

颜色特征加权滤波骨髓细胞分割方法

韩彦芳¹ 杨 娜¹ 缪 艳² 徐伯庆¹

(上海理工大学光电学院 上海 200093)¹ (上海儿童医学中心血液科 上海 200127)²

摘 要 颜色信息对骨髓细胞分割、分类非常重要。首先对已知类别的图像胞核、胞浆、成熟红细胞颜色采样进行 r, g, b 三值统计及线性关系分析,提取颜色特征;然后,采用 fcm 一次分割与多级阈值分割进行实验,针对边缘不完整和空洞问题,引入邻域颜色特征一致性系数,对特征图像进行加权极差滤波。实验表明,该方法可以简单、有效地实现骨髓细胞图像分割。

关键词 骨髓细胞分割,极差滤波,边缘增强,特征提取

中图法分类号 TP391 文献标识码 A

Bone Marrow Cell Segmentation Based on Color Feature Weighted Filter

HAN Yan-fang¹ YANG Na¹ MIAO Yan² XU Bo-qing¹

(University of Shanghai for Science and Technology, Shanghai 200093, China)¹

(Shanghai Children's Medical Center, Shanghai 200127, China)²

Abstract Color feature plays an important part in bone marrow cell segmentation as well as classification. Firstly, a color feature extraction and analysis method was discussed based on the relationship among r, g, b values of the sampling data from known classes. Then, fcm segmentation and multi-level thresholding were used for cell segmentation considering the color features. With respect to the defect of loss edges and holes inside, a weighted range filter was proposed by introducing one coefficients called in-neighbor consistency. Experiments were done and results show that this approach is simple and effective.

Keywords Bone marrow cell segmentation, Range filter, Edge enhancement, Feature extraction

1 引言

骨髓是人体最主要的造血系统。常见血液系统疾病,如急性髓系白血病(AML)、急性淋巴细胞白血病(ALL)及骨髓增生异常综合症(MDS)^[1]等,主要表现为骨髓细胞数量和形态的明显异常。而且,同一病种根据细胞形态和数量不同又分为不同亚类,如 ALL 中的 L1, L2, L3; AML 可分为 M0~M7。因此,骨髓细胞形态分析及计数对血液疾病诊断、鉴别具有重要价值^[2]。传统人眼细胞定位、识别及计数存在很强的主观性,诊断结果依赖于医生个人经验,而长时间肉眼观测将引起疲劳和分心,容易造成漏诊和误诊。基于图像处理技术的计算机辅助诊断,可为医生提供有效帮助。

临床上,通常依据血象和骨髓象诊断血液病。前者以外周血细胞为研究对象,后者则以骨髓细胞为研究对象。外周血细胞分割已比较成熟^[3-7],而由于骨髓细胞图像相对于外周血复杂得多,其分割仍是国内外研究的重点和难点。常用的骨髓细胞分割方法有阈值法^[8,9]、区域生长法^[10]、边缘检测法^[11-13]、基于模型的方法^[14,15]等。不同方法针对不同问题取得了一定成功,也存在一些缺陷。C. Garbay^[16]对松弛法、区域生长法和边缘检测进行了性能比较和分析,认为区域增长法在处理速度上不一定最优,但分割性能最好。C. E. Liedt-

ke^[17]则讨论了3种基于阈值的方法和基于分裂与合并的区域增长,指出基于分裂与合并的区域增长方法分割效果最佳。因此,结合流域算法的区域生长方法得到了广泛应用^[18,19]。但实验发现,阈值法会产生空洞以及一些小面积区域,因而需要基于数学形态学做大量后处理。边缘检测法对噪声干扰敏感,导致检测结果不稳定,得到的边缘常常是孤立点或分小段连续,导致边界不闭合,而得到的边缘宽度比实际大,以致引起相邻细胞边缘重叠(粘连)。模型法一般建立在边缘提取的基础上,也面临同样问题。区域增长法分割性能虽好,但种子点的确定受胞核形态多样性(既有单核又有多核)以及胞浆颗粒性的干扰,导致同一细胞被过度分割,同时逐像素的生长过程对噪声敏感,而算法速度同待处理图像面积成正比。

