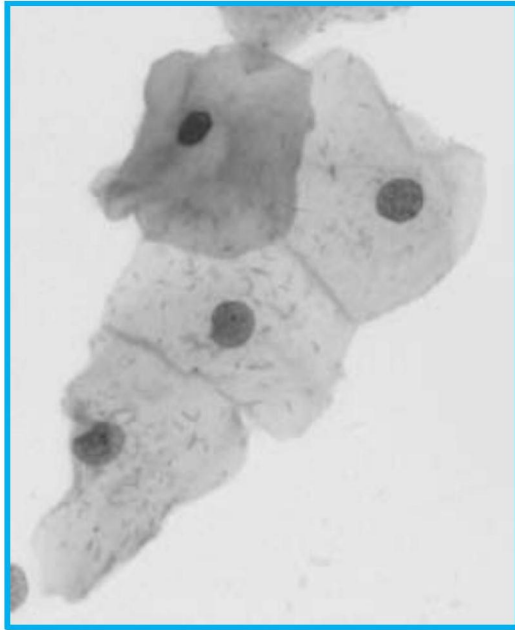


图1a

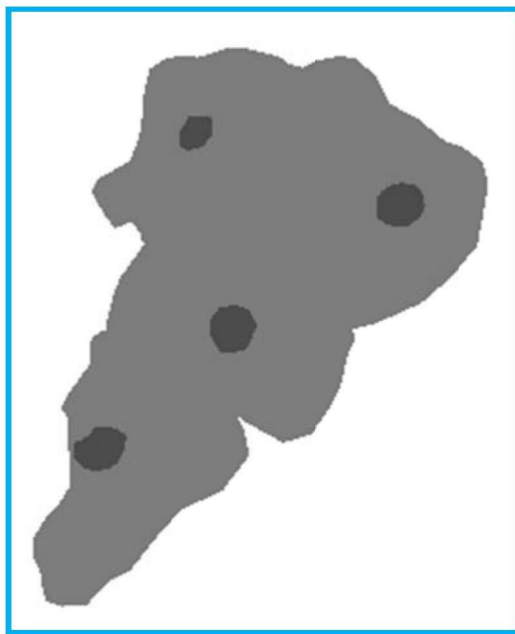


图1b

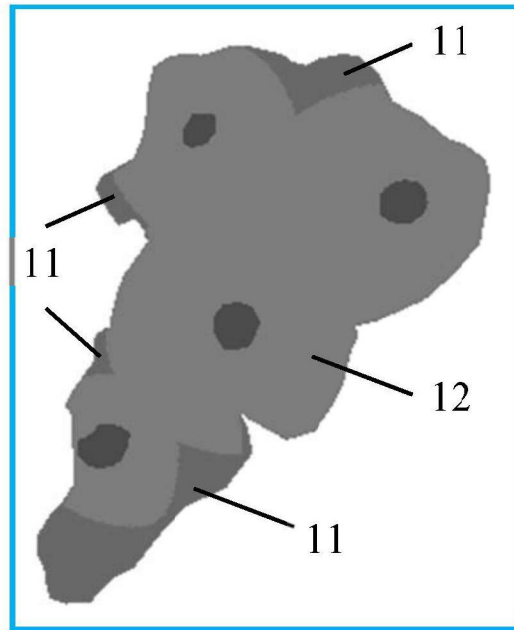


