

细胞学图像中细胞核染色质特征提取方法研究

徐传运^{1,2} 张杨³ 王森¹

(重庆理工大学计算机科学与工程学院 重庆 400050)¹ (重庆大学计算机科学与工程学院 重庆 400044)²
(重庆师范大学数学与计算机科学学院 重庆 400047)³

摘要 针对宫颈细胞学涂片图像分析中需要描述细胞核染色质颗粒大小和染色质分布均匀度的需求,提出了一种基于数学形态学的颗粒分析方法来描述细胞核的染色质颗粒大小,使用以细胞核质心为起点的各条辐射线上的颜色深度的方差来描述染色质分布的均匀度。文中具体描述了两种染色质特征提取的算法,并用实验证明了用提取的染色质特征描述细胞核图像的染色质特征的有效性和可行性。

关键词 数学形态学,颗粒分析,细胞学涂片,染色质特征

中图分类号 TP311.5 文献标识码 A

Cytology Images Chromatin Feature Extraction Method

XU Chuan-yun^{1,2} ZHANG Yang³ Wang Sen¹

(College of Computer Science and Engineering, Chongqing University of Technology, Chongqing 400050, China)¹

(Department of Computer Science and Engineering, Chongqing University, Chongqing 400044, China)²

(College of Computer Science and Information Science, Chongqing Normal University, Chongqing 400047, China)³

Abstract This paper focused on the requirements of describing particle size and distribution uniformity of nuclear chromatin in image analysis of cervical cytology, proposed a mathematical morphology-based particle analysis method to describe particle size of nuclear chromatin. In this method, the distribution uniformity of chromatin is described using the variance of color depth of each radiation line whose starting point is the centroid of nuclei. It described two specific chromatin feature extraction algorithm. The experiments prove that the proposed chromatin described features can be more effective nuclear chromatin image characteristics.

Keywords Mathematical morphology, Particle analysis, Cytology, Chromatin features

1 引言

宫颈细胞学涂片检查是普遍使用、经济有效、副作用小的早期宫颈癌筛查方法,而采用模式识别技术对宫颈细胞学涂片扫描图像进行计算机辅助判读,能有效降低检查成本,提高诊断的准确性。其中,胞核的染色质颗粒大小和分布均匀度是重要判读依据,因为细胞在发生病变时,常常伴随染色质异常。因此,细胞学图像处理的相关研究者提出了多种方法来描述染色质的特征,文献[1]先把细胞核区域根据染色深度分为高、中、低三个子区域,再把每个区域面积占细胞核面积的比例作为染色质的分布特征,还用各个子区域的质心与细胞核区域的质心的欧氏距离来描述染色质分布的均匀度,但是该方法没能准确描述染色质的颗粒大小与分布均匀度,特别是用质心距离来描述均匀度的方法过于笼统,文献[2]利用细胞图像在 0° 、 45° 、 90° 、 135° 四方向上的灰度共生矩阵的能量、能量均值、能量方差等作为纹理特征,还利用 Tamura 纹理特征[3]来描述细胞的纹理,但灰度共生矩阵只是对描述有规律

的纹理比较有效,而对染色质这种自然纹理的描述能力较差。

本文先采用数学形态学中的颗粒分析方法提取染色质的颗粒大小,再使用以细胞核质心为起点的各个辐射线上的颜色深度的方差来描述染色质分布的均匀度。

2 染色质的颗粒大小

染色质分布于细胞核内部,根据细胞特性的不同,可能呈均匀分布或者颗粒状分布,本文把两种情况都归于对颗粒大小的度量。如果呈现均匀分布,则认为是颗粒的大小和细胞核的大小一致,即染色质只有一个颗粒,并且这个颗粒的大小等于细胞大小。

数学形态学的颗粒分析方法[4,5]是度量颗粒大小的有效工具。颗粒分布可以形象地描述为:用一系列网眼大小递增的筛子过滤颗粒,并度量每次过滤后的剩余颗粒数量,每过滤一次,剩余颗粒就减少一部分,如果某次过滤后剩余颗粒减少较大,那么说明颗粒中有较多的颗粒大小与本次过滤的筛子网眼大小一致,因此通过分析每次过滤后剩余颗粒的减少量

本文受国家自然科学基金项目:主动视觉注意的语义认知计算模型的研究(60975015),国家自然科学基金项目:基于主动视觉注意的网络视频感知质量评估方法研究(61202348)资助。

徐传运(1979—),男,博士,讲师,主要研究领域为信息检索、图像语义、软件工程, E-mail: xuchuan-yun0@qq.com; 张杨(1981—),女,博士,主要研究领域为软件度量、可信计算。

就能得到物质中颗粒的大小。在颗粒分析中,通常采用颗粒分析曲线来描绘被测颗粒中相对于每个筛子网眼大小的颗粒分布。在图像分析中,用数学形态学分析的开运算来模拟筛子的过滤过程。

