

中图法分类号: TP391.41 文献标志码: A 文章编号: 100874-8961(2011)10-1791-11

论文索引信息: 马莉, 秦波. 基于小波分析的皮肤肿瘤轮廓结构不规则特征分类器设计 [J]. 中国图象图形学报, 2011, 16(10): 1791-1801

## 基于小波分析的皮肤肿瘤轮廓结构 不规则特征分类器设计

马莉, 秦波

(杭州电子科技大学生命信息与仪器工程学院, 杭州 310018)

**摘要:** 皮肤肿瘤轮廓的结构不规则性对黑色素瘤临床早期诊断具有重要意义。针对皮肤肿瘤轮廓结构不规则性度量和特征分类器设计问题, 提出了一种基于小波子带分析的轮廓结构分量获取及多尺度特征神经网络分类器构建方法。利用肿瘤轮廓小波子带能量的 Hausdorff 距离寻找显著性小波子带, 进而重构肿瘤轮廓的结构分量; 给出基于显著性小波子带的轮廓多尺度结构不规则特征描述——基于统计和几何的轮廓结构不规则性度量。对单尺度/多尺度、小样本/大样本, 以及特征选择前后神经网络分类器性能进行了实验。结果表明, 在小样本情况下, 基于显著性小波子带多尺度特征描述扩展了样本的特征空间维数, 使得特征选择后分类器的灵敏度和特异度指标分别等于和优于大样本分类器。

**关键词:** 黑色素瘤检测; 轮廓不规则性; 多尺度描述子; 显著性小波子带

### Design of classifiers based on wavelet analysis using features of structural irregularity for skin lesion's contours

Ma Li, Qin Bo

(College of Life Information Science & Instrument Engineering, Hangzhou Dianzi University, Hangzhou 310018 China)

**Abstract:** Boundary Structural irregularity of skin lesions has shown the most significance in the early diagnosis of melanomas in clinic. Aiming to measure structural irregularities of skin lesion contours and designing feature oriented classifiers, a method for extracting structural components of contours based on wavelet sub-band analysis and neural network classifiers using multi-scale features is proposed in this paper. A set of salient wavelet sub-bands are searched using Hausdorff distance evaluations on energy at wavelet subbands. Then structural components of skin lesion contours are reconstructed. Furthermore, multi-scale descriptions of the contour's structural irregularity are given using statistical and geometrical measures based on the salient wavelet sub-bands. Finally, various experimental schemes are taken on single scale/multi-scale features, for small training sets and large training sets as well as and before and after feature selection for a comparison on performances of the neural network classifiers. It is shown from experiments that the classifier's sensitivity and specialty after feature selection using small sample sets equal and exceed that of large training sets respectively as dimensions of feature spaces are expanded by the proposed multi-scale independent features at salient wavelet sub-bands.

**Keywords:** melanoma detection; structure irregularity; multi-scale descriptors; salient wavelet sub-bands

收稿日期: 2010-09-26; 修回日期: 2010-11-25

基金项目: 国家自然科学基金项目(60775016); 浙江省重大科技专项基金项目(2007C13062)。

第一作者简介: 马莉(1954—), 女, 教授, 硕士生导师。1998年于中南大学获控制理论与控制工程专业博士学位, 主要研究方向为图像分析、模式识别。E-mail: mali@hdu.edu.cn。

# 0 引言

恶性黑色素瘤是一种恶性肿瘤,由表皮或黏膜细胞组成,具有易转移、恶性程度高,预后不良等特点。目前,临床判断主要依据 ABCD 法则进行诊断<sup>[1]</sup>,即瘤体的不对称性(asymmetry)、轮廓的不规则性(boundary irregularity)、瘤体颜色(color)和瘤体直径(Diameter)。

