

光学-计算机混合方法识别尘肺

陈一苇 刘其真 何永保
(复旦大学计算机系)

吴中亚 于兆林
(安徽省职业病防治所) (中国科学院合肥智能机械研究所)

提 要

本文应用光学与计算机相结合的光学功率谱分析方法进行尘肺自动识别的研究。文中提出了一个新的特征抽取方法——逐次二维最优变换,并采用锥型分类函数作为判别函数,最后列出了三类样本的分类结果,实验表明,对于识别正常/病变两类问题的正确识别率达到94%。

关键词: 尘肺, 光学功率谱, 光学/计算机混合图像处理。

一、引 言

自动诊断尘肺的研究在国外已有近十年的历史, 基本研究方法可归纳为统计纹理分析度量法^[1,2]和不透明阴影直接识别法^[3]两大类, 前一类按所用设备不同又可分为全数字处理方法和光学/计算机混合处理方法; 阴影直接识别法也是一种主要依赖数字计算机处理图片的全数字方法。一般说来, 由于光学/计算机混合处理方法具有高速, 平行处理等特点, 故适用于快速筛选或粗分类场合, 而全数字方法适宜于病情分析等细分类场合。国外报道的对于后前位胸部 X 线胶片所进行的正常/尘肺两类问题的分类正确率, 混合方法的最好结果小于90%。

本文应用光学/计算机混合图像处理系统, 即光学功率谱分析系统, 进行尘肺自动识别的研究。提出了一个新的用于尘肺自动分类的特征抽取方法——逐次二维最优变换。实验表明, 对于识别正常/病变两类问题, 正确率达94%。

二、光学/计算机图像处理系统

OPSA-I 型系统*是一个光学/计算机混合图像信息处理系统^[4], 它由光学系统和计算机系统两部分组成, 前一部分是相干光傅里叶变换装置, 它能以光速从图片平行产生傅里叶光学功率谱, 由楔环光电探测器^[5]转换成32个楔能量和32个环能量, 经模拟/数字转换后, 送入数字计算机分析处理。

收稿日期: 1985年9月24日; 收到修改稿日期: 1986年3月5日

* 复旦大学研制。

三、特征抽取方法

在本研究中,我们提出一个新的特征抽取方法——逐次二维最优变换,它是基于一个判断两类样本聚类可分程度的准则

$$R(d_1, d_2) = \frac{(d_1^T \Delta)^2}{d_1^T A d_1} + \frac{(d_2^T \Delta)^2}{d_2^T A d_2}, \quad (1)$$

其中 d_1 和 d_2 是两个标准正交向量,它们是样本数据所在的空间中待求的两个投影方向; Δ 为两类样本数据的均值差,它反映类间的自然重心间距; A 为各类的协方差估计矩阵之和,是反映类内聚合程度的量。准则 R 综合 Δ 和 A 投影到由 d_1 和 d_2 所张成的二维特征空间中分布情况,二维最优特征空间便由使准则 $R(d_1, d_2)$ 达到最大的 d_1, d_2 一次张成。如果在已求得特征空间中不断重复这一过程,即可产生一个逐次二维最优的特征空间。我们得到第 $k+1$ 组解 d_{2k+1} 和 d_{2k+2} 可由前 k 组解 $d_1, d_2, \dots, d_{2k-1}, d_{2k}$ ($2k \ll n, n$ 为空间维数);递归产生这样一个解析结果,公式为

$$d_{2k+1} = \frac{(B_k A B_k + \alpha_k I)^{-1} (B_k \Delta - \theta_k x_k)}{\|(B_k A B_k + \alpha_k I)^{-1} (B_k \Delta - \theta_k x_k)\|},$$

$$d_{2k+2} = \frac{(B_k A B_k + \alpha_k I)^{-1} (B_k \Delta + \theta_k x_k)}{\|(B_k A B_k + \alpha_k I)^{-1} (B_k \Delta + \theta_k x_k)\|},$$

其中 $B_k = I - \sum_{i=1}^{2k} \alpha_i \alpha_i^T$, $B_0 = I$; α_k 是 $B_k A B_k$ 的最小非零特征向量; x_k 是 α_k 所对应的 $B_k A B_k$ 的单位特征向量; $\theta_k = 2\alpha_k \sqrt{\Delta^T (B_k A B_k + \alpha_k I)^{-2} \Delta}$ 。由于 A 对称正定, B_k 对称,故 $B_k A B_k$ 对称非负定, α_k 为正值,因而 $(B_k A B_k + \alpha_k I)$ 仍为对称正定,其逆存在。

