

显微细胞图像有形成分自动识别^①

漆鹏杰¹, 刘秀波¹, 仲兆准², 管 淼², 谢光伟¹

¹(苏州大学 机电工程学院, 苏州 215021)

²(苏州大学 沙钢钢铁学院, 苏州 215021)

摘 要: 研究了一种显微细胞图像有形成分自动识别方法, 首先针对显微细胞复杂成分的现状, 运用了相关的图像预处理方法, 明显的提高了图像的对比度和清晰度. 随后着重研究了在显微细胞图像处理时的图像分割、特征提取以及细胞特征识别等方法, 得到了一套算法流程, 它能够快速、准确的识别细胞. 最后用该方法对常见的四种细胞图像样本进行测验, 实验结果验证了该方法的有效性.

关键词: 显微细胞; 有形成分; 自动识别; 图像处理

Visible Components Automatic Recognition of Microscopic Cell Image

QI Peng-Jie¹, LIU Xiu-Bo¹, ZHONG Zhao-Zhun², GUAN Miao², XIE Guang-Wei¹

¹(College of Mechatronic Engineering, Soochow University, Suzhou 215021, China)

²(College of Iron & Steel, Soochow University, Suzhou 215021, China)

Abstract: Automatic identification of microscopic cell image is studied in this paper. Firstly, considering the status of the complicated cellular components, the corresponding method is applied to image preprocessing, the contrast and resolution of the image is significantly improved. Then, the image segmentation, feature extraction and cell identification are emphatically studied, gaining a set of algorithms process to identify cell speedy and precisely. Finally, this method is tested by using four kinds of common cells samples, the results verify the effectiveness of the method.

Key words: microscopic cells; visible components; automatic identification; image processing

显微细胞检查包括血液检查、尿常规检查、体液检查等, 它是指通过用显微镜观察并分析细胞成分中的红细胞、白细胞、上皮细胞、结晶等有形成分来判断人体的身体健康状况. 传统的显微细胞检测大多为人工镜检, 这种方法操作时间长、工作强度大、易污染环境; 受主观的影响, 人为误差大、标准不统一, 所以手工镜检已经难以快速、准确、客观的进行检查, 不利于临床动态观察^[1,2].

近年来, 随着人工智能研究的发展和生物学图像处理技术的提高, 对显微细胞图像采用图像处理技术进行有形成分的自动识别已成为现实. 在面对细胞成分复杂、粘连严重、散焦严重、对比度不高的情况下, 提出了与之对应的图像预处理、图像分割、特征

提取和自动识别技术算法. 这大大降低了医务人员的工作强度, 提高了诊断的有效性和准确性^[3]. 本文主要通过研究和比较传统的图像预处理方法、阈值分割法和边缘检测法以及数学形态学等方法在显微细胞特征提取中的应用, 提出了一种基于 Sobel 算子边缘提取结合形态学操作的特征分割方法. 并利用 BP 神经网络误差训练方法对显微细胞特征进行识别并分类.

1 显微细胞图像识别方法的基本流程

本文研究的显微细胞有形成分自动识别方法的流程主要包括显微细胞图像获取、图像的预处理、图像分割、特征提取和有形成分识别等五个方面, 其自动识别流程如图 1 所示.

^① 基金项目:国家自然科学基金(61304095,51204115);江苏省自然科学基金(BK20130317);江苏省博士后科研资助计划(1302103B)

收稿时间:2014-09-01;收到修改稿时间:2014-10-24

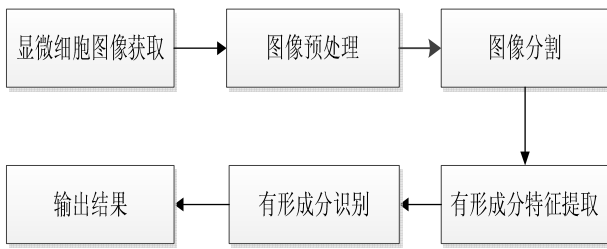


图 1 显微细胞识别方法基本流程

首先使用分辨力极高的显微成像系统自动采集细胞有形成分的图像，并保存在计算机中就获得了有形成分的图像^[4]。再经过适当的预处理，从而改善图像质量。然后对预处理图像有形成分定位，分割出其有效边缘，以便于下一步的特征提取并分类识别^[5]。最后通过适当的选取特征，就可以使用分类器模型对显微细胞进行分类识别了。