大多数方法基于灰度图像或 HIS 空间处理^[20]。基于 rgb 空间的处理一般分别利用 r, g, b 3 个通道的特征。实际上,一方面骨髓涂片制作过程中,细胞不同种类及生长阶段对染色基质亲和性不同,从而呈现不同颜色。如原、早幼、中幼和晚幼红细胞胞浆颜色随发育阶段不同而呈不透明深蓝、灰蓝到粉红等一系列变化;而嗜酸、嗜碱以及中性粒细胞胞浆根据发育阶段不同也呈不同颜色;胞核却基本呈紫红色。因此, rgb 颜色信息对区分细胞种类及发育阶段非常重要。另一方面,像素点的不同颜色由 r, g, b 按不同比例混合而成,因此,

到稿日期:2010-08-12 返修日期:2010-12-24 本文受国家自然科学基金面上项目(60972122)资助。

韩彦芳(1974-),女,博士,讲师,主要研究方向为图像处理、模式识别、医学图像处理。

r, g, b 之间的关系对颜色分析不容忽视。

本文从彩色骨髓细胞显微图像的 r, g, b 值出发,对胞浆、胞核、成熟红细胞 rgb 值进行采样,并分析这 3 值之间的统计与线性关系,提取颜色特征,进行分割。在此基础上,针对核内空洞即边界不完整问题,提出一种邻域加权极差滤波器,其综合考虑像素点颜色特征及邻域分布,进行胞核和胞浆分割。

2 骨髓细胞颜色特征提取

2.1 颜色统计特征

特征是数据分析的基础。一般,统计特征可分为 3 类:一类表示数据集中性,如平均数、中位数、众数等;一类表示数据分散性,如均方根差、极差、变异系数等;还有一类表示分布形态特点,如偏度、峰度等。将图像看作 rgb 空间的一个点集,则每个像素的颜色值由 r, g, b 大小决定。根据细胞染色知识,进行瑞氏染色后,胞核一般呈紫红色,胞浆随细胞种类及发育阶段不同呈不同颜色。因此,选取合适颜色特征,使不同类别像素在特征空间尽量分散,对图像分割事半功倍。

为此,本文首先对彩色显微细胞图片胞核、胞浆、成熟红细胞 r, g, b 值分别进行抽样,得到 3 类已知类别、各 50 组颜色数据 $I_i (i=1, 2, 3 \text{ 为类别数})$,则 I_i 为以 r, g, b 为维度的三维向量,大小为 $N (N=50)$ 。然后,计算 I_i 中每个样本点的 r, g, b 统计特征,考虑相互间线性关系,寻找使不同类别数据尽量分散的特征。需要说明的是,骨髓细胞图像包括胞核、胞浆、成熟红细胞和背景几部分,由于背景接近全白,较易分割,故只讨论前 3 类数据的区分。

基于上述思路,分别计算每个像素点 r, g, b 的均方差、变异系数、极差、 $r-g$ 值、 $g/(r+g+b)$ 值、 $2b-r$ 值,得到胞浆、胞核和成熟红细胞的特征值曲线图,如图 1 所示。其中绿色实线、蓝色折线及红色虚线分别为胞浆、胞核和红细胞特征曲线。