图1c

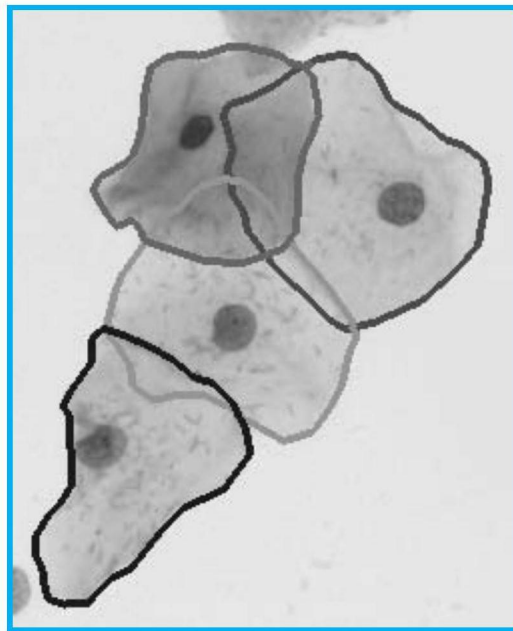


图1d

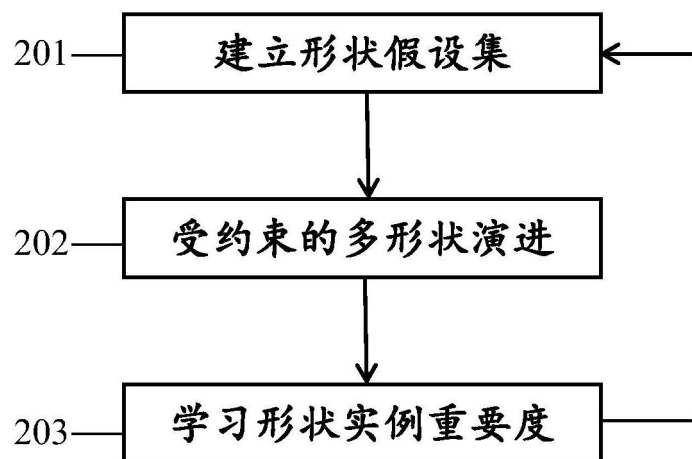


图2