过去十几年来,对黑色素瘤轮廓不规则性的研究主要有几何法<sup>[2]</sup>和分形维法<sup>[3]</sup>。皮肤肿瘤轮廓由纹理部分和结构部分构成,而肿瘤轮廓的不规则性可分解为纹理的不规则和结构的不规则性。前者指沿瘤体轮廓细微的变化,后者表示轮廓上大的凹凸起伏<sup>[2]</sup>。纹理不规则性与成像系统的噪声有关,而结构不规则性由表征皮肤肿瘤的异性增生引起,在恶性黑色素瘤诊断过程中具有重要的临床诊断意义。图 1 给出了良恶性肿瘤的轮廓差异。良性肿瘤轮廓平滑(图 1(a)),轮廓具有微小的变化;而恶性肿瘤轮廓呈现大的波动,具有明显凹凸起伏(图 1(b))。在临床诊断中,结构不规则性的度量具有重要诊断意义。Lee 提出用高斯滤波方法不断平滑肿瘤轮廓,通过建立的多尺度轮廓曲率模型确定合适的尺度级进行结构不规则性度量<sup>[2]</sup>,但计算十分复杂。

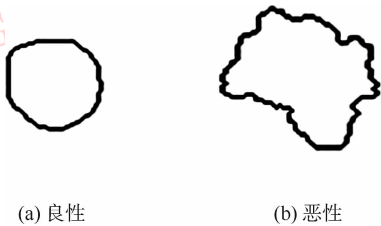


图 1 皮肤肿瘤轮廓样本  
Fig. 1 Contour samples of skin lesions

小波变换是一种强大的多分辨率分析工具<sup>[4]</sup>。根据信号频谱分析,肿瘤轮廓的纹理和结构不规则性分别表现为高频和低频段。由于临床中已获病理检测确认的恶性黑色素瘤样本较少,基于计算机辅助诊断的小样本分类器研究成为所关注的热点问题之一。小波多尺度分析可以扩展信号在多尺度空间的特征表述,进而为构建具有高性能的小样本肿瘤分类器提供了可行性。本文的研究目标是探讨基于小波多尺度分析方法的皮肤肿瘤轮廓结构不规则性检测,以及基于轮廓结构不规则特征的恶性肿瘤分类器设

计。本文的基本思想是:根据皮肤肿瘤轮廓的结构和纹理分量在傅里叶频谱中的可分离性和小波分析的多分辨率特点,借助小波多分辨率子带分析寻找实现分离皮肤肿瘤轮廓纹理和结构分量的合适尺度,接着利用小波子带肿瘤轮廓中隐含的重要信息作为分类器的特征向量,从而提高分类的准确性。

## 1 皮肤肿瘤轮廓结构分量提取

### 1.1 肿瘤轮廓小波分解

原始皮肤肿瘤图像轮廓可描述为  $L = \{x_1, y_1, x_2, y_2, \dots, x_N, y_N\}$ 。将 2 维信号转换为 1 维信号有两种方法:1) 将轮廓  $L$  分解成两个 1 维子向量  $L_x = \{x_1, x_2, \dots, x_N\}$  和  $L_y = \{y_1, y_2, \dots, y_N\}$ ; 2) 将轮廓  $L$  转换为  $L_r = \{r_1, r_2, \dots, r_n\}$  其中  $r_i = \sqrt{(x_i - x')^2 + (y_i - y')^2}, i = 1, 2, \dots, n, (x', y')$  描述了轮廓直角坐标系下闭合曲线  $L$  的质心。本文采用第 2 种方法。

#### 1.1.1 尺度归一化

通常情况下,恶性肿瘤的轮廓长度要比良性肿瘤轮廓长度长得多。若良、恶性肿瘤轮廓点构成的数据集尺寸不同,会导致其频域分辨率不一致,进而影响小波子带的分辨率。肿瘤轮廓尺度归一化可以保证良、恶性肿瘤在各频率段具有可比性。轮廓归一化公式为

$$r'_i = r\left(i \frac{n}{N}\right) + \mu_1 - \mu_2 \quad i = 1, 2, \dots, N \quad (1)$$

式中,  $n$  和  $N$  分别描述了归一化前后的轮廓尺度,  $\mu_1$  和  $\mu_2$  表述数据集中良恶性肿瘤的径向距离均值。

#### 1.1.2 小波分解

小波多分辨率分析通过一对高通 H 和低通 L 滤波器,将原信号分解为小波域下的近似分量和细节分量,其中近似分量作为进一步分解的输入信号,依次生成一棵分解树。1 维小波分解结构如图 2 所示<sup>[5]</sup>。