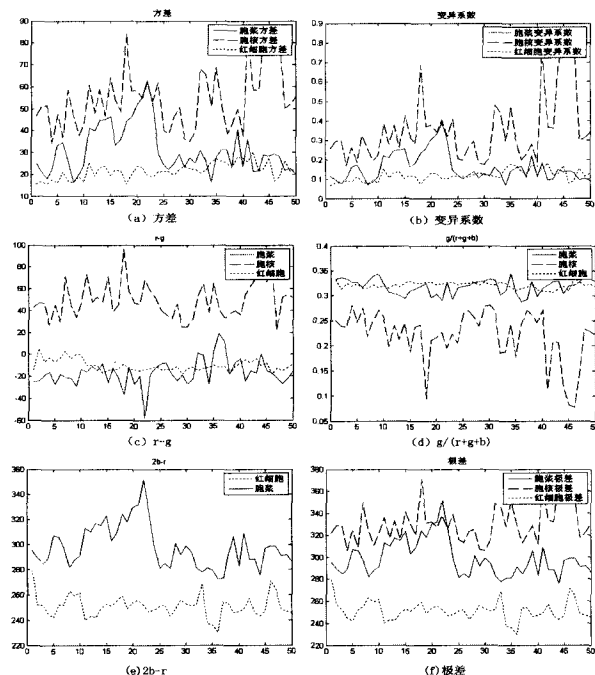


图1 胞核、胞浆、成熟红细胞特征值曲线

由图 1(a)、(b) 可知,方差和变异系数可以将胞核和成熟红细胞分开,但胞浆和两者都有交叠;由图 1(c)、(d) 可知, $r-g$

g 特征值和 $g/(r+g+b)$ 特征值可以很显著地将胞核分离出来,但胞浆和红细胞的 $r-g$ 值交叠在一起,而图(e)、(f)的 $2b-r$ 和极差则能将胞浆和成熟红细胞分离开。总之,没有一种特征能同时将 3 类数据分离开,所以骨髓细胞分割可采用多级分割或通过多特征组合聚类实现。另外,观察所有图中成熟红细胞的特征曲线可以发现,虽然成熟红细胞和胞浆特征曲线有交叠,但其特征曲线的波动幅度相对于胞浆特征曲线小很多,而这种波动幅度的大小可以通过方差或极差来体现。

2.2 特征相关性分析

假设训练样本集中有 M 个不同类别的样本。令 N_j 表示第 j 类的样本数,第 j 类中第 i 个样本的两个特征分别记为 x_{ij} 和 y_{ij} 。首先计算第 j 类每个特征均值:

$$\hat{\mu}_{xj} = \frac{1}{N_j} \sum_{i=1}^{N_j} x_{ij} \quad (1)$$

$$\hat{\mu}_{yj} = \frac{1}{N_j} \sum_{i=1}^{N_j} y_{ij} \quad (2)$$

式中, $\hat{\mu}_{xj}, \hat{\mu}_{yj}$ 分别为第 j 类特征 x 和特征 y 的均值, μ_{xj}, μ_{yj} 上的联系号($\wedge$)表示这两个值仅是基于训练样本的估值,不是真实的类均值。

特征方差估值为:

$$\hat{\sigma}_{xj}^2 = \frac{1}{N_j} \sum_{i=1}^{N_j} (x_{ij} - \hat{\mu}_{xj})^2 \quad (3)$$

$$\hat{\sigma}_{yj}^2 = \frac{1}{N_j} \sum_{i=1}^{N_j} (y_{ij} - \hat{\mu}_{yj})^2 \quad (4)$$

则第 j 类特征 x 和特征 y 的相关系数为:

$$\hat{\sigma}_{xyj} = \frac{\frac{1}{N_j} \sum_{i=1}^{N_j} (x_{ij} - \hat{\mu}_{xj})(y_{ij} - \hat{\mu}_{yj})}{\hat{\sigma}_{xj} \hat{\sigma}_{yj}} \quad (5)$$

其范围为 -1 到 1。如果该值为零,说明两个特征之间没有相关性。如果此相关系数绝对值接近 1,则说明两个特征相关性都很强。

根据式(1)一式(5),对于胞核有两个特征: $g/(r+g+b)$, $r-g$,其相关系数为 -0.7197;对于红细胞,有两个特征:极差、 $2b-r$,其相关系数为 0.9758。可见,每一类中的两个特征相关性很强。

2.3 基于颜色特征的骨髓细胞分割

根据以上分析,选择极差和 $r-g$ 作为颜色特征,分别采用 fcm 聚类方法和多级阈值方法对图 2 所示的骨髓细胞图像进行分割。

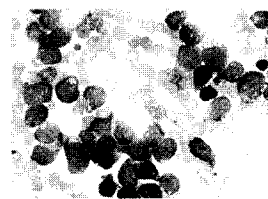


图2 骨髓细胞原始图像

2.3.1 fcm 聚类

图 2 为骨髓细胞原始图像。将图像 A 看作一个三维数据集,依据上述方法计算颜色特征 Jc 和 Rg ,采用 fcm 聚类,聚类中心个数为 3,得到的胞核和胞浆区域如图 3 所示。



(a) 胞核区域 (b) 胞浆区域

图3 fcm 分割结果

2.3.2 多级阈值分割

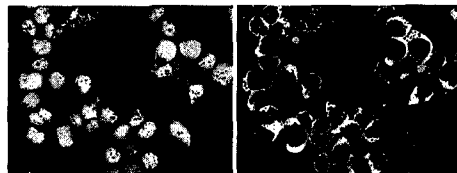
以图2为原始图像,多级阈值分割步骤如下:

1) 设原始图像为 A , 计算每个像素点 $A(i, j)$ 的 rgb 极差, 用 J_c 表示; 计算每个像素点 $r-g$ 值, 用 R_g 表示。

2) 以 J_c 为特征, 采用最大类间内距离比阈值法求阈值 th_{J_c} 。低于该阈值部分为成熟红细胞及背景, 高于该阈值部分为包括胞核、胞浆的感兴趣区域 roi 。