2 显微细胞图像预处理与分割

2.1 图像预处理

在显微细胞图像获取的过程中，由于各方面原因的影响，在得到的图像中难免的会混入噪声。这不仅会对造成图像质量的下降，还会对后续的图像处理带来不便，因此需要去除噪声。图像噪声一般表现为高频形式，因此采用低通滤波对图像进行去噪。常见的去噪方法主要有均值滤波和中值滤波。均值滤波能有效抑制噪声，但会使图像边缘模糊；中值滤波是一种去除噪声的非线性处理方法，它能够在抑制随机噪声的同时不模糊边缘，但在处理细节特别多的图像时效果不太理想，同时也会带来降低图像分辨率的影响^[6]。本文经过实验对比后采用中值滤波去噪，选取了合适的中值滤波窗口尺寸对图像进行处理，既消除了显微细胞图像在数字化和传输过程中所混入的噪声又不致使图像模糊加重而对后续的处理造成影响。

为了增强图像的对比度，运用图像增强的方法对图片进行预处理，主要方法有灰度变换、直方图均衡化、拉普拉斯滤波、高低帽变换等。本文采用的是直方图均衡化的方法，它是利用灰度变换来自动调节图像对比度质量，其基本思想是通过灰度级的概率密度函数求出灰度变换函数。在经过直方图均衡化后的图像具有更加清晰，能够显示出更多细节信息的优点。

本文选取了四种不同类别较为典型的显微细胞图像来进行该方法的实验，图 2 为细胞原图与经中值滤

波去噪和对比度增强后取得的图像。

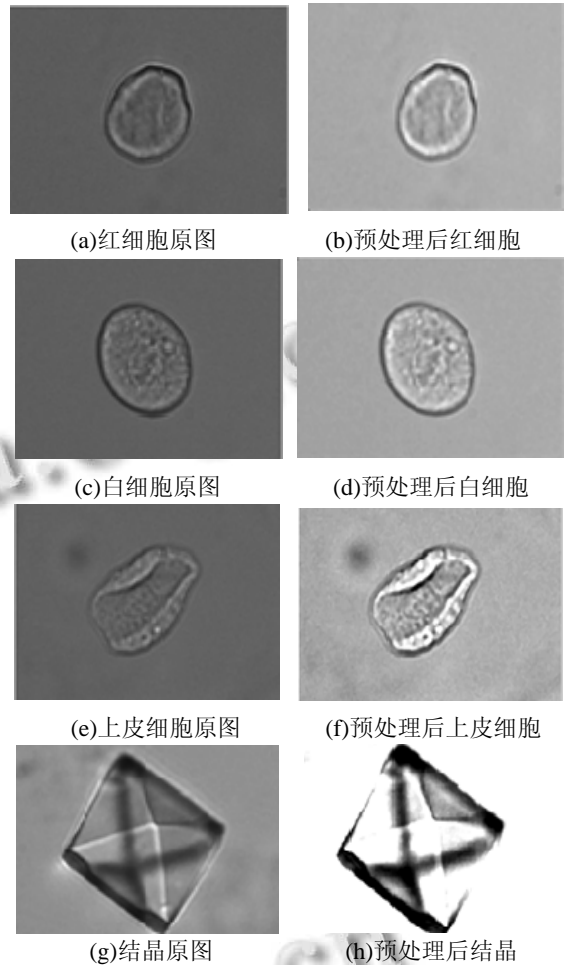


图 2 显微细胞原图和预处理后图

由上图可以看出，在经过中值滤波和图像增强之后，图像的对比度和清晰度明显提高了，这有利于后面进一步的处理。

2.2 图像分割

原始图像经预处理后，就需要对其进行图像分割，以定位出所需的部分。区分图像中的背景与细胞并将其分离，尽可能完整地提取细胞特征，从而将细胞从图像背景中提取出来^[7]。这是由图像处理到图像分析的关键步骤，对特征测量有重要的影响。本文主要采用阈值分割法(迭代阈值、Otsu 阈值)、边缘检测法(Sobel、Roberts、Canny 算子)以及数学形态学方法(膨胀、闭合、填充)来对图像进行处理。

2.3 分割方法

1. 迭代全局阈值分割: 开始时选择一个阈值作为初始估计值，然后按某种策略不断的改进这一估计值，

直到满足给定的准则为止. 好的策略应具备两个特征: 一是能够快速收敛, 二是在每一个迭代过程中, 新产生的阈值优于上一次的阈值, 其步骤如下.

- 1) 选择图像灰度的中指作为初始阈值 T_0 .
- 2) 利用阈值 T 把图像分割成两个区域— R_1 和 R_2 ,

用下式计算区域和的灰度均值 μ_1 和 μ_2 :

$$\mu_1 = \frac{\sum_{i=0}^{T_i} i n_i}{\sum_{i=0}^{T_i} n_i}, \quad \mu_2 = \frac{\sum_{i=T_i+1}^{L-1} i n_i}{\sum_{i=T_i+1}^{L-1} n_i} \quad (1)$$

- 3) 计算出 μ_1 和 μ_2 后, 用下式计算出新的阈值 T_{i+1} :

$$T_{i+1} = \frac{1}{2}(\mu_1 + \mu_2) \quad (2)$$

- 4) 重复上面步骤(2)和(3), 直到 T_{i+1} 和 T_i 的差小于某个给定值.