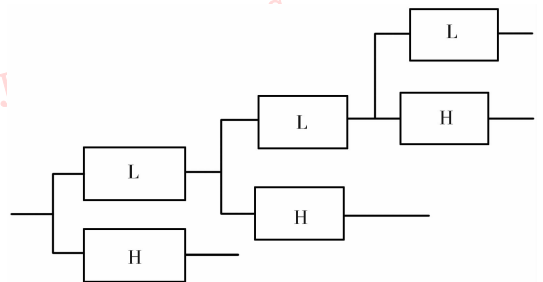


图 2 树状小波分解结构  
Fig. 2 Tree structure of wavelet decomposition

对轮廓  $L_r$  进行 1 维平稳小波分解,分解后第  $i$  层的近似系数和细节系数分别记为  $A^i = \{a_1^i, a_2^i, \dots, a_N^i\}$  和  $D^i = \{d_1^i, d_2^i, \dots, d_N^i\}$ 。随着尺度的增加,近似系数进一步被分解,在频域中将低频进一步分解为低频和高频。一般来说,轮廓的纹理部分位于高频带,能量较低,而结构部分位于低频带,能量较高。肿瘤轮廓  $L_r$  通过小波分解至  $s$  层,被分解为子带信号  $A^s, D^s, D^{s-1}, \dots, D^1$ , 其对应子带信号的频带分别为  $[0, 1/2^s f_{max}]$ ,  $[1/2^i f_{max}, 1/2^{i-1} f_{max}]$ ,  $i = s, s-1, \dots, 2, 1$ , 描述了频带之间的级联关系,其中  $f_{max}$  为采样频率的一半。本文研究描述肿瘤轮廓结构分量的小波子带位置以及确定具有较强甄别良恶性肿瘤能力的频带范围<sup>[6-7]</sup>。

### 1.2 小波子带选择

任意 1 维信号 FFT 变换的幅频特性描述了其在

频域中的幅值分布。频谱与轮廓点位置的无关性是 FFT 变换的不足之处,并且变换时窗无穷大;而小波变换的窗口函数是可变的。图 3 给出了 FFT 和小波分析的对比,其中纵坐标为幅值,横坐标为离散频率。图 3(a)描述了良性肿瘤轮廓的变换,其中第 1 幅图给出了良性肿瘤轮廓 FFT 变换后的频谱图(中间为零频率),难以观察中频和低频部分的变化;第 2 幅图描述了经过 2 级小波变换,在频域中表现的频带子带分割图; $A^1, A^2, D^1, D^2$  分别描述了信号的近似系数和细节系数,它们各自的子带系数分别为图 3 前两列的  $A^1, A^2, D^1, D^2$ , 可以看出,良性肿瘤的高频  $D^1$  和次高频  $D^2$  有微小的波动。图 3(b)描述了恶性肿瘤轮廓的情况。从图 3 中可以发现,恶性肿瘤在  $D^1$  和  $D^2$  频段的小波细节系数的幅度比良性高,与良性轮廓的  $D^1$  和  $D^2$  有明显的差异。