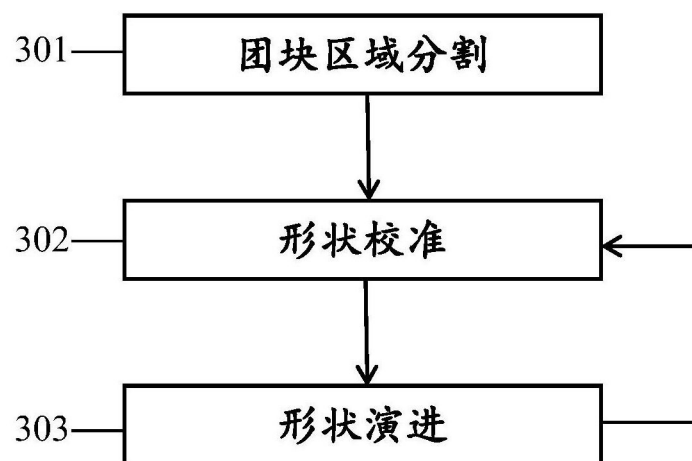


图3

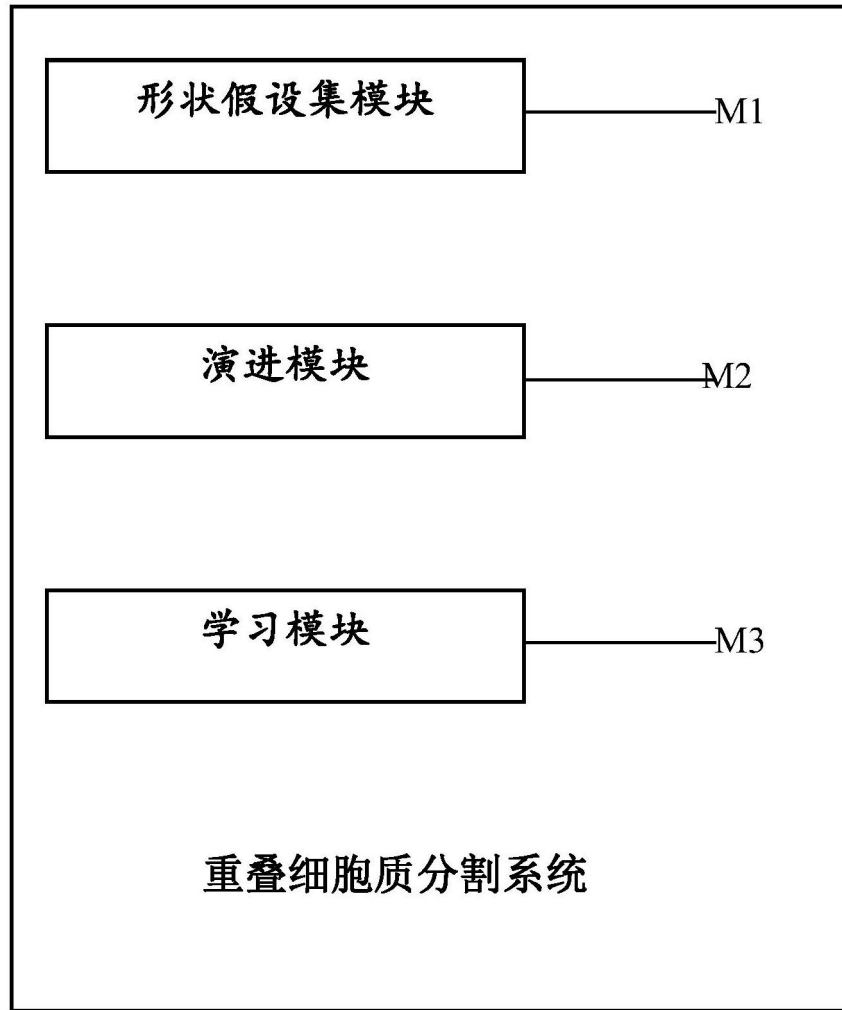


图4

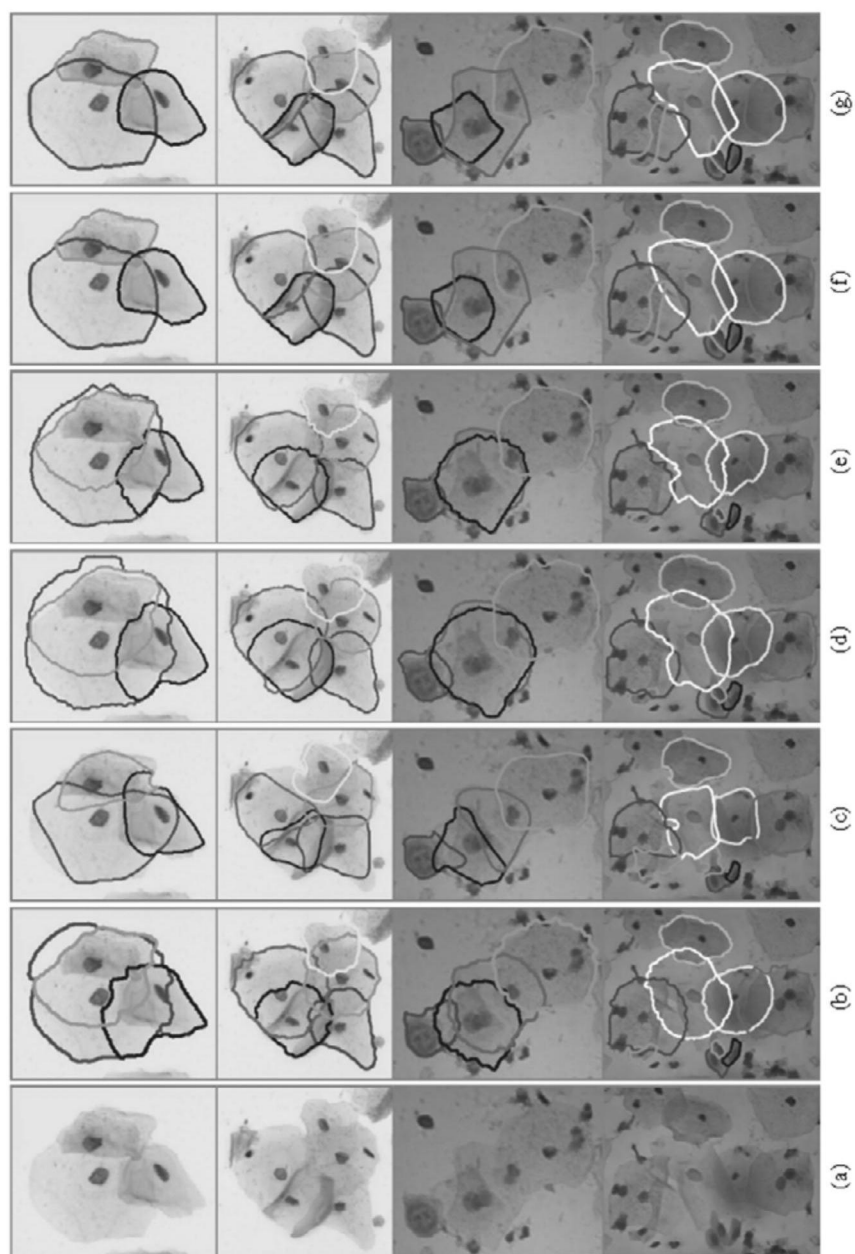


图5