其分割图如图 3 所示.

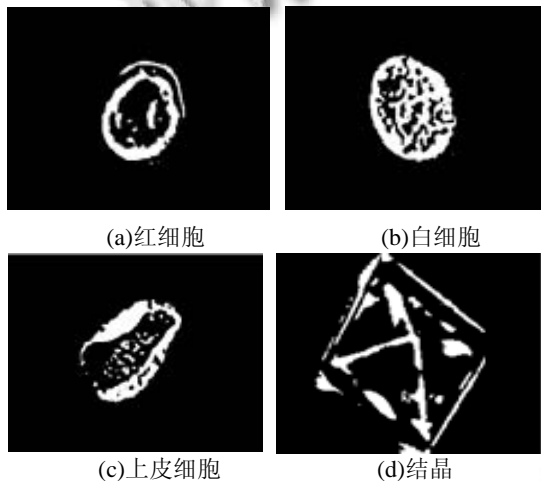


图 3 全局阈值分割

2. Otsu 阈值分割^[8]: Otsu 法是一种使类间方差最大的自动确定阈值的方法, 该方法具有简单、处理速度快的特点, 是一种常用的阈值选取方法. 其基本思想如下: 设图像像素为 N , 灰度范围为 $[0, L-1]$, 对应灰度级 i 的像素为 N_i , 几率为:

$$p_i = n_i / N, i = 0, 1, 2 \dots L-1 \quad (3)$$

把图像中的像素按灰度值用阈值 T 分成两类 C_0 和 C_1 , C_0 由灰度值在 $[0, T]$ 之间的像素组成, C_1 由灰度值在 $[T+1, L-1]$ 之间的像素组成, 对于灰度分布几率, 整幅图像的均值为: $u_T = \sum_{i=0}^{L-1} i P_i$ (4) 则 C_0 和 C_1 的均值为:

$$u_0 = \sum_{i=0}^T i P_i / \varpi_0, \quad u_1 = \sum_{i=T+1}^{L-1} i P_i / \varpi_1 \quad (5)$$

$$\text{其中: } \varpi_0 = \sum_{i=0}^T P_i, \quad \varpi_1 = \sum_{i=T+1}^{L-1} P_i = 1 - \varpi_0 \quad (6)$$

由上面的式子可得:

$$u_T = \varpi_0 u_0 + \varpi_1 u_1 \quad (7)$$

类间方差定义:

$$\sigma^2 = \varpi_1 \varpi_0 (\mu_0 - \mu_1)^2 \quad (8)$$

让 T 在 $[0, L-1]$ 范围依次取值, 使 σ^2 最大的 T 值即为 Otsu 法的最佳阈值. 其分割图如图 4 所示.

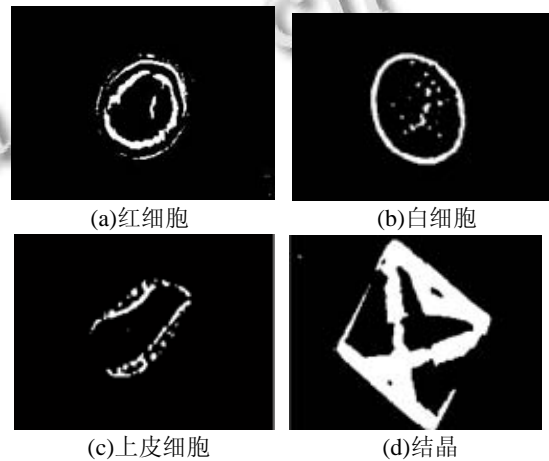


图 4 Otsu 阈值分割

实验结果表明, 经典二值法进行分割时容易受背景干扰, 且形成的二值图的边缘不够连续, 特征不够完善. 为了更好的得到有形成分的有效特征, 下面采用边缘检测算子来对预处理细胞图像进行分割实验.

3. 索贝尔(Sobel)边缘算子: 索贝尔算子是一组方向算子, 从不同的方向检测边缘. 它不是简单求平均再差分, 而是加强了中心像素上下左右 4 个方向像素的权重^[9], 运算结果是一副边缘图像, Sobel 算子模板如图 5 所示.