二值图像的颗粒过滤可以定义为:

$$\psi_G(t) = G \circ tE \quad (1)$$

式中, E 为形态学基本结构元素, G 表示二值图像, \circ 表示形态学的开运算, tE 定义为基本结构元素与自身进行 t 次膨胀运算,如式(2)所示:

$$tE = \overbrace{E \circledast E \circledast \dots \circledast E}^t \quad (2)$$

式中, \circledast 表示形态学中的膨胀操作。式(1)中的开运算定义为:

$$A \circ E = (A \circledast E) \ominus E \quad (3)$$

式中, \ominus 表示形态学中的腐蚀操作符。根据式(3)可以看出,颗粒过滤可以定义为以下递推形式:

$$\begin{aligned} \psi_G(t) &= \psi_G(t-1) \circ E \\ \psi_G(0) &= G \end{aligned} \quad (4)$$

式中, G 表示原始二值图像。

颗粒分析曲线通常通过计算两次连续开运算之间的前景像素损失数量来绘制,像素损失数量的计算公式为:

$$\alpha(t) = \Omega(\psi_G(t-1)) - \Omega(\psi_G(t)) \quad (5)$$

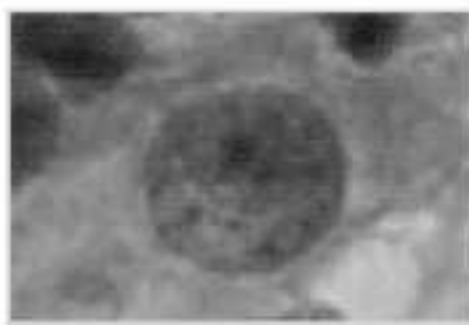
式中, $\Omega(\cdot)$ 表示求图像前景像素的数目。可以证明,当 $t \rightarrow \infty$ 时,必然有 $\lim_{t \rightarrow \infty} \Omega[\psi_G(t)] = 0$,因而当 $t \rightarrow \infty$ 时, $\lim_{t \rightarrow \infty} \alpha(t) = 0$ 。当 $\alpha(t)$ 为最大值时,图像中有最多的颗粒与此时结构元素大小一致,因而可认为图像的颗粒大小就是此时结构元素的大小。

对于灰度图像,式(1)~式(5)中所有的形态学操作全部要换成灰度图像下的形态学操作,式(1)中 $\Omega(\cdot)$ 函数定义为图像像素的积分:

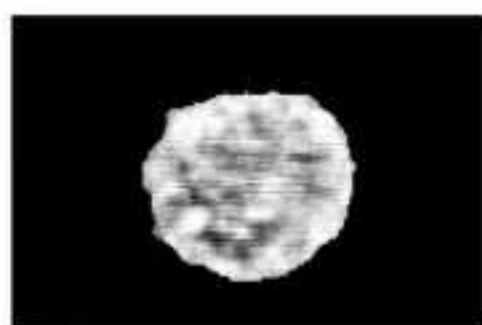
$$\Omega(g) = \int_{x=1}^N \int_{y=1}^M g(x,y) dx dy \quad (6)$$

式中, M, N 分别表示图像的长和宽, $\Omega(\cdot)$ 函数离散形式的定义为:

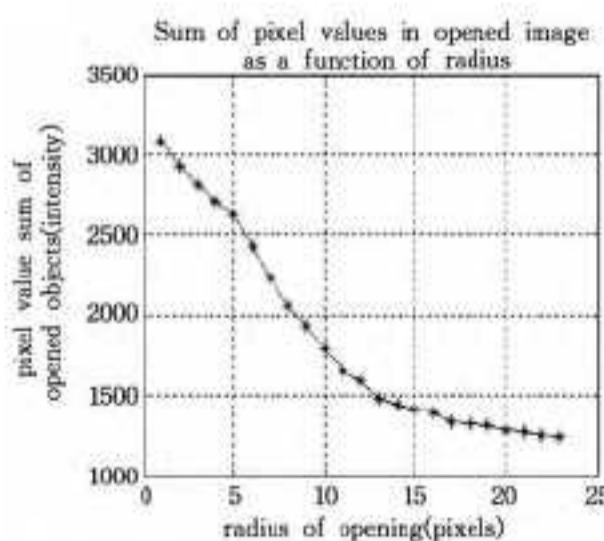
$$\Omega(g) = \sum_{x=1}^N \sum_{y=1}^M g(x,y) \quad (7)$$