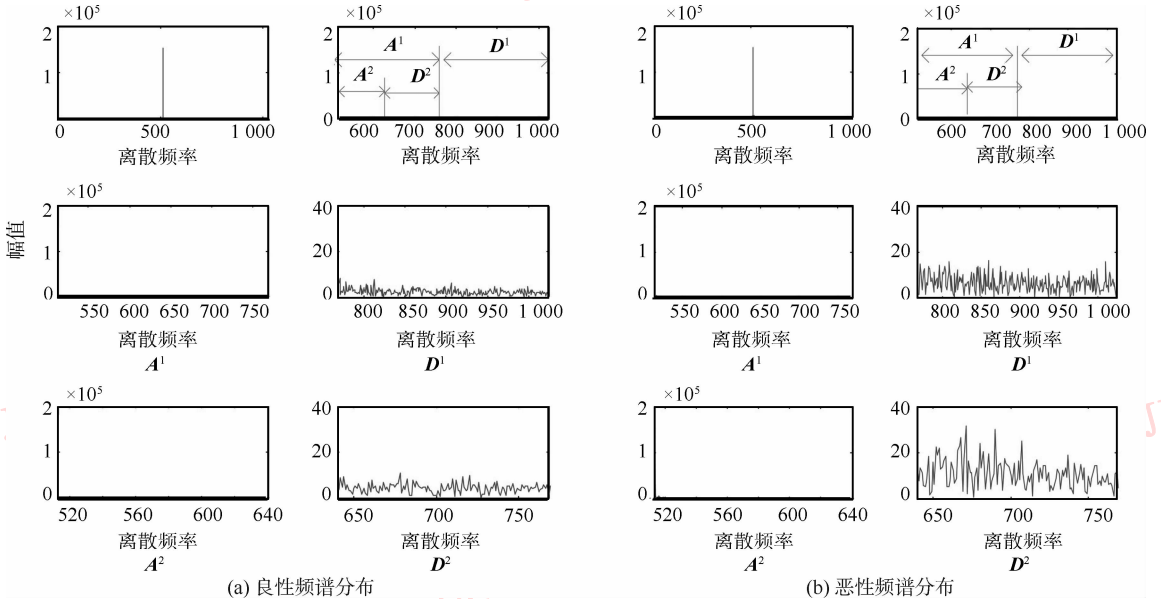


图 3 良恶性轮廓频谱比较

Fig. 3 Comparison of frequency spectrum for a benign and a malignant contours

#### 1.2.1 具有高类间区分度的显著性子带

根据小波重构原理,肿瘤轮廓通过小波分解后近似系数和若干层细节系数重构可得到轮廓的结构分量。在轮廓信号分解过程中,各层细节总能的大小代表了该小波子带相对原轮廓信号的重要程度。基于小波分解各层细节系数的能量值,通过比较良、恶性样本数据集之间的 HD(hausdorff distance)<sup>[8]</sup> 值分布来选择重要的小波子带,即类间区分度显著性子带。小波子带  $D^j$  的小波能为<sup>[4]</sup>

$$E_j = \sum_i (d_i^j)^2 \quad j = 1, 2, \dots, n \quad (2)$$

#### 1.2.2 类间区分度显著性子带选择

- 1) 选择若干具有分类标识的良、恶性黑色素瘤样本,并分别均分为 3 组。设置最大分解层数  $n$ 。
- 2) 对各组样本分别进行小波分解,得到各层小波子带细节信号  $D^n, D^{n-1}, \dots, D^1$ , 并应用式(2)计算小波能。
- 3) 分别计算良、恶性肿瘤各组的小波子带平均能量。
- 4) 计算良恶性肿瘤在各小波子带平均能量的类间 HD。
- 5) 选择具有较大 HD 值的小波子带。

HD 值表征相应小波子带类间区分能力。由于良、恶性肿瘤轮廓的主要差异在于结构分量,则基于 HD 选择的显著性子带构成肿瘤轮廓结构分量的细节部分。

### 1.2.3 肿瘤轮廓结构分量获取

由小波重构理论可知,当轮廓分解至  $s$  层,则原轮廓可以用  $A^s, D^s, D^{s-1}, \dots, D^1$  重构获得。为了得到重构后的肿瘤轮廓结构分量,需要确定分解层数  $s$  和分界点  $s_i$  ( $s > s_i$ )。即位于  $[s, s_i]$  的低频子带描述了肿瘤轮廓结构分量,而位于  $[s_{i-1}, 1]$  的高频子带表征肿瘤轮廓的纹理分量。肿瘤轮廓的结构分量  $C_s$  和纹理分量  $C_i$  表示为

$$C_s = A^s + D^s + D^{s-1} + \dots + D^{s_i} \quad (3)$$

$$C_i = D^{s_{i-1}} + D^{s_{i-2}} + \dots + D^1 \quad (4)$$

然而选择合适的  $s$  和  $s_i$  非常困难。 $s$  过大,小波子带信号  $D^s$  过于微弱而不足以表示信号的低频的变化,而且计算过程的计算量大大增加;若  $s$  过小,小波子带信号  $D^s$  将混合有结构和纹理的信息,难以准确实现肿瘤轮廓结构分量的提取。同样,  $s_i$  过大或过小将分别导致轮廓结构分量缺失或结构分量中混有纹理不规则的成分。1.2.2 节描述的小波子带选择确定了分解层数  $s$  和分界点  $s_i$ 。该方法在小波子带能量度量基础上,利用肿瘤轮廓样本数据的 HD 距离选择具有高类间区分度的显著性子带。其中  $s$  和  $s_i$  分别为显著性子带所覆盖的小波子带尺度最大和最小值。