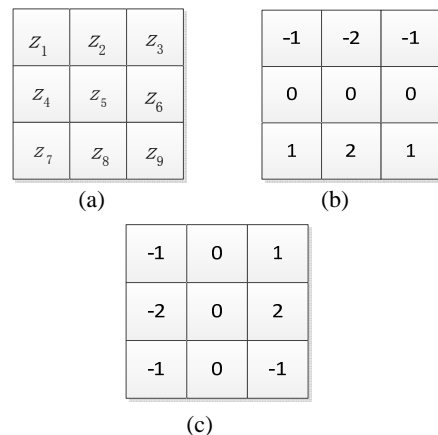


图 5 Sobel 算子模板(Z 为灰度值)

该算子的计算公式如下:

$$G_y = (Z_3 + 2Z_6 + Z_9) - (Z_1 + 2Z_4 + Z_7) \quad (9)$$

$$G_x = (Z_7 + 2Z_8 + Z_9) - (Z_1 + 2Z_2 + Z_3) \quad (10)$$

$$|G(x, y)| = \sqrt{G_x^2 + G_y^2} \approx |G_x| + |G_y| \quad (11)$$

其分割图如图 6 所示.

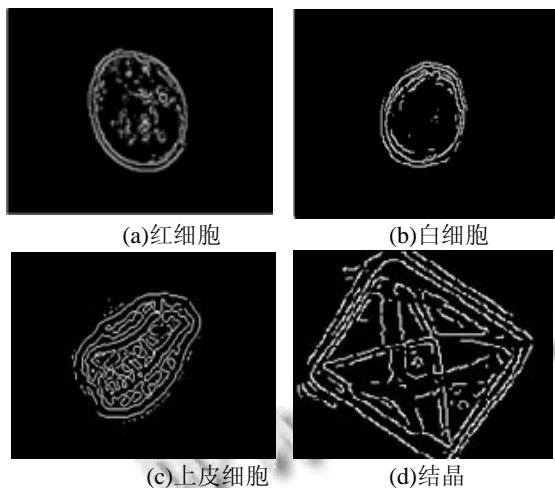
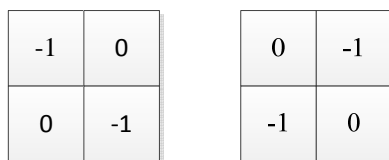


图 6 Sobel 算子分割

4. 罗伯特(Roberts)边缘算子: 罗伯特算子是一种斜向差分的梯度计算方法, 梯度的大小代表边缘的强度, 梯度的方向与边缘走向垂直. Roberts 交叉梯度算子由图 7 模板 (a)和模板(b)分别计算出 G_x 和 G_y , 然后求得梯度向量.



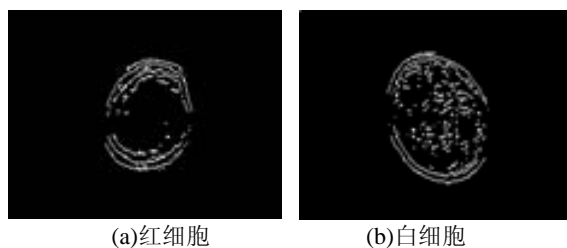
(a) (b)
图 7 Roberts 算子模板

该算子的计算公式如下:

$$G_x = (Z_9 - Z_5), \quad G_y = (Z_8 - Z_6) \quad (12)$$

$$|G(x, y)| = [(Z_9 - Z_5)^2 + (Z_8 - Z_6)^2]^{\frac{1}{2}} \quad (13)$$

其分割效果如图 8 所示.



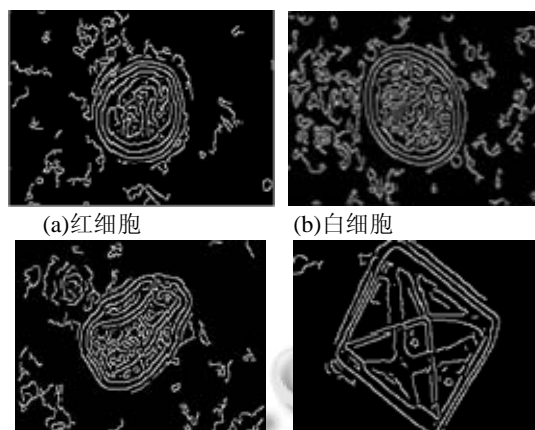
(a)红细胞 (b)白细胞



(c)上皮细胞 (d)结晶

图 8 Roberts 算子分割

5. 坎尼(Canny)边缘算子: 先对处理的图像选择一定的高斯滤波器进行平滑滤波, 抑制图像噪声; 然后采用一种称之为“非极值抑制”(Nonmaxima Suppression)的技术, 细化平滑后的图像梯度幅值矩阵, 寻找图像中的可能边缘点; 最后利用双门限检测通过双阈值递归寻找图像边缘点, 实现边缘提取. 其分割效果如图 9 所示.



(a)红细胞 (b)白细胞 (c)上皮细胞 (d)结晶

图 9 Canny 算子分割

实验结果表明, Roberts 算子提取的边缘不够完善, 没能很好的提取出图像的轮廓. Canny 算子提取的边界过于细致, 且杂点影响较多, 干扰太大. Sobel 算子提取的边界相对来说较好, 较完整的表达了细胞的轮廓, 所以本文选取 Sobel 算子分割.