最后的分类是采用锥型分类器,在二维情形下,它是一种分段线性分类器^[6]。

四、尘肺病自动分类实验结果

供实验用的原始样本是一组 $35 \text{ mm} \times 24 \text{ mm}$ 的后前位胸部 X 线胶片,它们分为三类:其中被医学专家诊断为正常(0 期如 0-I 期)范围的有 20 张,系初期(I 期)尘肺的有 21 张,属晚期(II 期和 III 期)尘肺的有 19 张。实验中分别将每类样本随机划分成一组设计样本集和一组测试样本集。设计集用来设计有关特征抽取和分类所需的参数,测试样本集用来测试这些参数确定的特征空间的合理性和分类精度等。表 1 给出了具体样本集名及其所含样本类型和个数。

鉴于传统的人工诊断方式是将后前位 X 线胸片所显现的肺野划分为右上、右中、右下和左上、左中、左下这六个肺区来衡量病情在整个肺部的大致分布范围和严重程度,我们也分别对这 6 个肺区进行光学功率谱采样;针对肺野的特殊形状,采样时在胶片前置放一副椭圆形光阑,以求避开噪声区域,将采样范围限于信息量较大的部分。图 1 给出了各类样本集的典型后前位 X 线胸片,典型纹理图形及对应于胸片右中区域的傅里叶光学功率谱。现在,每个样本胶片就产生了六组光学功率谱经采样量化后的数据,为了提高处理速度,我们仅采用环分量作为特性数据,然后再将各肺区的特性数据作数学平均,作为用复合多孔光阑对整张

Table 1 The make-up of sample sets

sample categories	sample sets	training sets		test sets	
		name	number	name	number
normal	0	D_1	9	T_1	2
	0-I		6		3
simple disease	I	D_2	14	T_2	7
severe disease	II & III	D_3	14	T_3	5

