

文章编号:1671-6833(2015)04-0015-04

# 骨髓细胞图像的小波变换与 K-means 聚类分割算法

苏士美<sup>1</sup>, 吕雪扬<sup>2</sup>

(郑州大学 电气工程学院, 河南 郑州 450001)

**摘要:** 为了能准确地分割出骨髓细胞涂片中的各类细胞, 提出一种基于小波分析的聚类分割方法. 首先采用小波变换消除散焦噪声, 然后通过对彩色图像  $G$  分量进行小波系数多尺度分解, 提取特征参数信息, 根据图像  $G$  分量与  $S$  分量的差异性并结合得到变换图像  $STG$ , 二值化处理提取白细胞胞核, 最后为 K-means 聚类方法提供优化的初始聚类中心, 从而对各类红细胞、白细胞进行分割和分离. 通过对比分析和实验测试, 该算法有效克服了骨髓细胞显微图像的复杂散焦、细胞种类繁多以及目标区分度低而导致图像分割的困难, 准确率达 94.15%.

**关键词:** 细胞分割; 小波变换; K-means 聚类; 颜色特征

**中图分类号:** TP317.4 **文献标志码:** A **doi:**10.3969/j.issn.1671-6833.2015.04.004

## 0 引言

骨髓细胞涂片的形态分析对血液病诊断和鉴别具有重要价值. 诊断结果依赖于医生肉眼识别和经验, 存在主观性, 易造成漏诊和误诊. 基于图像处理技术的计算机辅助诊断, 可为医生的早期诊断提供帮助<sup>[1]</sup>. 由于骨髓细胞显微图像散焦较严重, 成分复杂, 一些成血细胞的边缘较为模糊, 且与背景的对比度低, 相对于其他细胞图像较为复杂, 其分割一直是国内外研究的重点. 近年来, 常用的细胞分割方法有阈值法、区域生长法、边缘检测法、基于模型的方法等. Madhloom H 等<sup>[2]</sup>把自适应局部阈值分割算法应用于复杂背景下的血细胞分割图像, 对于细胞中颜色不均匀处分割时易误生成空洞; Kong H 等<sup>[3]</sup>针对病理细胞涂片图像采用基于颜色和纹理的综合框架分析的有监督的分割方法, 分割过程需要人机交互, 且速度较慢; 模型法针对一幅特定的图像难以选择适合的特征向量, 而这直接影响到图像分割的效果. 针对染色后的骨髓细胞涂片呈现不同颜色的红细胞、白细胞, 其胞核与胞浆边界模糊、形状不规则、散焦噪声影响等分割难题, 经过多次试验, 笔者确定首先运用小波变换去除显微图像的背景噪声与散

焦影响, 为下一步的分割提供准确的精度, 再对彩色图像  $G$  分量直方图进行小波变换多尺度分解, 提取特征参数, 为改进的 K-means 聚类算法提供优化的初始聚类信息.

## 1 图像预处理

### 1.1 小波变换

小波变换是一种信号的时间-尺度分析方法, 在时频两域均具有表征信号局部特征的能力. 小波是函数空间  $L^2(R)$  中, 满足所谓的“容许条件”(即:  $C_\psi = \int_{-\infty}^{+\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty$ ) 的函数.

对任意信号  $f(t) \in L^2(R)$ , 连续小波变换:

$$WT_f(a, b) = \frac{1}{\sqrt{a}} \int_R f(t) \Psi^*\left(\frac{t-b}{a}\right) dt. \quad (1)$$

其离散小波变换中时间仍为连续变量, 将尺度参数和平移参数离散化, 得到分析小波:

$$\Psi_{(j,n)}(t) = 2^{-\frac{j}{2}} \Psi(2^{-j}t - n). \quad (2)$$

离散小波变换定义为:

$$WT_f(j, n) = \int_R f(t) \Psi_{(j,n)}^*(t) dt. \quad (3)$$

### 1.2 图像去噪

小波分析最主要的特点是自动变焦、多尺度

收稿日期: 2014-12-30; 修订日期: 2015-02-19

基金项目: 河南省自然科学研究计划资助项目(2011A510022)

作者简介: 苏士美(1965-), 女, 河南永城人, 郑州大学副教授, 硕士, 主要从事信号检测与处理, E-mail: smsu@zhu.edu.cn.