2.4 形态学操作

在经过图像分割之后, 得到的分割图往往有一些断裂点以及不均衡的地方, 为了更好的提取细胞特征, 就需要对二值图断裂点的连接、填充以及平衡等操作^[10]. 这些操作可以通过数学形态学的方法做到, 本文用到的形态学方法有膨胀、闭合和填充. 膨胀是将与物体接触的所有背景点合并到该物体中, 使边界向外部扩张的过程. 通过膨胀, 可以填充图像

中的小孔及在图像边缘处的小凹陷部分. 结构元素 B 对 A 的膨胀, 记作 $A \oplus B$, 定义为:

$$A \oplus B = \{X : \hat{B}_x \cap A \neq \emptyset\} \quad (14)$$

闭合是用来填充物体内部细小空洞、连接邻近物体、平滑其边界的同时并不明显改变其面积. 结构元素 B 对 A 的开运算, 记作 $A \bullet B$, 定义为:

$$A \bullet B = (A \oplus B) \ominus B \quad (15)$$

填充是从边界中求得目标区域, 以准确定位出所需的部分, 以上形态学操作处理效果如图 10、11、12 所示.

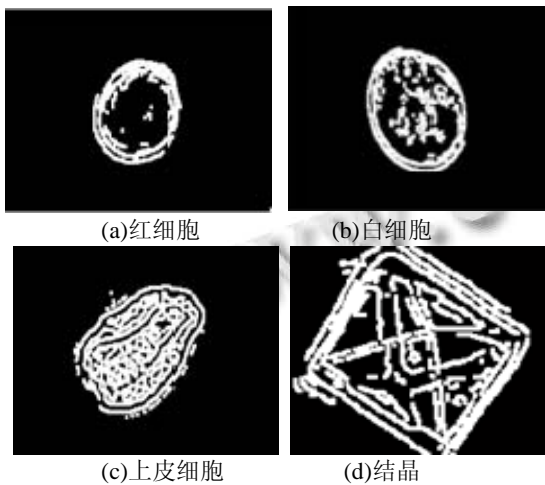


图 10 膨胀后图

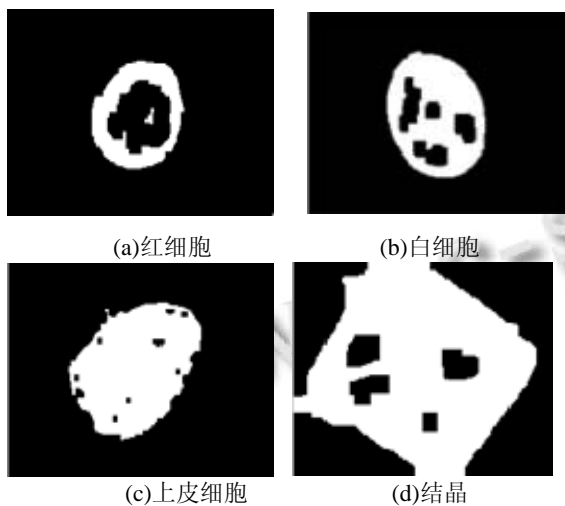


图 11 闭合后图

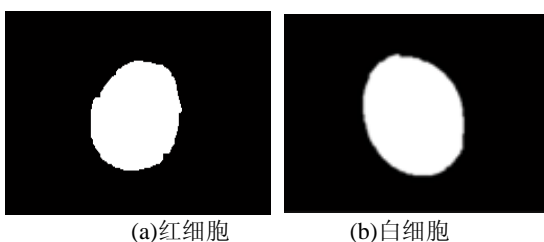


图 12 填充后图

从以上实验结果可以看到, 经过形态学处理后我们得到了较为精细的有形成分二值图像, 从中可以得到显微细胞的形态特征, 说明本文的分割方法还是比较好的.

3 显微细胞图像的特征提取

把显微细胞图像有形成分分割出来后, 为了进行后面的模式识别, 即使用计算机系统, 从图像中提取有用的数据或信息, 生成非图像的描述或表示, 如数值、符号等, 从而识别图像. 所以这里需要对分割的有形成分进行特征提取与选择, 将这些特征转换为计算机能识别的数字信息, 这样计算机才能对有形成分进行自动识别.

3.1 主要图像特征

图像的特征是指图像的原始特征或属性, 每一幅图像都有其本身的特征, 常用的图像特征有几何形状特征、灰度特征和纹理特征.