## 2 多尺度不规则特征描述子

目标轮廓的不规则度量对提取轮廓结构部分后的分类问题至关重要。采用两种轮廓不规则特征的描述子:1) 基于统计量的特征描述子;2) 基于几何学的特征描述子。结构不规则特征描述子仅在显著性子带层  $j$  中提取,  $j \in \Omega$ 。 $\Omega$  为在 1.2.2 节中得到的小波子带域。

### 2.1 统计学特征描述子

除了将显著性子带能量均值  $D^j$  作为特征描述子外<sup>[6]</sup>,以下的各描述子也作为肿瘤轮廓不规则性描述子:

#### 1) 小波能量熵<sup>[9]</sup>

$$w_j = - \sum_{i=1}^N p_i^j \log(p_i^j) \quad (5)$$

式中,  $p_i^j$  描述了小波系数为  $D^j$  的第  $i$  个值的能量概率,即

$$p_i^j = \frac{E_i^j}{E^j} \quad E_i^j = |d_i^j|^2$$

小波能量熵描述了信号的紊乱程度。

#### 2) 极限宽

对任意信号  $X = \{x_1, x_2, \dots, x_N\}$ , 极限宽定义为

$$w = \frac{2 \times \sigma}{\mu} \quad (6)$$

式中,  $\mu$  和  $\sigma$  分别表示信号  $X$  的均值和标准差。极限宽值越大表示信号越不规则,反之越规则。

## 2.2 基于几何度量的不规则描述子

在显著性子带第  $j$  层,轮廓结构部分  $C_j$  可通过小波系数重构得到:  $C_j = A^s + D^s + D^{s-1} + \dots + D^j$ ,  $s \geq j \geq s_i$ 。除方差<sup>[5]</sup>特征之外,不规则描述子为

#### 1) 径向偏差度<sup>[10]</sup>

$$RA = \frac{1}{N} \sum_{i=1}^N |r_i - \bar{r}| \quad (7)$$

式中,  $\bar{r}$  表示轮廓径向距离均值。

#### 2) 轮廓粗糙度<sup>[10]</sup>

$$R = \frac{1}{N} \sum_{i=1}^N |r_i - r_{i+1}| \quad (8)$$

#### 3) 轮廓不规则描述子

$$IM = \frac{A(C_j \oplus S_m)}{A(S_m)} \quad (9)$$

式中,  $S_m$  表示近似系数  $A^m$  在重要子带  $m = s$  层重构得到的轮廓所围成的区域面积,  $\oplus$  表示异或操作,  $A(\quad)$  为求面积函数。

## 3 特征选择

小波多分辨率分析提供了多尺度肿瘤轮廓结构不规则特征描述,对所构成的原始特征向量进行选择。即去除冗余,保留具有较好区分能力的重要特征,改善分类准确性和提高分类效率。特征选择分为两步:1) 首先进行相关性分析,根据特征之间的相关性去除相关性较强的特征;2) 对去相关性后得到的特征,统计其特征概率密度分布,选择那些具有较小分类误差的特征,去除分类效果较差的特征。

### 3.1 相关性分析

特征相关性分析是一种常见的特征选择方法。经过相关性分析以去除相关性较强的冗余特征<sup>[11]</sup>。对于特征集  $T = \{t_1, t_2, \dots, t_n\}$  和样本集  $S = \{x_1, x_2, \dots, x_m\}$ , 任意两个特征  $t_i, t_j$  的相关性系数为

$$\rho_{ij} = \frac{\sum_{p=1}^m (t_{pi} - z_i)(t_{pj} - z_j)}{\sqrt{\sum_{p=1}^m (t_{pi} - z_i)^2 \sum_{p=1}^m (t_{pj} - z_j)^2}} \quad (10)$$