分析,被称为数学分析的“显微镜”. 因此利用小波变换实现频率选择和多尺度分解,可起到抑制背景噪声,消除散焦影响和增强目标的作用.

信号的分解公式为

$$A_{i+1,k} = \sum_n h(n-2k)A_{i,n}; \tag{4}$$

$$D_{i+1,k} = \sum_n g(n-2k)D_{i,n}. \tag{5}$$

式(4)为低频信号的分解过程,第*i*级低频信号  $A_{i,k}$  分解得第  $i+1$  级离散低频信号  $A_{i+1,k}$ ,*h* 为低通滤波器;式(5)为高频信号的分解过程,第 *i* 级高频信号  $D_{i,k}$  分解得第  $i+1$  级离散高频信号  $D_{i+1,k}$ ,*g* 为高通滤波器<sup>[4]</sup>. 利用小波变换去除低频噪声保留高频信号,如图 1 所示.

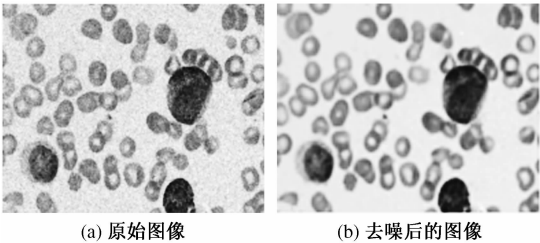


图 1 原始和去噪后的图像  
Fig. 1 Original and denoising images

1.3 小波多尺度分解

骨髓细胞图像存在背景、红细胞浆、白细胞核 3 大区域,其灰度值也由此分为低、中、高 3 大等级. 将彩色细胞图像的 RGB 各分量提取出得图 2 的各个图像,图 3 所示为各分量的直方图呈现一些“三峰”形状. 经过对比分析细胞图像的 *G* 分量更易于实验应用分割.

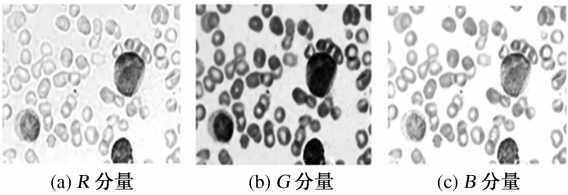


图 2 RGB 各分量图  
Fig. 2 Images of RGB components

如果把图像 *G* 分量的直方图看作一堆信号,则用小波变换的过零点检测对它进行多尺度分析和特征提取<sup>[5]</sup>. 用小波变换的过零点和极值点来检测信号的局部突变,选用在原信号上表现良好,且不会随尺度增加产生新零交叉点的高斯函数作检测用小波,得到直方图的每一个波峰点与波谷点,为细胞分割的 K-means 聚类算法提供初始聚

类中心参数.