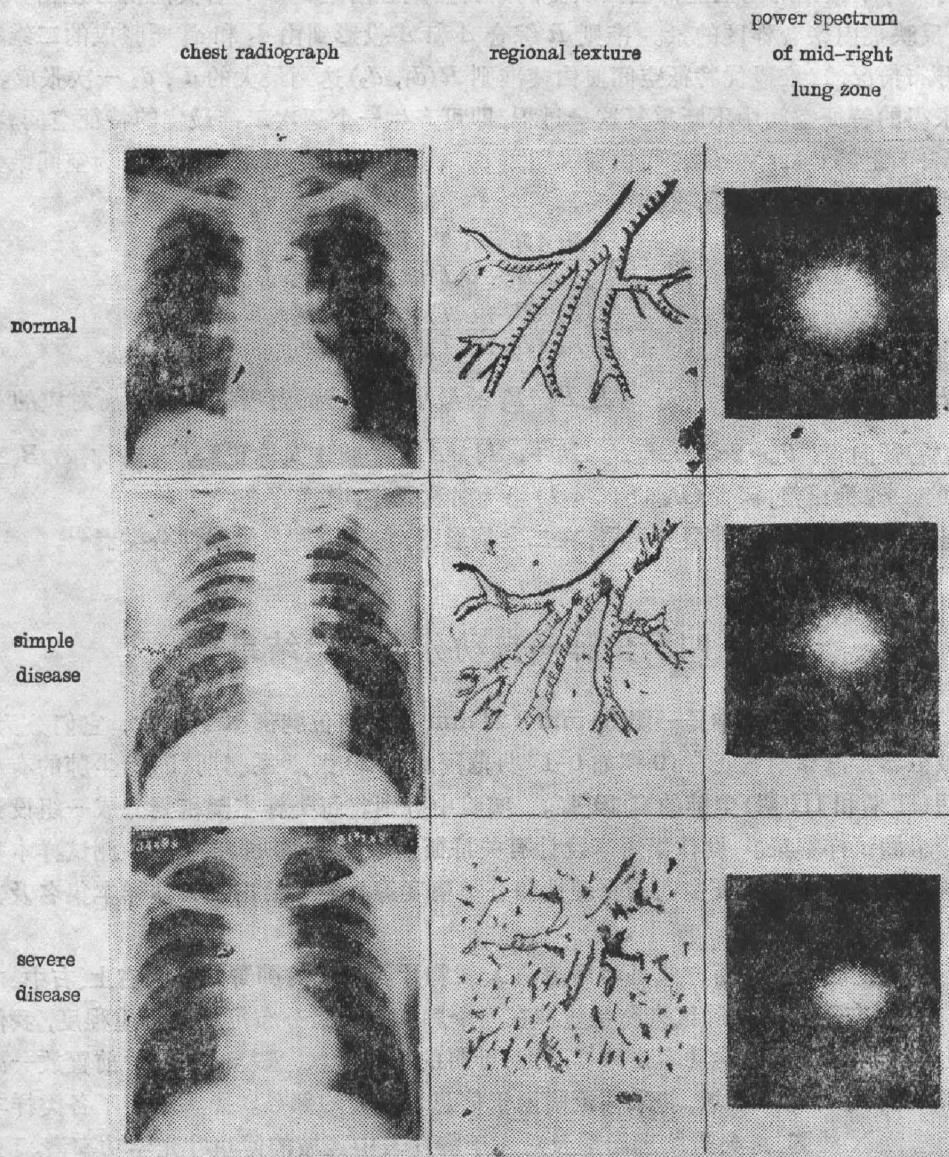


Fig. 1 Typical X-ray films, texture features and optical power spectra of several types of mid-right lung zone

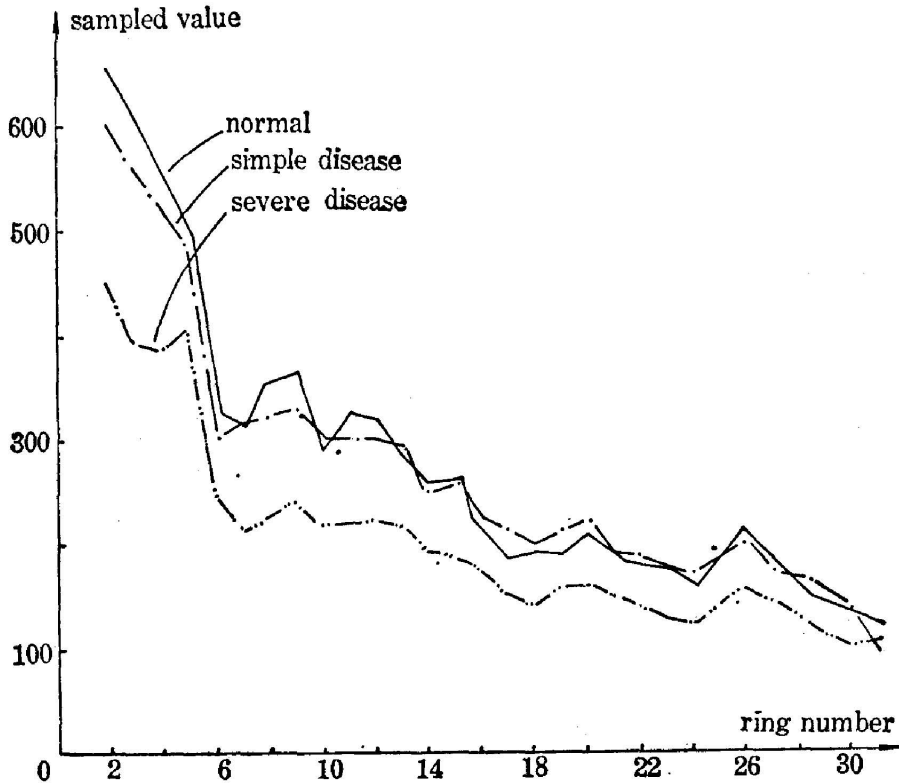


Fig. 2

胸片一次性采样的近似,将数据依低频至高频排列为 r_1, r_2, \dots, r_{32} 。由于 r_1 和 r_2 是功率谱中心周围高能积分,失真较大,故未予采用,剩下的分量组成供进一步分析处理用的样本数据向量对于维数 $n=30$ 的数据空间中的向量 \mathbf{d} 为

$$\mathbf{d} = (r_3, r_4, \dots, r_{32})^T。$$

实际上,作为样本数据向量集合的设计集 D_i 或测试集 T_i ($i=1, 2, 3$) 都是由它们各自所含的样本经过如上处理后的样本数据向量 \mathbf{d} 组成。图 2 显示了三个设计集的各自样本数据向量平均,也即平均环分量的分布情况。

整个数据分析和处理过程是在计算机上完成的,其基本过程如下:

(1) 计算每个样本的六个肺区的环分量数据的平均值,然后进行数据归一化处理。

(2) 对于正常类设计集 D_1 和病变类设计集 $D_2 \cup D_3$ 作二维最优变换,从而设计出供区分正常/病变所需要的二维特征空间;其次再由 D_2 和 D_3 作变换设计出供区分初期尘肺/晚期尘肺的二维特征空间,于是我们已将 30 维样本数据空间压缩成二维。