式中,  $z_i$  描述了样本集合中特征为  $t_i$  的均值,  $t_{pi}$  描述了样本  $x_p$  对应特征  $t_i$  的取值。相关系数  $\rho_{ij}$  的绝对值大小描述了特征  $t_i$  和  $t_j$  之间相关程度的高低,  $\rho_{ij}$  绝对值越大, 表示相关程度越高, 反之越低。

在实验中, 通过计算任意两个特征的相关性, 可以得到相关性矩阵<sup>[7]</sup>

$$\mathbf{Z}_{ij} = \begin{bmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1N} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{N1} & \rho_{N2} & \cdots & \rho_{NN} \end{bmatrix} \quad (11)$$

根据特征之间的相关性去除冗余的算法步骤如下:

- 1) 计算相关性矩阵。
- 2) 设定相关性阈值  $T$ , 得到相关系数值大于阈值  $T$  的特征向量对  $(t_i, t_j)$  ( $i \neq j, 1 \leq i, j \leq n$ ), 并按相关系数值大小降序排列。
- 3) 用逐一法去除特征向量对  $(t_i, t_j)$  中的一个特征, 即对于已删除了一个特征的特征对, 无条件保留另一个特征。

### 3.2 概率密度分析

对于去除了强相关后的特征集合, 需要进一步考察这些特征的分类能力。根据从已知类别的样本图像中所提取的特征量, 度量各特征对分类贡献的大小<sup>[12]</sup>。即给定已知类别的肿瘤样本集: 良性肿瘤  $w_1$  和恶性肿瘤  $w_2$ , 统计任一特征的概率分布函数  $P(t_i | w_2)$ ,  $P(t_i | w_1)$ 。假设特征对分类的先验概率  $p(w_1)$  和  $p(w_2)$  相等, 则后验概率  $p(w_1 | t_i)$  和  $p(w_2 | t_i)$  接近于  $p(t_i | w_1)$  和  $p(t_i | w_2)$ 。类条件概率密度为

$$p(w_i | t) = \frac{p(t | w_i)p(w_i)}{\sum_{i=1}^2 p(t | w_i)p(w_i)} \quad (12)$$

图 4 描述了不同特征概率分布对良恶性肿瘤分类的影响。若特征  $t$  使概率分布函数  $p(t | w_2) = 0$  处  $p(t | w_1) \neq 0$ , 如图 4(a), 则两类为完全可分的, 对应分类误差为零; 相反, 若特征  $p(t | w_2) = p(t | w_1)$ , 如图 4(b), 则两类完全不可分<sup>[12]</sup>。

通常情况下, 两类样本的特征  $t_i$  分布具有重叠区域。重叠区域与分类误差紧密相关, 重叠区域越大表示特征  $t_i$  的分类能力越差。针对 3.1 节中去除相关性后的特征集合, 通过考察良恶性样本特征的概率分布函数曲线的交叠程度, 选择具有良好类间

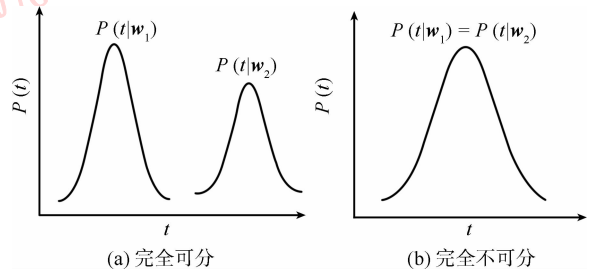


图 4 两种概率分布情况

Fig. 4 Two types of probability distributions

可分性的特征<sup>[13]</sup>。具体选择过程如下:

- 1) 分别计算良恶性肿瘤每个特征的概率分布密度曲线;
- 2) 考察任一特征  $t_i$  的概率分布  $P(t_i | w_1)$  和  $P(t_i | w_2)$  之间的交叠程度。若存在阈值  $T$  使得累加概率良性  $p_b = \sum_{i=1}^k p_i^b \geq T$ , 恶性  $p_m = \sum_{i=k+1}^N p_i^m \geq T$ , 则保留特征  $t_i$  且其错分率为  $1 - T$ , 否则剔除特征  $t_i$ 。 $T$  为准确区分良、恶性的设定阈值,  $p_i^b$  和  $p_i^m$  分别为良性和恶性概率分布  $P(t_i | w_1)$  和  $P(t_i | w_2)$  在点  $i$  的值。