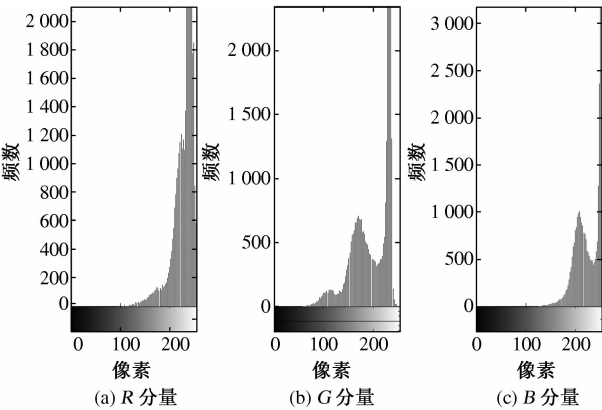


图 3 RGB 各分量直方图  
Fig. 3 Histograms of RGB components

2 细胞图像分割算法

2.1 K-means 聚类算法

算法的主要步骤包括<sup>[6]</sup>: (a) 随机初始化 *k* 个聚类中心; (b) 计算每个数据对象到聚类中心的距离,将每个点聚类到离该点最近的聚类中去; (c) 重新计算 *k* 个聚类中每个点的平均值,作为新的聚类中心; (d) 反复执行(b)、(c)两步骤,直到算法收敛为止.

2.2 改进的 K-means 聚类算法

传统算法是基于划分的方法,能简单、可靠、有效地分割图像. 然而聚类结果对算法选取的初值要求较高,选取不同的初值往往会导致不同的聚类结果,且目标函数极易陷入局部极小值,致使图像聚类分割不具备针对性,在相当程度上影响分割的效率和准确性.

通过统计彩色图像 RGB 各分量值出现的频数,体现在直方图中“波峰”与“波谷”中,即该像素点出现的概率的大小. 图 3(b)为图像 *G* 分量直方图,由于背景噪声以及光线和粘连的影响,存在多个伪“波峰”和“波谷”(即图中出现的各种“毛刺”现象),通过小波多尺度分解,筛选出像素出现频率较高的“波峰”处像素,提取较大的空间“密集程度”点作为改进算法的初始聚类中心,分别聚类为红细胞浆、白细胞核成分.

2.3 白细胞核提取

随着细胞有形成分从骨髓细胞涂片图像中分割出来,分割面临的下一个问题即是 将白细胞从图像中分离. 瑞式染色的骨髓显微图像大多以 RGB 彩色格式保存,然而胞核的染色浓度相当高于胞浆,利用饱和度参数 *S* 恰恰可以最佳的描述其变化,且 HIS 模型由于其亮度和色度的独立性,

使之较适应人眼的视觉特征,根据公式(6)将 RGB 空间转换到 HIS 空间<sup>[8]</sup>:

$$\begin{cases} H = \arctan\left[\frac{\sqrt{3}(G-B)}{(R-G)+(R-B)}\right]; \\ I = \frac{(R+G+B)}{3}; \\ S = 1 - \frac{\min(R,G,B)}{I}. \end{cases} \quad (6)$$

由图 3 骨髓图像各分量直方图对比显示,  $G$  分量更适宜于对比饱和度分量进行白细胞核的提取,经过多幅涂片图像观察,可以看出  $G$  分量与  $S$  分量正好相反,胞核在  $S$  分量中具有最高的亮度,但在  $G$  分量中却表现出最低的亮度;通过骨髓细胞图像在  $G$  分量与  $S$  分量中的变化规律,为突出胞核部分,构造出变换图像 STG,如式(7);运用公式(8)基于阈值的算法对 STG 图像进行二值化处理得  $BH$ ,将白细胞核提取出来.

$$STG = S/G. \quad (7)$$

$$BH = \begin{cases} 1, STG(i,j) > T; \\ 0, STG(i,j) \leq T. \end{cases} \quad (8)$$

其中  $T$  为 STG 直方图中的灰度阈值,图像矩阵的像素点坐标为  $(i, j)$ ,白细胞核区为大于  $T$  部分,背景与胞浆为小于  $T$  的部分. 部分骨髓细胞图像中的白细胞核提取如图 4 所示.