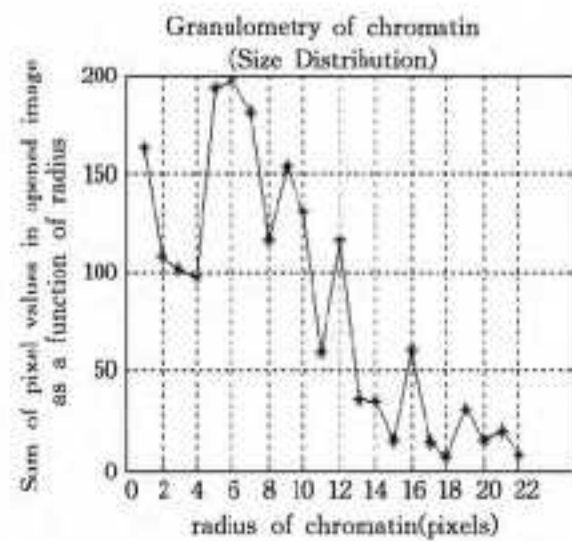
(a) 原始图像



(b) 对比度加强的灰度图像



(c) 开运算后的图像累加



(d) 开运算后的像素损失

图1 细胞核的颗粒分析示例

图1显示了细胞核颗粒分析的示例。由于染色质与其背

景的对比度较低,因此在进行颗粒分析之前,必须增强图像对比度。本文在实验时采用了对比度受限自适应直方图均衡(CLAHE)^[6]方法来增强对比度,效果比较好。

从图1(d)可以看出,在结果元素大小为6时,像素损失最大,因此认为图像的染色质颗粒大小为6。

3 染色质均匀度

染色质均匀分布表现为在细胞核的各个方向上染色质数量比较接近。如图(2)所示,在abcde 5个方向上染色质数量比较接近,意味着5个方向上的染色数量的方差接近于0。本文采用以细胞核的质心为起点的各条辐射线上的染色质数量的方差作为染色质分布均匀性的特征。

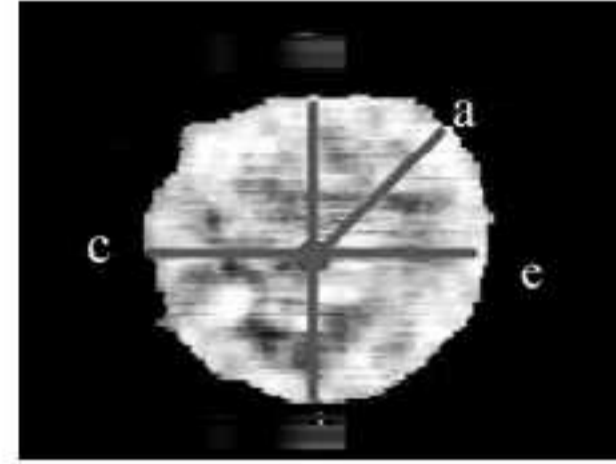


图2 辐射线的染色质

染色质在各条辐射线上的染色质数量的方差(简称径向方差)定义为:

$$\begin{aligned} dr &= \frac{1}{n-1} \sum_{i=1}^n (\Omega_i - \bar{\Omega})^2 \\ \bar{\Omega} &= \frac{\sum_{i=1}^n \Omega_i}{n} \end{aligned} \quad (8)$$

式中, Ω_i 表示第 i 根辐射线上的染色质数量, n 表示辐射线的数量,等于细胞核轮廓的周长。 dr 值越大,表示染色质越不均匀。由于各条辐射线的长度不相等,因此式(8)归一化后重定义为:

$$\begin{aligned} dr &= \frac{1}{n-1} \sum_{i=1}^n \left(\frac{\Omega_i}{l_i} - \bar{\Omega} \right)^2 \\ \bar{\Omega} &= \frac{\sum_{i=1}^n \frac{\Omega_i}{l_i}}{n} \end{aligned} \quad (9)$$

式中, l_i 表示辐射线的长度。在细胞核灰度图像中,染色质数量可以用辐射线上的像素灰度值求和得到,如式(10)所示。

$$\Omega_i = \sum_{j=1}^{l_i} g_{l_i}(j) \quad (10)$$

式中, $g_{l_i}(j)$ 表示辐射线上第 j 个像素的灰度值。式(9)的方差受到细胞核中染色的总体染色深度影响,不利于对不同图像进行对比,因此需要进行归一化处理,归一化后的方差计算公式为:

$$dr = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{\Omega_i}{l_i} - \bar{\Omega} \right)^2 \quad (11)$$

计算式(8)中的辐射线上的染色质数量 Ω_i 需要先计算第 i 条辐射线的位置,计算效率较低。为了提高计算效率,可以先把细胞核图像从笛卡尔坐标系转换到极坐标系,再计算极坐标系中图像每列上的像素,以得到 Ω_i 。图像从笛卡尔坐标系转换到极坐标系的计算过程如式(12)~式(15)所示。

$$\theta = \arctan\left(\frac{y - yc_c}{x - xc_c}\right) \quad (12)$$

(下转第238页)