1) 几何形状特征: 用来描述有形成分的几何形状特性. 主要包括面积 A : 目标包含的像素数; 周长 P : 目标的边界包含的像素数; 圆形成度 C : 面积的 4π 倍与周长平方的比值; 矩形成度 R : 描述区域与矩形的偏离程度; 伸长度 E : 区域的高度和宽度之比, 为狭长度; 占空比 Ar : 图像的二值图非零元素的面积 Area_Binary 与最小外接矩形的面积 Area_Minrect 之比.

2) 灰度特征: 灰度特征主要是基于灰度直方图得到. 它主要有灰度平均值 \bar{f} : 图像的平均灰度; 灰度方差 σ^2 : 表示区域灰度的弥散程度; 灰度熵 HE : 图像所包含的灰度信息; 歪斜度 SK : 图像直方图的不对称程度; 峰态 KU : 图像的灰度分布在接近均值时的大致状态.

3) 纹理特征: 在显微细胞有形成分识别中纹理特征是比较重要的特征, 纹理描述了图像中的某种规律性模式, 它是由许多相互接近的、相互编织的元素构

成, 并常富有周期性, 也可以被认为是灰度(颜色)在空间以一定的形式变化而产生的图案(模式), 它是真实图像固有的特征之一^[11]. 纹理特征能很好的区分用形状特征不能区分的有形成分. 主要有对比度 *CON*: 反应图像的清晰程度; 能量 *WE*: 也叫角二矩阵, 它是灰度共生矩阵像素值平方的和, 反映的是图像灰度分布均匀性的度量; 相关系数 *COR*: 衡量灰度共生矩阵的元素在行的方向或列的方向的相似程度; 熵 *ENT*: 是图像所具有的信息量的度量; 逆差矩 *IDM*: 反映图像纹理的同质性, 度量图像纹理局部变化的多少.

3.2 特征的选择

为了能够快速、准确的识别细胞, 必须选择有效的特征. 对于一个特征而言, 评判的标准有 4 个方向, 分别是可区别性、可靠性、独立性好、数量少. 在特征选择时, 必须尽可能的减少整个识别系统的处理时间和错误识别率. 本文在通过实验之后, 选择了面积 *A*、周长 *P*、圆形度 *C*、矩形度 *R*、灰度平均值 \bar{r} 、灰度方差 σ^2 、灰度熵 *HE*、歪斜度 *SK* 以及以上的五个纹理特征值共 13 个特征量, 它们能很好的区分显微细胞图像.

3.3 特征的提取

在选择以上的几个主要特征量之后, 我们就要进行图像的特征提取, 这里我们借助于常用的软件 *Matlab* 来实现显微细胞有形成分主要特征量的提取. 特征提取主要是针对几何形状、灰度、纹理这三类特征来进行. 本文以红细胞为例, 来进行细胞特征的提取.

几何形状特征主要是基于显微细胞的灰度图像和分割后的二值化图像得到的, 通过它们的图像可以求得显微细胞的面积和周长, 再利用软件并根据公式就可以求出其它几类几何形状特征, 图 13 和图 14 分别是红细胞的灰度图和二值化图.

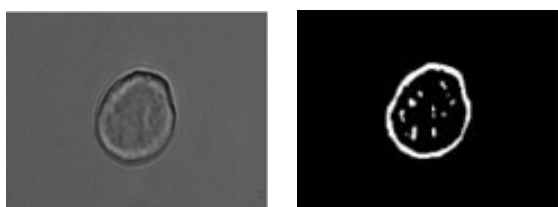


图 13 红细胞灰度图 图 14 红细胞二值图

灰度特征主要是根据细胞图像的灰度直方图求得,

得到红细胞图像的直方图后根据公式计算灰度特征. 红细胞的灰度直方图如图 15 所示.

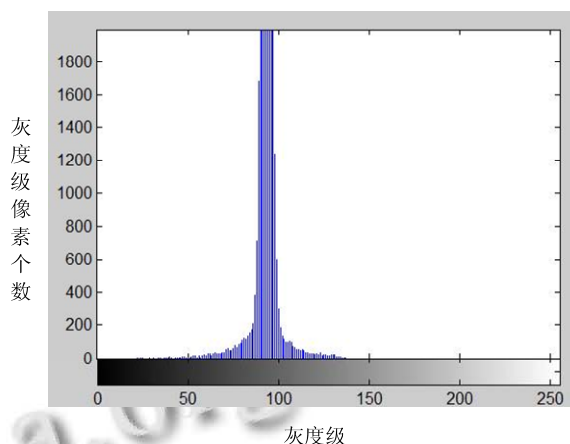


图 15 红细胞灰度直方图

纹理特征是对红细胞灰度图像进行压缩, 再求得四个灰度共生矩阵, 分别求得每个灰度共生矩阵的相关纹理特征, 对其取平均值最终得到纹理特征量.

根据上述的方法对其它几类细胞做同样处理, 就可以提取出它们的特征值. 最后提取出的特征量都在表 1 中给出, 表中所有特征量都进行了归一化处理.

