第4卷 第4期
1984年4月

光 学 学 报
ACTA OPTICA SINICA

Vol. 4, No. 4
April, 1984

利用功率谱作模式识别
——以肺癌细胞筛选为例

张以谟 黄 智
(天津大学精密仪器系)

提 要

本文讨论了用功率谱采样的方法对于同一类物体进行识别和筛选的原理。以肺癌细胞筛选为例，提出用长焦距、高倍率的傅里叶变换系统和线性判别处理。由微计算机控制采样过程并作统计判别，完成了光学计算机混合处理系统，实现了肺癌细胞的自动筛选。

利用光学功率谱采样作模型式识别，方法简单、价格便宜，在医学诊断、遥感图象分析、工业产品检验等许多方面有广泛的应用价值。曾有一些学者作过研究。美国 B. Pernick 采用相干光处理和 Baysian 数据处理方法，对宫颈细胞作了筛选[1]。国内复旦大学赵焕卿等人对带菌细胞的功率谱作了采样检测。

最近，针对癌症早期诊断和普查中的细胞学自动检查问题，我们以痰脱落细胞中癌细胞的自动识别为题，采用长焦距、高倍率的傅里叶变换系统和线性判别处理，用微计算机控制采样过程和细胞识别过程，对肺癌细胞的自动识别作了研究。完成了光学计算机混合处理系统，初步实现了肺癌细胞的自动筛选。