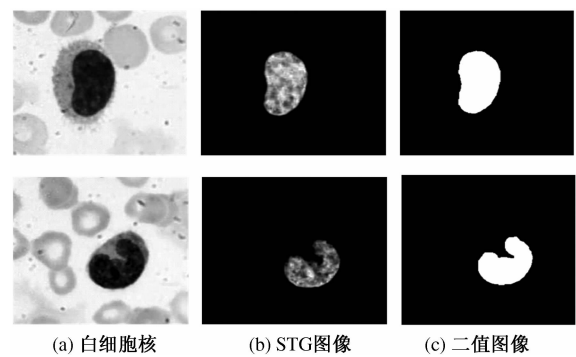


图 4 白细胞核提取

Fig. 4 White nucleus extract

3 实验对比分析

为了验证小波变换与改进 K-means 聚类算法对骨髓细胞显微图像的分割效果,笔者分别使用 Canny 边缘检测、Otsu 自适应阈值、距离变换分水岭、半监督聚类算法与本研究算法进行对比试验. 实验的平台是 Windows7、2.60 GHz CPU、2.91 GB 内存、500 GB 硬盘. 在 matlab7.0 环境下,对骨髓显微图像进行分割,其判据是算法能否完整保留细胞轮廓、提取图像各有形成分特征区域,以利于

后续的进一步处理. 实验结果如图 5 所示.

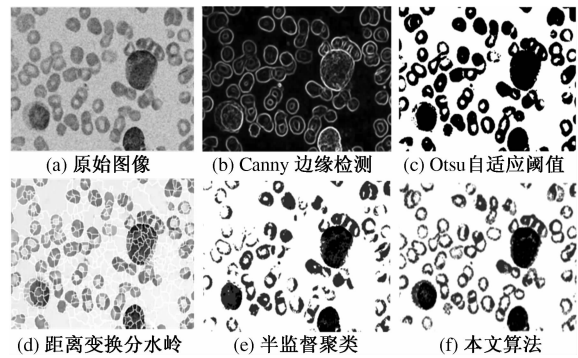


图 5 各算法分割效果图

Fig. 5 Renderings of various segmentation algorithm

通过以上几种算法的分割效果对比,得出以下结果分析:

(1) Canny 边缘检测算法<sup>[9]</sup>,对图像进行高斯滤波平滑,计算各像素正交方向梯度与幅值,“非最大化抑制”寻找细胞边缘,但不能有效处理粘连情况,如图 5(b)将粘连当成一个连通域,无法识别胞核与胞浆.

(2) Otsu 自适应阈值法<sup>[10]</sup>,计算像素邻域平均灰度,使像素的每个阈值依赖于以自身为中心的邻域窗口确定. 如图 5(c)呈现“非黑即白”的结果,无法获取信息,且对阈值依赖过大.

(3) 距离变换分水岭算法<sup>[11]</sup>,首先二值化图像,然后运用极限腐蚀处理得不同距离图,提取真实分水岭得目标图像. 但算法对微弱边缘较敏感,易生成封闭的单像素区域轮廓,造成过分割现象,难以判断真正边缘,如图 5(d)所示.

(4) 半监督聚类算法<sup>[12]</sup>,随机标记样本点隶属度赋初始值,把图像中每一个像素点  $p$  分配到与之相距最近的质心所代表的类中,用标记的样本点指导聚类中心的更新迭代计算. 无法准确分割出完整细胞轮廓,胞核、胞浆信息丢失严重,且背景散焦导致误分割如图 5(e)所示.

(5) 本算法在基于彩色 K-means 算法的基础上,首先使用小波分解去除噪声与散焦使细胞成分更加突出,有利于聚类正确的细胞成分,再应用小波多尺度分析对图像  $G$  分量直方图信息进行特征提取,为 K-means 聚类分割算法提供合适的初始聚类中心,得分割效果如图 5(f)所示.

为进一步验证本算法的稳定性,选用多幅骨髓细胞涂片图像进行试验测试,并将其中 3 个样本实验结果与上述各算法进行比较,得出各算法实现分割效果列表对比,结果表明本方法的效果最好,充分验证了本算法的有效性,结果如表 1 所示.