(3) 在各特征空间中设计出区分两类问题的二维锥型分类器。

(4) 在各特征空间中用相应的分类器对设计集本身进行分类测试,以考察各类在低维特征空间内的聚类特征和算法的相容性。最后用测试集就特征空间的合理性和分类器的有效性进行测试。

表 2 以正误矩阵形式给出了分类结果。结果表明,本文提出的逐次二维最优变换,应用于通过对胸部 X 光片的光学功率谱采样得到的高维数据空间的压缩,可以获得使各类样本

Table 2 Classification results

a		true class			b		true class		
		D_1	D_2	D_3			T_1	T_2	T_3
assigned class	D_1	15	0	1	assigned class	T_1	5	0	1
	D_2	0	14	0		T_2	0	6	0
	D_3	0	0	13		T_3	0	1	4

(a) Computer classification of the three training sets

(b) Computer classification of the three test sets

在锥面可分的意义下具有良好的聚类特性的低维特征空间;在这样的两个空间中,所有设计集包含的四十三个样本仅有一个错分,也说明了光学功率谱能有效地反映尘肺病的肺部纹理变化信息。用测试集进行测试,能以 94% 的正确识别率识别正常/病变两类问题,以高于 90% 的正确识别率识别初期/晚期尘肺的细分问题,从而证明了特征空间的合理性,分类器的有效性和光学/计算机混合图像处理方法自动诊断尘肺病的可行性。比较而言,光学功率谱方法对于正常/病变两类问题的粗分类更为有效,说明光学功率谱对于这两类的纹理变化更为敏感,也符合实用的要求。注意到无论设计集还是测试集中都有正常样本被错分成严重尘肺的情况,这和国外所报道的实验结果是相一致的。因为仅就纹理上来说,有的晚期尘肺病例因有融合块出现,胸片上的相应部位被大块灰度相近的阴影遮盖,所以在功率谱的强度变化上与某些正常胸片的发生接近的现象,导致相互错分。但有些指标,例如三类问题分类的综合平均误差比国外做的要小得多。

五、结 束 语

以上介绍了用光学/计算机混合处理方法进行尘肺病自动分类研究的实验情况。实验表明,将图像的功率谱幅度信息,经过数字化,在逐次二维最优变换这一特征抽取方法和维型分类方法的支持下,能够较为有效地选取鉴别正常、初期尘肺和晚期尘肺这三类 X 线胸片的特征,从而将高维空间进行了大规模的压缩,在低维空间中对三类样本实现快速自动分类,获得了较高的正确识别率,为应用光学/计算机混合处理方法自动识别尘肺病研究走向实用提供了依据。

本实验所用的样本胶片的原片由安徽省职业病防治所提供,该所专家赵衍同志将原片作了详细诊断并分类,复旦大学计算机科学系杨伟敏同志协助对原片的翻拍,并在实验过程中提供了其它许多帮助,在此向他们致谢。

参 考 文 献

- [1] P. R. Kruger *et al.*; *Appl. Opt.*, 1977, 16, No. 10 (Oct), 2637~2646.
- [2] H. Stark, D. Lee; *IEEE Trans. Syst. & Man Cybern.*, 1976, SMC-6, No. 11 (Nov), 788~793.
- [3] C. C. Li *et al.*; *Proc. Conf. on Computer Appl. in Rad. & Computer Anal. Radiol. Images*, 1983, 424~427.
- [4] 何永保等;《自动化学报》,1984, 10, No. 1 (Jan), 80~84.

- [5] 赵焕卿等;《光学学报》, 1982, 2, No. 5 (Sep), 458~462.
[6] K. Fukunaga;《Introduction to Statistical Pattern Recognition》, (Academic Press, New York, 1972), Chapter 7, Chapter 8.

Hybrid optical-digital diagnosis of pneumoconiosis

CHEN YIWEI, LIU QIZHEN, HE YONGBAO
(Computer Science Department, Fudan University)

WU ZHONGYA
(Center of Prevention and Cure of Occupational Disease of Anhui Province)

YU ZHAOLIN
(Research Institute of Intelligent Machinery, Hefei)

(Received 24 September 1985; revised 5 March 1986)

Abstract

This paper presents the results of a joint study of automatic diagnosis of chest radiographs of pneumoconiosis. The Hybrid Optical/Digital Image Processing System of Fudan University (OPSA-1) was used. New algorithms developed for feature selection and classification led to a higher efficiency through experiments, with 94% accuracy in discriminating normal/abnormal radiographs.

Key Words: Pneumoconiosis, Optical power spectrum, Hybrid optical-digital image processing.