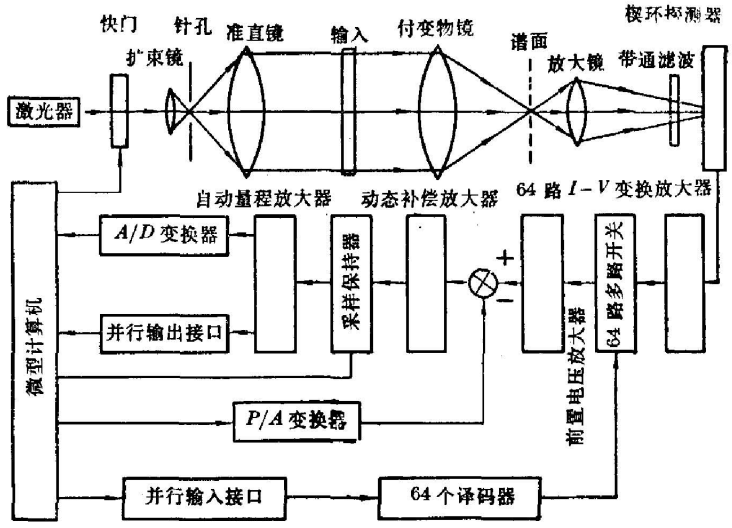


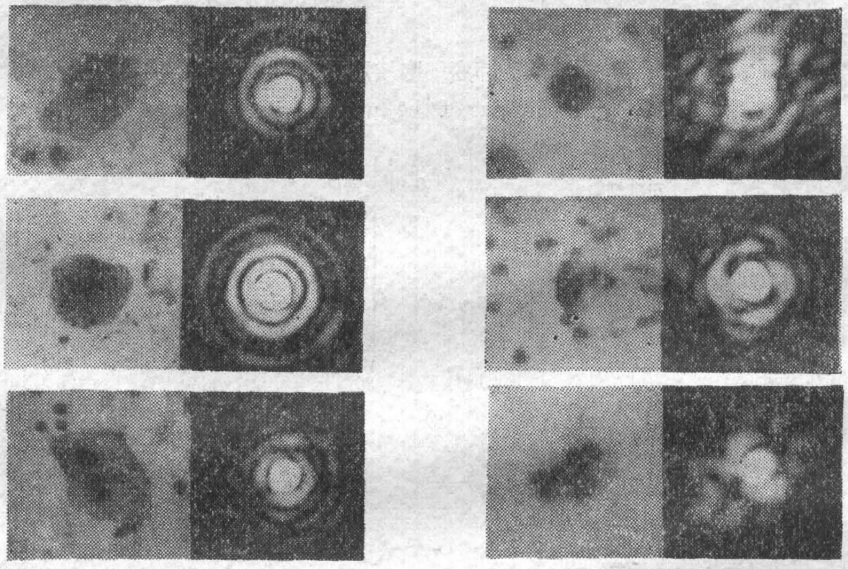
图1 光学计算机混合处理系统

Fig. 1 Hybrid processing system of optics and computer

收稿日期：1983年8月13日；收到修改稿日期：1983年11月2日

图 1 为光学计算机混合处理系统的示意图。将经显微摄影得到的黑白细胞底片放在输入位置上，在傅氏变换物镜的后焦面上可得到细胞图象的功率谱。通过适当的处理，可得到带有明显区别的正常细胞和癌细胞的功率谱。

图 2 给出了两类细胞的图谱照片。从图中可以看出正常细胞的谱由较规则的衍射圆环组成，它基本保持了圆形物体谱的特征。正常细胞的衍射环有粗细相间的现象。可以这样理解，整个谱是由底片上细胞核区、浆区和背景区所形成的谱迭加而成的。迭加后有的衍射环加强，有的减弱，加强的变粗，减弱的变细，而结果使衍射环变得粗细相间。



(a)　　(b)

图 2　细胞及其功率谱

Fig. 2　Cells and its spectrum

(*a*) Normal cells and its spectrum; (*b*) Cancer cells and its spectrum

癌细胞的谱同其本身的形态相联系，变化很大。总的来看，其谱基本失去了圆形物体谱的特征。衍射花样杂乱无章。谱的中心级较大，两个衍射极大之间没有明显的极小出现。谱的方向性较强。

根据多次实验的结果，我们认为细胞的谱主要受细胞核的几何形状及核内物质均匀程度的影响。底片上，细胞核区的透光率最强，在透过底片的光场中占的比例最大。正常细胞的衍射环主要是基本上呈圆形的细胞核形成的，而癌细胞衍射花样杂乱无章很大程度是受细胞核内脱氧核糖核酸团块的影响。

我们观察了带有核糖核酸团块的癌细胞核的谱。图 3 给出了其图谱照片。可以看到细胞核的谱杂乱无章，成散射状。这个结果说明了癌细胞的谱变得杂乱确是受其核内物质的影响。

图 4 为细胞功率谱的测量曲线。测量是以极坐标进行的。图 4(*a*)为径向曲线，4(*b*)为角向曲线。其中径向测量是在 80～130(周/mm)的频域中进行的，角向测量是在 70～150(周/mm)的频域中进行的。上述频域是经多次实验所确定的两类细胞差异较为突出的区域。