3) 以 R_g 为特征, 采用与上述相同的阈值法求阈值 th_{R_g} , 高于该阈值部分为胞核, 如图4(a)所示。

4. 从 roi 中减掉胞核部分, 得到胞浆区域, 如图4(b)所示。



(a) 胞核区域 (b) 胞浆区域

图4 多级阈值分割结果

从上述实验可见, 分级阈值法分割效果比 fcm 好很多, 主要原因在于: 所提取的颜色特征虽可使胞浆、胞核和红细胞中的一类与其他两类分离开, 但对每个特征, 其中总有两种是交叠的。因此, 同时采用多个特征一次分割将导致交叠区误分割, 而分级分割可以每次分离出与其他两种显著不同的一部分, 最终通过代数运算成功分离出各部分。此外, 该方法与 fcm 相比还具有简单、高效的优点。

但是, 从上图看到, 两种方法得到的胞核边缘都不完整, 而且核内存在较多空洞。虽然胞核染色粗糙是一个直接原因, 但算法本身没有考虑像素的邻域以及弱边缘也是一个重要因素。如果通过邻域处理平滑噪声、增强弱边缘, 则可以改善这种情况。从图1的特征曲线发现, 虽然成熟红细胞和胞浆特征曲线有交叠, 但其特征曲线的波动幅度相对于胞浆特征曲线小很多, 而这种波动幅度的大小可以通过方差或极差来体现。因此, 如果平滑噪声的同时对特征邻域求极差, 则特征幅度波动较大的胞核和胞浆可以从红细胞和背景中分离出来。

3 邻域加权极差滤波器

3.1 极差滤波器

极差滤波器是一种次序统计滤波器。它以中心像素邻域极差为输出, 实现有效边缘检测。设滤波器 G 窗口大小为 $m \times n$, 当前像素点 (i, j) 邻域内特征值分别为 $x_1, x_2, \dots, x_{mn}$ 。当滤波器窗口在图像上移动时, 极差滤波器输出为:

$$G(i, j) = \max(x_1, x_2, \dots, x_{mn}) - \min(x_1, x_2, \dots, x_{mn}) \quad (6)$$

因此, 极差滤波器输出反映了窗口内像素特征值的差异。若输出值大, 则表示特征值差异大, 该位置可能存在边缘; 反

之, 则表示特征值差异小, 对应图像平坦区域。此方法计算简单, 容易实现, 但对噪声敏感, 且随滤波窗口宽度增大, 边界将变宽。

3.2 邻域加权极差滤波

3.2.1 加权滤波器定义

针对极差滤波的缺陷, 如果能够根据邻域特征对极差滤波的输出进行加权, 则可以平滑邻域内部噪声, 同时增强边缘点处的输出。为此, 设计滤波器的冲激响应如下:

$$H = (\delta + H_w) * F = \delta * F + WG * F \quad (7)$$

式中, H 为滤波器输出, F 为被滤波图像, δ 为脉冲响应, G 为极差滤波器, W 为权矩阵。

加权滤波器输出 H_w 为:

$$H_w(i, j) = (WG)^{-1} \sum \sum F(i, j) (WG) \quad (8)$$

权矩阵 W 为:

$$W = \begin{bmatrix} w(i-1, j-1) & w(i-1, j) & w(i-1, j+1) \\ w(i, j-1) & w(i, j) & w(i, j+1) \\ w(i+1, j-1) & w(i+1, j) & w(i+1, j+1) \end{bmatrix} \quad (9)$$

3.2.2 权矩阵计算

为使 W 起到增强边缘的作用, 引入邻域特征一致性测度 s , 其计算过程如下:

设 f 为特征图像, 以 (i, j) 为中心的 $L \times L$ 邻域 $R = \{(m, n)\}$ 中, 经过点 (i, j) 做方向 $\theta (0 < \theta \leq \pi)$ 的直线将 R 分成两个区域 $R_1(\theta)$ 和 $R_2(\theta)$, 则邻域特征一致性测度为:

$$s_{ij} = \max_{\theta} [|f_{R_1}(\theta) - f_{R_2}(\theta)| / |R_1(\theta)|] \quad (10)$$