表 1 样本中各算法的分割结果对比  
Tab.1 The results contrasts of various algorithms

实验算法	算法实际分割数(胞浆/胞核)			准确率/ %
	样本一 (26/1)	样本二 (58/3)	样本三 (114/5)	
Canny 检测	18/1	46/2	92/3	79.19
Otsu 阈值	22/0	53/1	109/2	69.97
分水岭	30/0	63/0	128/2	86.70
聚类	20/1	49/2	102/4	89.97
本算法	23/1	52/3	109/4	94.15

4 结论

笔者针对骨髓显微图像的特点,利用小波变换多尺度分析作预处理,并为聚类提供初始化参数,以彩色分量 *G* 分量为聚类对象,利用 K-means 聚类算法分割细胞,取得了不错的分割效果.克服了传统单一分割算法对胞核与胞浆识别率低、误分割、过分割的问题,准确率达 94.15%,通过实验验证,该算法对骨髓细胞显微图像的分割是有效的.

参考文献:

[1] 韩彦芳,杨娜,缪艳,等.颜色特征加权滤波骨髓细胞分割方法[J]. 计算机科学,2011,38(7):283 - 286.  
[2] MADHLOOM H,KAREEM S,ARIFFIN H,et al. An automated white blood cell nucleus localization and segmentation using image arithmetic and automatic

threshold[J]. Appl SCI,2010,10(11):959 - 966.  
[3] KONG H, GURCAN M, BELKACEM K. Partitioning histopathological Images: an integrated framework for supervised color-texture segmentation and cell splitting[J]. IEEE Trans Med Imaging, 2011,30(9):1661 - 1677.  
[4] 陈莉. 基于小波变换的图像增强算法[J]. 陕西理工学院学报,2014,30(1):32 - 37.  
[5] 张聚,王陈,程芸. 小波与双边滤波的医学超声图像去噪[J]. 中国图象图形学报,2014,19(1):126 - 132.  
[6] ERISOGLU M,CALIS N,SAKALLIOGLU S. A new algorithm for initial cluster centers in K-means algorithm[J]. Pattern Recognition Letters, 2011, 32(14):1701 - 1705.  
[7] 顾广华,崔冬. 白细胞图像的自动分割算法[J]. 仪器仪表学报,2009,30(9):1874 - 1879.  
[8] 王小俊,刘旭敏,关永. 基于改进 Canny 算子的图像边缘检测算法[J]. 计算机工程,2012,38(14):196 - 198,202.  
[9] 倪麟,龚劬,曹莉. 基于自适应加权中值滤波的二维 Otsu 图像分割算法[J]. 计算机应用研究,2013,30(2):598 - 600.  
[10] 戴丹. 基于改进分水岭算法的粘连颗粒图像分割[J]. 计算机技术与发展,2013,23(3):19 - 22.  
[11] 孙刚灿,李苹苹,申金媛,等. 基于半监督聚类理论 MQAM 信号的调制识别[J]. 郑州大学学报:工学版,2014,35(4):84 - 87.

Segmentation Algorithm for Bone Marrow Cell Image Based on The Wavelet Transform and K-means Clustering

SU Shi-mei<sup>1</sup>, LV Xue-yang<sup>2</sup>

(School of Electrical Engineering, Zhengzhou University, Zhengzhou 450001, China)

**Abstract:** A segmentation method based on the wavelet analysis is proposed in this paper for segmenting visible components from the bone marrow cell. Firstly, the wavelet transform is used to erase the effect of defocusing. Secondly, *G* component is chosen as clustering the information of characteristic parameters by using the threshold multi-scale wavelet analysis, which is combined with saturation component according to different distribution characteristic of leukocyte nucleus to construct a transpositional image named STG. The nucleus are extract by thresholding the image. Ultimately, it can provide an optimal initial clustering centers for K-means clustering method to segment the red blood cells and the white blood cells from the background as well as separate them from each other. Through comparative analysis and algorithm testing, this method can effectively overcome the difficulties of complex components、poor discrimination and complicated defocusing during the cells segmentation, the accuracy rate reaches 94.15%.

**Key words:** cell segmentation; wavelet transform; K-means clustering; color feature