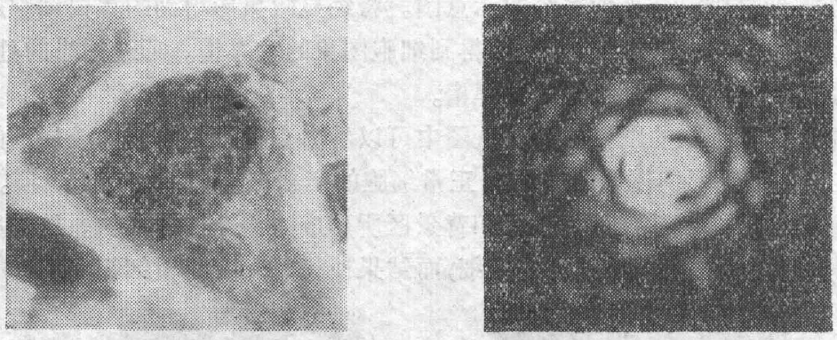


图 3 肺癌细胞核及其谱

Fig. 3 Cancer nucleus of the lung and its spectrum

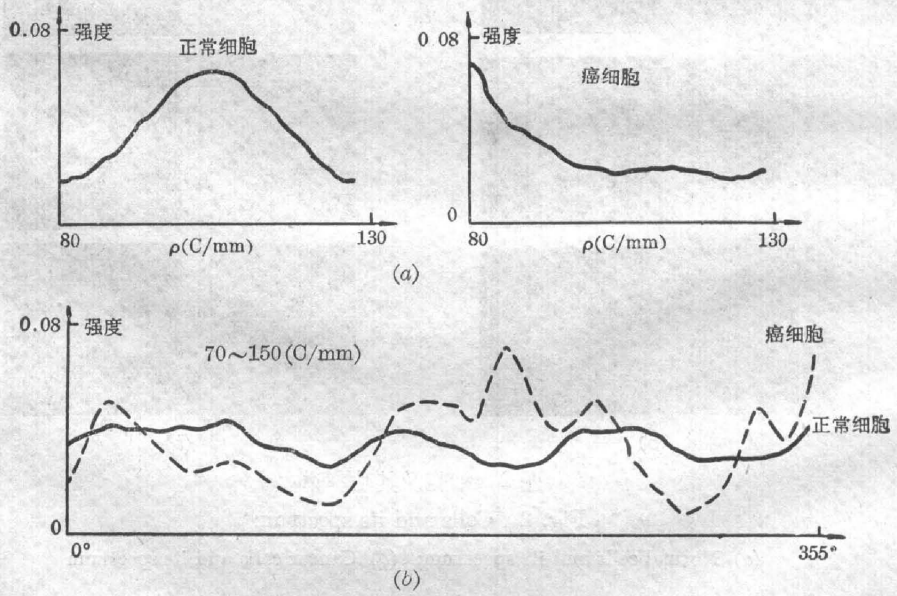


图 4 谱的扫描曲线

Fig. 4 Scanning curves of spectrum

(*a*) Radial scanning curve; (*b*) Angular scanning curve

可以看到在径向曲线中，对于正常细胞曲线是向上凸的，而癌细胞曲线是向下凸的。两类曲线起伏的程度也不同。谱的角向曲线差异主要表现在离散程度上。正常细胞的曲线抖动小，比较平稳；而癌细胞的曲线跳动很大。这说明癌细胞的谱方向性较强。

要由计算机自动完成细胞识别，必须把两类谱的差异表示成数字化的特征。

在模式识别中，特征抽取十分重要。无论选择哪种判别方法，都是以良好的特征抽取为基础的。但在模式识别的理论中，至今还没有特征抽取的一般性方法[2]。实际的抽取往往是一个物理模型分析、统计计算和判别效果检验相结合的复杂过程。我们采用的是物理模型分析和统计计算相结合的方法。最后选出了六个特征表示两类谱的差异。

细胞涂片中正常细胞染色淡，癌细胞染色浓。反映在细胞底片上癌细胞区域密度小，透过率大；正常细胞密度大，透过率小。这说明入射光场透过底片总的光能量会有差异。故选择总能量作为第一个参量。分析谱的径向分布可以得到三个参量。曲线的凸向不同说明斜

率有差异；平滑程度不同说明离散有差异。又因用径向频率加权后差异更为突出，故选择斜率值为第二个参量，选择径向分布离散及加权的径向分布离散为第三个和第四个参量。从谱的角向分布中可以得到最后两个参量。癌细胞沿谱方向性强，而正常细胞角向分布比较均匀。故选择角向分布离散为第五个参量；角向分布最大和最小值之差为第六个参量。为消除外界随机因素比如激光器输出功率、环境光的微小变化所产生的干扰，除能量参量外，所有的参量都对总能量归一。

下面给出六个特征参量的理论公式。

1. 总能量 x_1

$$x_1=\int_0^{\pi}d\theta\int_0^{\rho}\rho|F(\rho,\theta)|^2d\rho, \tag{1}$$

其中 ρ、θ 分别为极坐标系下的空间频率坐标和角坐标，$F(\rho,\theta)$为谱上(ρ,θ)处的光振幅。

2. 斜率 x_2（用 ρ 加权）

$$x_2=\frac{1}{x_1}\left\{\int_0^{\pi}d\theta\int_a^{b}\rho^2|F(\rho,\theta)|^2d\rho-\int_0^{\pi}d\theta\int_c^{d}\rho^2|F(\rho,\theta)|^2d\rho\right\}, \tag{2}$$

其中 a、b、c、d 为预先确定的频域区间。式 (2) 是斜率的近似表示。分母中略去了表示频宽的因子。这是因为对于两类细胞频宽相等。略去共同的比例因子对识别没有影响。