式中, $|R_1(\theta)|$ 表示 $R_1(\theta)$ 中像素个数。

$$f_{R_k}(\theta) = \sum_{(m, n) \in R_k(\theta)} f(m, n), k=1, 2 \quad (11)$$

则 s_{ij} 反映了邻域内直线两侧特征的总体差别。若邻域中存在过点 (i, j) 的边界, 当直线与边界重合时, s_{ij} 取最大值; 若邻域是一平坦区域, 则无论直线取何方向, s_{ij} 都将较小。由于 $f_{R_k}(\theta)$ 是 $R_k(\theta)$ 区域中每个像素特征之和, 求和具有平滑作用, 因此噪声对其影响不大。

$$W(i, j) = s_{ij} \quad (12)$$

4 实验结果及分析

基于以上分析, 实验步骤如下:

1) 分别计算原始图像极差特征图像 F_c 以及 $r-g$ 特征图像 F_g ;

2) 计算邻域一致性测度, θ 分别取 $0, 45, 90$ 和 135 度 4 个方向;

3) 采用 3×3 邻域加权滤波器滤波, 增强特征图像;

4) 根据特征图像进行阈值分割。

为验证本文方法的有效性, 对不同种类疾病的图像与控制标记符分水岭分割进行比较。本方法能克服传统分水岭方法过分割的缺点, 是一种有效的细胞分割方法。实验结果如图4所示。图中分别为 ALL-L2 骨髓像和 ALL-L1 骨髓像的滤波增强、本文方法得到的 roi 以及控制标记符分水岭分割得到的 roi 。

从图4可见, 控制标记符分水岭分割克服了传统分水岭方法的过分割缺点, 但由于灰度图像中胞浆和红细胞灰度值相近, 控制标记符分水岭方法不能得到完整的胞浆, 而且误将背景红细胞分割为目标。相对而言, 本文方法得到了较好的分割效果。

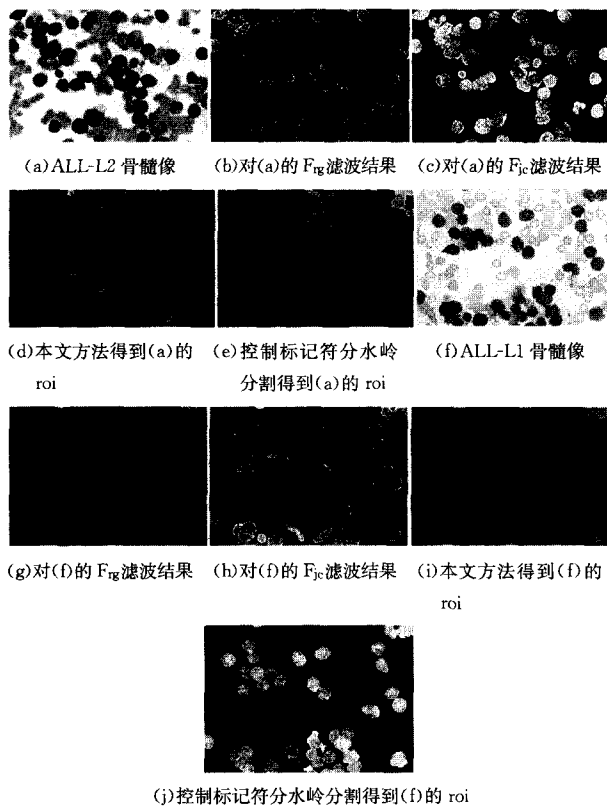


图4 本文分割及控制标记分水岭分割得到的 roi

在感兴趣区域基础上,采用多级阈值法分割细胞核和胞浆,结果如图5所示。

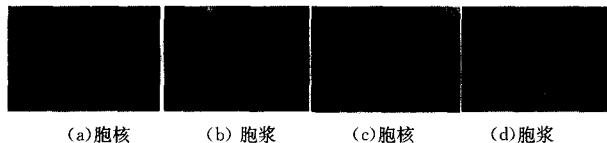


图5 细胞核和胞浆分割结果

基于邻域加权极差滤波的方法引入了邻域一致性系数,因而能较好平滑核内空洞,同时增强图像边缘。从图5可见,对增强图像进行多级阈值分割后,克服了由于核染色粗糙引起的空洞问题,达到了较好的效果。

结束语 颜色信息对骨髓细胞图像分割及分类非常重要。本文尝试从彩色细胞图像的 r, g, b 3 值的统计及线性关系出发,分析并提取颜色特征。分别采用一次 fcm 分割和多级阈值分割。由于细胞核染色粗糙性及胞浆颗粒性会导致分割结果存在空洞以及边界不完整,为此,提出一种邻域加权极差滤波器。实验表明,本文方法简单有效。