[7] 王尧,余祖俊,朱力强,等. 基于脉冲耦合神经网络和 Markov 随机场的立体匹配研究[J]. 仪器仪表学报, 2013, 34(7): 1540-1545

[8] Hyvärinen A, Hoyer P O, Inki M. Topographic Independent Component Analysis[J]. Neural Computation, 2001, 13(7): 1527-1558

[9] 赵彦明. 基于生物视觉信息的 PCNN 参数自适应设置方法及模型改进[J]. 计算机科学, 2013, 40(6): 291-294

[10] Olshausen B A, Field D J. How close are we to understanding V1? [J]. Neural Computation, 2005, 17: 1665-1699

[11] Serre T, Wolf L, Bileschi S, et al. Robust Object Recognition with Cortex-Like Mechanisms[J]. IEEE Transactions on Pat-

tern Analysis and Machine Intelligence, 2007, 29(3): 411-426

[12] Tatani K, Nakamura Y. Dimensionality reduction and reproduction with hierarchical NLPKA neural networks—extracting common space of multiple humanoid motion patterns[C]// Proc. of IEEE International Conference on Robotics and Automation, 2003: 1927-1932

[13] Eckhorn R, Reitboeck H J, Arndt M, et al. Feature linking via synchronization among distributed assemblies; Simulations of results from cat visual cortex[J]. Neural Computation, 1990(2): 293-307

[14] 杨光,王恒,徐鹏,等. 基于 NSCT 和 MPCNN 的人脸特征提取[J]. 计算机工程, 2012, 38(22): 152-158

(上接第 219 页)

$$r = \sqrt{(y - y_{c_c})^2 + (x - x_{c_c})^2} \quad (13)$$

$$y_p = r \quad (14)$$

$$x_p = \frac{\theta}{2\pi} N_p \quad (15)$$

式中, x_{c_c}, y_{c_c} 表示细胞核的质心, (r, θ) 分别对应极坐标系的极径和极角, (x_p, y_p) 对应变换后的 (x, y) 坐标。图 3 展示了图 2 在极坐标变换后的辐射线。



图 3 极坐标系中的辐射线

4 实验与分析

本实验从 Herlev^[7] 宫颈细胞图像数据集选取了 5 张细胞图像来分析染色质颗粒大小与均匀度, 实验结果见表 1。

表 1 染色质颗粒大小与均匀度实验结果

编号	颗粒大小	均匀度
C01	20	4.82
C02	5	0.39
C03	7	5.20
C04	5	0.73
C05	5	1.33

从表 1 中可以看出, 细胞 C01 由于染色质完全融合在一起, 几乎没有颗粒的存在, 因此颗粒大小值为 20; 相比之下细胞 C02、C04、C05 颗粒大小值为 5, 表现为有明显的颗粒, 并且颗粒大小相差不大。细胞 C01、C03 的染色质偏向一边, 因而均匀度最差, 表现为度量值最大; 而细胞 C02、C04 的均匀度是最好的。

本次实验表明, 本文提出的颗粒大小和均匀度量方法与人类的认知基本一致。

结束语 针对宫颈细胞学涂片图像分析中对于描述细胞核染色质颗粒大小和染色质分布均匀度的需求, 本文借鉴数学形态学中的颗粒分析方法, 用以细胞核质心为起点的各条辐射线上的颜色深度的方差来描述染色质分布的均匀度。本文详细描述了两种染色质特征提取的算法, 并用相关实验及其结果充分证明了本文提取的染色质特征能够较为有效地描述细胞核图像的染色质特征。

参考文献

[1] 庞宝川. 基于计算机视觉的肿瘤早期诊断方法研究[D]. 武汉: 华中科技大学, 2009

[2] 范金坪. 宫颈细胞图像分割和识别方法研究[D]. 广州: 暨南大学, 2010

[3] Tamura H, Mori S, Yamawaki T. Texture features corresponding to visual[J]. IEEE Transactions on System, Man and Cybernetics, 1978, 8(6): 460-473

[4] Wang Xiao-peng, Hao Chong-yang, Wang Yun. Noise image segmentation based on morphological granulometry[C]// The Sixth International Conference on Electronic Measurement and Instrument. Taiyuan, China: ICEMI, 2003, 1: 301-304

[5] Soille P. Morphological Image Analysis: Principles and Applications (Second Edition)[M]. Springer-Verlag, 2008

[6] Zuiderveld, Karel. Contrast Limited Adaptive Histogram Equalization. Graphic Gems IV[M]. San Diego: Academic Press Professional, 1994: 474-485

[7] Jantzen J, Norup J, Dounias G, et al. Pap-smear benchmark data for pattern classification[C]// Proceedings of the NiSIS 2005, Nature inspired Smart Information Systems (NiSIS). Albufeira, Portugal, 2005: 1-9