3. 径向离散 x_3

$$x_3=\frac{1}{x_1^2}\operatorname*{avg}_{\rho=e\to f}\left\{\int_0^{\pi}d\theta\int_{\rho}^{\rho+\Delta\rho}\rho|F(\rho,\theta)|^2d\rho-\operatorname*{avg}_{\rho=e\to f}\left[\int_0^{\pi}d\theta\int_{\rho}^{\rho+\Delta\rho}\rho|F(\rho,\theta)|^2d\rho\right]\right\}^2, \tag{3}$$

其中 $\Delta\rho$ 表示微小的频域。$\operatorname*{avg}_{\rho=e\to f}\{\ \}$表示在 e 列 f 的频域中取平均值。

4. 径向离散 x_4（用 ρ 加权）

$$x_4=\frac{1}{x_1^2}\operatorname*{avg}_{\rho=e\to f}\left\{\int_0^{\pi}d\theta\int_{\rho}^{\rho+\Delta\rho}\rho^2|F(\rho,\theta)|^2d\rho-\operatorname*{avg}_{\rho=e\to f}\left[\int_0^{\pi}d\theta\int_{\rho}^{\rho+\Delta\rho}\rho^2|F(\rho,\theta)|^2d\rho\right]\right\}^2。 \tag{4}$$

(4)式同(3)式类似，只是用 ρ 作了加权。

5. 角向离散 x_5

$$x_5=\frac{1}{x_1^2}\operatorname*{avg}_{\theta=0\to\pi}\left\{\int_{\theta}^{\theta+\Delta\theta}d\theta\int_g^{h}\rho|F(\rho,\theta)|^2d\rho-\operatorname*{avg}_{\theta=\ \to\pi}\left[\int_{\theta}^{\theta+\Delta\theta}d\theta\int_g^{h}\rho|F(\rho,\theta)|^2d\rho\right]\right\}^2, \tag{5}$$

其中 $\Delta\theta$ 表示微小的角域。g, h 为频域区间。

6. 角向最大和最小值之差 x_6

$$x_6=\frac{1}{x_1}\left\{\max_{\theta=0\to\pi}\left[\int_{\theta}^{\theta+\Delta\theta}d\theta\int_g^{h}\rho|F(\rho,\theta)|^2d\rho-\min_{\theta=0\to\pi}\left[\int_{\theta}^{\theta+\Delta\theta}d\theta\int_g^{h}\rho|F(\rho,\theta)|^2d\rho\right]\right\}, \tag{6}$$

这里 $\max_{\theta=0\to\pi}[\ ]$ 和 $\min_{\theta=0\to\pi}[\ ]$分别为在 g 到 h 的频域中，从 0 到 π 的极角范围内，微小角域上最大和最小的能量值。

由上述六个表达式，通过近似可得到由楔环探测器输出表示的六个参量，方便地编成程序，由微计算机完成参量的计算。经过统计计算和实际的判别检验，这六个特征参量的质量是较好的。

判别分析是细胞识别的最后一步。判别分析的关键是选择恰当的分辨函数。基于判别的目的是要区分两类细胞，采样和判别分析由同一微型机完成，两类谱的差异信息较大。我们选择了线性判别函数。线性判别函数是模式识别中一种常用的，比较可靠的判别函数。其

优点是：如果两类现象有明显的差异，那么只要特征参量选得正确，就能实现分辨。而且判别分析的计算量较小，易于实时处理[3,4]。

线性判别的中心思想是根据选定的特征参量[设为 $x_k(k=1, 2, \cdots, p)$]，作线性组合，构成一个线性函数，作为判别函数。

即：

$$y=\sum_{k=1}^{p} c_k x_k, \tag{7}$$

其中 p 为参量的个数，c_k 为待定系数。

由于两类现象 x_k 的数值不同，y 亦不同。这样就自然可找到一个介于两个 y 值之间的 y_0 作为判别指标，以此区分两类现象。

一般讲使用单个的指标进行判别比较困难，而使用综合指标比较容易。使用综合指标可以减少两类现象间的掺杂部分，使差异更加明显，分辨更为方便和准确。

建立综合指标具体就是根据两类现象的实测数据，确定待定系数 c_k。基本原则是使综合指标满足两类间的离散最大，同类间的离散最小的条件。该条件可用数学式子表示成：

$$I=\frac{(\bar{y}_A-\bar{y}_B)^2}{\sum_{i=1}^{n_1}(y_{Ai}-\bar{y}_A)^2+\sum_{i=1}^{n_2}(y_{Bi}-\bar{y}_B)^2}, \tag{8}$$