实验中发现,根据颜色特征曲线,任一特征同时只能分离一类数据与其他两类数据,不能同时分离 3 类数据。针对这一问题,本文采用了两级阈值分割方法,得到了比一次分割更好的效果。进而,对该问题的解决可以从寻找同时分离细胞核、胞浆和红细胞的特征入手,但由于骨髓细胞图像的分布特点,这种方法在颜色空间以及位置空间较难实现。因此可以将特征映射到其他空间,如核空间,以达到分离数据的目的。

另外,粘连细胞分离及细胞分类问题,将是我们下一步的研究内容。

参考文献

[1] 吴晓芝. 血液病诊断与鉴别诊断图谱[M]. 北京: 人民卫生出版

社, 2009(10)

- [2] 王敏, 石军, 陶英, 等. 幼稚前体细胞异常定位结构特点与急性髓系白血病早期复发的相关性分析[J]. 临床医学杂志, 2009, 22(7): 345-347
- [3] Giraldi G, Strauss E, Oliveira A. Dual_T_Snakes model for medical imaging segmentation[J]. Pattern recognition letters, 2003, 24: 993-1003
- [4] Nipon T-U, Dougherty E R, Gader P D. non-homothetic granulometric mixing theory with application to blood cell counting[J]. Pattern recognition, 2001, 34: 2547-2560
- [5] Nipon T-U. Patch-based white blood cell nucleus segmentation using fuzzy clustering[J]. ECTI Trans. on Electrical ENG. , Electronics, and Communications, 2005, 3(1): 15-19
- [6] Wang Shi-tong, Chung K F L, Fu Duan. Applying the improved fuzzy cellular neural network IFCNN to white blood cell detection[J]. Neurocomputing, 2007, 70: 1348-1359
- [7] Altman M B, Wang S J, et al. Cell detection in phase-contrast images used for alpha-particle track-etch dosimetry: a semi automated approach[J]. Phys. Med. Biol, 2005, 50: 305-318
- [8] Ballaro B, et al. An automated image analysis methodology for classifying megakaryocytes in chronic myeloproliferative disorders[J]. Medical image analysis, 2008, 12: 703-712
- [9] Poon S S S, Rabab K W, Pacic B. Automated image detection and segmentation in blood smears[J]. Cytometry, 1992, 13: 766-774
- [10] Zhang Xi-wen, Song Ji-qiang, Michael R L, et al. Extraction of karyocytes and their components from microscopic bone marrow images based on regional color features[J]. Pattern Recognition, 2004, 37: 351-361
- [11] Yang F, Jiang T. Karyocyte image segmentation with kernel-based dynamic clustering and an ellipsoidal karyocyte shape model[J]. Informatics, 2001, 34(2): 67-73
- [12] Garrido A, dela Blanca N P. Applying deformable templates for cell image segmentation[J]. Pattern Recognition, 2000, 33(5): 21-832
- [13] 冯强, 于盛林, 黄晓晴, 等. 一种新颖的有核边缘检测方法[J]. 中国图象图形学报, 2009, 14(10): 2004-2009
- [14] Park J, Keller J M. Snakes on the watershed[J]. IEEE Trans PAMI, 2001, 23(10): 1201-1205
- [15] Cong G, Parvin B. Model-based segmentation of nuclei[J]. Pattern recognition, 2000, 33(8): 1383-1393
- [16] Garbay C. Image structure representation and processing: a discussion of some segmentation methods in cytology[J]. IEEE Trans on PAMI, 1986, 8(2): 140-160
- [17] Liedtke C E, Gahm T, Kappei F. Segmentation of microscopic cell scenes[J]. Analytical and Quantitative Cytology and Histology, 1987, 9(3): 197-211
- [18] Malpica N, Solorezano C O, Vaquero J J, et al. Applying watershed algorithms to the segmentation of clustered nuclei[J]. Cytometry, 1997, 28: 289-297
- [19] Comaniciu D, Meer P. Mean shift: a robust approach toward feature space analysis[J]. IEEE Trans on PAMI, 2002, 24(5): 603-619
- [20] Montseny E, Sobrevilla P, Romani S. A fuzzy approach to white blood cells segmentation in color bone marrow images[J]. IEEE FUZZ, July 2004: 25-19