表 1 细胞特征提取

特征代号	红细胞	白细胞	上皮细胞	结晶
面积 <i>A</i>	258.687	483.263	546.314	374.265
周长 <i>P</i>	58.420	81.049	112.093	90.358
圆形度 <i>C</i>	0.952	0.924	0.5461	0.5887
矩形度 <i>R</i>	0.687	0.630	0.738	0.9427
灰度平均值 \bar{r}	92.9161	95.7031	97.6889	75.5537
灰度方差 σ^2	7.2467	8.3038	7.0404	18.1704
灰度熵 <i>HE</i>	4.1580	4.2119	3.0416	5.4178
歪斜度 <i>SK</i>	0.0349	0.2470	0.1476	0.4329
对比度 <i>CON</i>	0.3699	0.3501	0.1784	0.2413
相关性 <i>COR</i>	3.0745	3.1299	2.5739	2.9174
纹理熵 <i>ENT</i>	3.1297	3.6453	2.0140	3.0476
平稳度 <i>IDM</i>	1.4355	1.4962	3.7189	3.9432
能量 <i>WE</i>	0.1427	0.3984	0.5341	0.2468

4 显微细胞图像识别

在提取了细胞有形成分的各个特征之后, 还需要将其特征输入到分类器中对其进行分类并识别. 本文中, 使用了反向传播(back propagation, BP)神经网络^[12]作为纹理特征的分类器, 来对各个有形成分进行识别.

4.1 BP 神经网络概述

BP 神经网络是一种按误差逆向传播算法训练的多层前馈网络,是目前应用最广泛的神经网络模型之一.该网络的训练采用误差反向传播算法,即神经元的连接权重的训练是从最后一层(输出层)开始,然后反向依次更新前一层连接权重.它的学习规则是使用快速下降法,通过反向传播来不断调整网络的权重,使网络的误差平方和最小.它的拓扑结构包括输入层、隐含层和输出层^[13].其网络结构如图 16 所示.

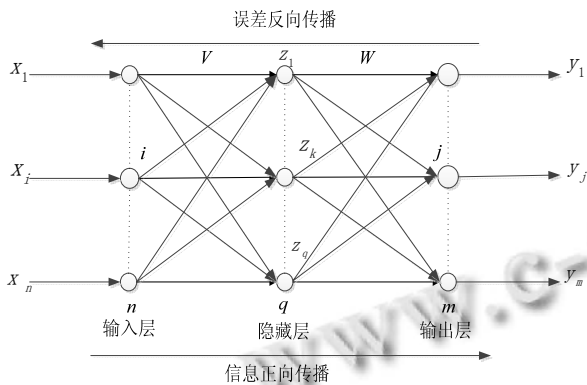


图 16 BP 神经网络结构图

把可靠的训练样本集依次输入 BP 神经网络,经过反复训练,当 BP 神经网络的误差低于设定的值时,训练过程结束.识别时,只需要把待识别的样本输入到训练好的神经网络,从网络的输出即可判定该样本的类型.

4.2 基于 BP 神经网络的显微细胞图像识别过程

本文利用 BP 神经网络分类器对细胞有形成分进行识别,专门设计了一个工作流程(如图 17 所示),以完成对输入的特征向量进行分类,从而达到识别的目的.

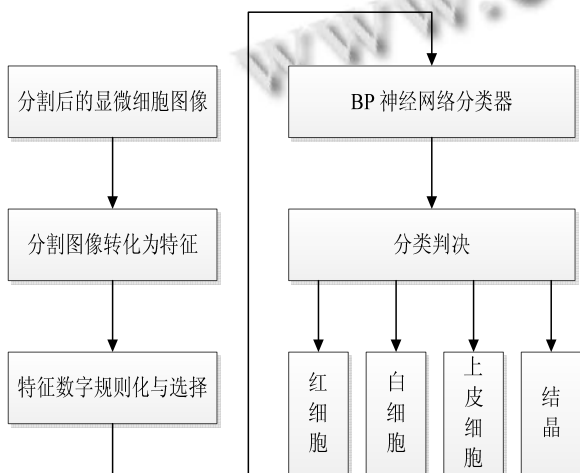


图 17 BP 神经网络分类器识别分类过程

4.3 BP 神经网络分类器结构设计及对显微细胞图像的识别实验

本文设计的 BP 神经网络分类器采用的是由输入层、隐藏层和输出层三部分组成的.对该 BP 神经网络分类器进行设计首先要确定分类器输入层、隐藏层和输出层各层中的神经元个数.通常,样本的特征数目是多少,输入层神经元的个数就为多少,本文中共选取了 13 个细胞特征量,所以输入层神经元的个数设置为 13.中间层为隐藏层,设计隐藏层神经元个数是个难点,若设计个数较少,神经网络从样本中获取特征的能力就减弱,不足以体现和概括训练样本集的特征;若设计个数过多,又可能将样本中的非特征内容记入神经网络中,出现过适应状态,从而降低神经网络的泛化能力.目前常用实验试凑法来确定隐藏层神经元个数,经过多次的实验验证,最终确定其最佳神经元个数为 9 个.输出神经元的个数即为待分类细胞样本的种类个数,设置为 4 个.