其中 A、B 分别代表正常细胞和癌细胞，$\bar{y}_A$、$\bar{y}_B$ 分别为 A 和 B 的判别指标，n_1 和 n_2 为两类细胞的样本数。

对(8)式中的 c_k 作偏微分运算，并令其偏导数为零，可得到以特征参量的相关因子 s_{kl} 为系数阵的 c_k 的 p 阶线性方程组。

$$\sum_{l=1}^{p} c_k s_{kl}=\beta d_k \quad (k=1, 2, \cdots, p), \tag{9}$$

其中：

$$\left.\begin{aligned} d_k&=\bar{x}_k(A)-\bar{x}_k(B),\\ \bar{x}_k(A)&=\frac{1}{n_1}\sum_{i=1}^{n_1} x_{ki}(A), \quad (k=1, 2, \cdots, p)\\ \bar{x}_k(B)&=\frac{1}{n_2}\sum_{i=1}^{n_2} x_{ki}(B), \end{aligned}\right\} \tag{10}$$

β 为独立于 k 的因子，可视具体情况取值。这里为计算方便取作 1。解方程组(9)，可求出 c_k。进而求出 $\bar{y}_A$ 和 $\bar{y}_B$ 最后可得加权平均的综合判别指标 y_0。

其中：

$$\left.\begin{aligned} \bar{y}_A&=\sum_{k=1}^{p} c_k\bar{x}_k(A),\\ \bar{y}_B&=\sum_{k=1}^{p} c_k\bar{x}_k(B),\\ y_C&=\frac{n_1\bar{y}_A+n_2\bar{y}_B}{n_1+n_2}。\end{aligned}\right\} \tag{11}$$

有了综合指标，便可使用(7)式判别了。

实验中，我们以 40 张细胞底片为样本，其中 20 张为正常细胞底片，20 张为癌细胞底片。在图 1 所示的系统上，用六十四路楔环探测器采样，并同时用微计算机算出六个参量，

建立了综合指标 y_0，做了参量的质量检验和显著性检验。在此基础上，对 100 张细胞底片作了筛选，正确识别率达 83%，其中假阴性率 8%，假阳性率 9%。

考虑可采取两个措施提高识别率。第一，实验中，样本为 40 个，比较少。扩大样本，重新建立综合指标，可得到更加准确的 c_k 和判别函数，从而提高识别率。第二，针对细胞检查要求假阴性率越低越好，可以对经计算得到的 y_0 作适当的调整，使假阴性率降低，使结果更加适合细胞检查的需要。

实验中使用的细胞样本是天津市肿瘤研究所提供的。在细胞底片的制备过程中，得到了金家瑞教授、王惠芳大夫的热情支持和帮助，在此表示感谢。

参 考 文 献

[1] R. E. Kopp; *J. Histochem. & Cytochem.*, 1974, **22**, No. 7 (Jul), 598.
[2] A. M. 罗斯著；《信息和通信理论》，(人民邮电出版社，1979)，221.
[3] 王宗皓，李麦村等编；《天气预报中的概率统计方法》，(科学出版社，1978)，102.
[4] C. R. Rao; «*Linear Statistical Inference and Its Application*», (John Wiley & Sons, 1965), 476.

Pattern recognition by means of power spectrum samples ——examples of screening cancer cells of human lung

ZHANG YIMO AND HUANG ZHI
(*Precision Instrumentation Department Tianjin University*)

(Received 13 August 1983; revised 2 November 1983)

Abstract

This article discusses the principle on which optical power spectrum is made use of identifying the same type of objects and gives an example of screening cancer cells of the lung. In that case, Fourier transform system with long focal length and high power lens and method of linear decision are proposed. A hybrid system of optics and computer which controls spectrum sample and completes statistical decision has been composed. The automated screening abnormal cells has been realized.