最后将提取的显微细胞有形成分样本的特征量送入 BP 神经网络分类器进行识别分类. BP 神经网络分类器训练采用 Levenberg-Marquardt 的改进算法,借助于 Matlab 仿真软件实现.本文选取了四类细胞各 50 个作为训练样本,对其进行训练,下图 18 反应了神经网络的训练过程.

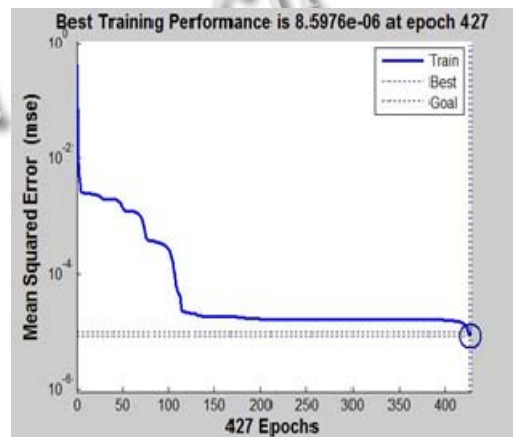


图 18 神经网络训练过程

在 BP 神经网络分类器通过细胞样本训练学习后,就可以用细胞测试样本对该分类器进行测试了.将已经提取的各类细胞特征量送入到 BP 神经网络器进行识别与分类.用 BP 神经网络分类器对表 2 中的样本进行测试,测试结果由表 3 给出.

表2 测试样本数目

序号	细胞类型	样本数目
1	红细胞	180
2	白细胞	95
3	上皮细胞	19
4	结晶	48
合计		342

表3 识别测试结果

样本类别	测试样本数目	正确识别数目	识别率
红细胞	180	154	85.6%
白细胞	95	84	88.4%
上皮细胞	19	17	89.5%
结晶	48	42	87.5%
合计	342	297	86.8%

由上表可知,此显微细胞的识别率基本在85%以上,识别的准确率能够达到预期的效果,该细胞自动识别方法能满足实际的识别要求。

5 结语

本文通过对显微细胞图像处理方法的研究比较,提出了一种基于 Sobel 边缘提取结合形态学操作的二值化分割方法,有效的实现了显微细胞有形成分的分割、定位。再结合相关图像处理技术,很好的选取了特征,最后采用 BP 神经网络识别器进行识别,得到了较高的识别准确率。所以本文的显微细胞图像识别方法在医学上有一定的实用价值。

参考文献

- 1 顾可梁.尿有形成分的识别与检查方法的选择.中华检验医学杂志,2005,28(6):572-575.
- 2 梁光明.体液细胞图像有形成分智能识别关键技术研究[硕士学位论文].长沙:国防科学技术大学,2008.
- 3 刘志文,安兴,李衡,时永刚,黄亚丽,杨婷.显微细胞图像分析方法的研究进展.北京理工大学学报,2014,34(5):441-453.
- 4 胡晓泊,张鹏,安稷,张丹丹,于君君,林杰.基于计算机视觉的医学图像自动识别技术研究.微计算机信息,2012,28(10):113-115.
- 5 赵少敏.尿沉渣图像的分割与识别算法的研究[硕士学位论文].重庆:重庆大学,2010.
- 6 陈建军,于军,廖红华.细胞显微图像处理研究.农机化研究,2011,33(2):24-27.
- 7 李盛阳.医学细胞图像分割分析方法研究[硕士学位论文].济南:山东科技大学,2003.
- 8 吴丹,蔡晓东,谢月飞,曾威.基于直方图分析和 OTSU 算法的文字图像二值化.计算机与现代化,2013,(7):117-119,234.
- 9 赵芳,栾晓明,孙越.数字图像几种边缘检测算子检测比较分析.通信与信息处理,2009,28(3):71-73.
- 10 沈晶,刘海波,周长健.Visual C++ 数字图像处理典型案例详解.北京:机械工业出版社,2012.
- 11 许宗敬,胡平.显微图像纹理特征提取方法综述.微计算机应用,2009,30(6):6-13.
- 12 蔡永军,刘伟玲,虞启琰.遗传神经网络在尿沉渣识别中的应用.医疗卫生装备,2004,25(11):1-2.
- 13 梁曼君,石竹.提高BP神经网络学习速度的算法研究.合肥工业大学学报,1995,12(4):59-63.