## 4 实验结果

实验用皮肤肿瘤样本图像共 134 幅, 其中 62 幅良性痣, 72 幅恶性黑色素瘤。实验样本均来自网络数据库 <http://www.dermnet.com/> 和当地医院。图像分辨率为  $120 \times 120$  像素, 每幅图像在预处理阶段均转换为灰度图像。在分类器设计实验中, 任意选择 31 幅良性痣和 36 幅恶性黑色素瘤图片作为训练集, 其余作为测试集。实验环境为 Matlab7.6。

实验包含 3 个部分: 基于显著性子带选择的轮廓结构分量重构, 特征选择和分类器设计。在分类器设计中, 进一步检验了多尺度特征的小样本分类能力。

### 4.1 显著性子带选择和轮廓结构分量重构

显著性子带选择。将训练集样本的良、恶性肿瘤图片集分别均分为 3 组。对每幅图像提取肿瘤轮廓, 经尺度归一化后转换为 1 维信号。肿瘤轮廓小波分解过程中选择“db3”小波, 分解层次设置为 13 层。接着, 计算小波分解后肿瘤轮廓在第  $j$  个小波子带  $D^j$ ,  $j = 1, 2, \dots, 13$  的能量, 并统计图像样本在各组的小波子带能量均值, 进而得到类间 HD 值<sup>[6]</sup>分布(如图 5 所示)。从图中可以看出, HD 分布表征基于小波子带能量的类间分类能力, 其在两端较低, 中间较高。

小波子带  $D^1$  至  $D^5$  位于肿瘤轮廓的高频部分,这部分为轮廓的纹理部分,由于恶性肿瘤均含有纹理的不规则性,所以高频部分的分类能力很低。小波分解层在  $j=6$  之后具有较高的 HD 值,对应中频段(恶性肿瘤的结构分量部分)。这是由于良恶性肿瘤在轮廓的结构分量上存在较大差异造成的。小波子带从  $j \geq 10$  之后,HD 逐渐减小。我们知道,小波子带  $D^{10}$  至  $D^{13}$  位于低频部分,在这些低频频段良、恶性肿瘤均趋于平滑,类间差异减小。因此,中频带即为包含有重要分类信息的频带,称为显著性子带。

图 6 描述了当  $j=9$  的近似系数  $A^9$  分别与  $j=6 \sim 9$ (见图 6(c))、 $j=7 \sim 10$ (见图 6(d)) 细节系数

重构得到肿瘤轮廓结构部分,图 6(b) 为由近似系数  $A^9$  重构的轮廓,可以看出,已接近良性轮廓,含有的高频部分已接近零。由图 6(d) 明显可以看出,由于缺少小波子带  $D^6$ ,轮廓虽然去除了原图像中的细微波动(纹理部分),但轮廓的凹凸起伏部分(结构不规则部分)稍有缺失。这说明小波子带  $D^6$  对轮廓的结构分量具有重要作用,而小波子带  $D^{10}$  则无作用( $A^9$  已经接近平滑)。图 6(c) 准确描述了轮廓的结构不规则性。依据 HD 分布和实验分析,取  $j=6 \sim 9$  为肿瘤的类型间区分度显著性子带,由于这个子带范围显著地描述了轮廓的结构不规则性,亦称为结构子带。

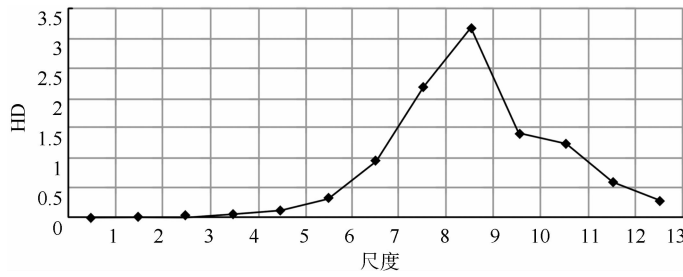


图 5 基于小波各层细节能量的 HD 值分布

Fig. 5 HD distribution based on the energy of detail coefficients against scales

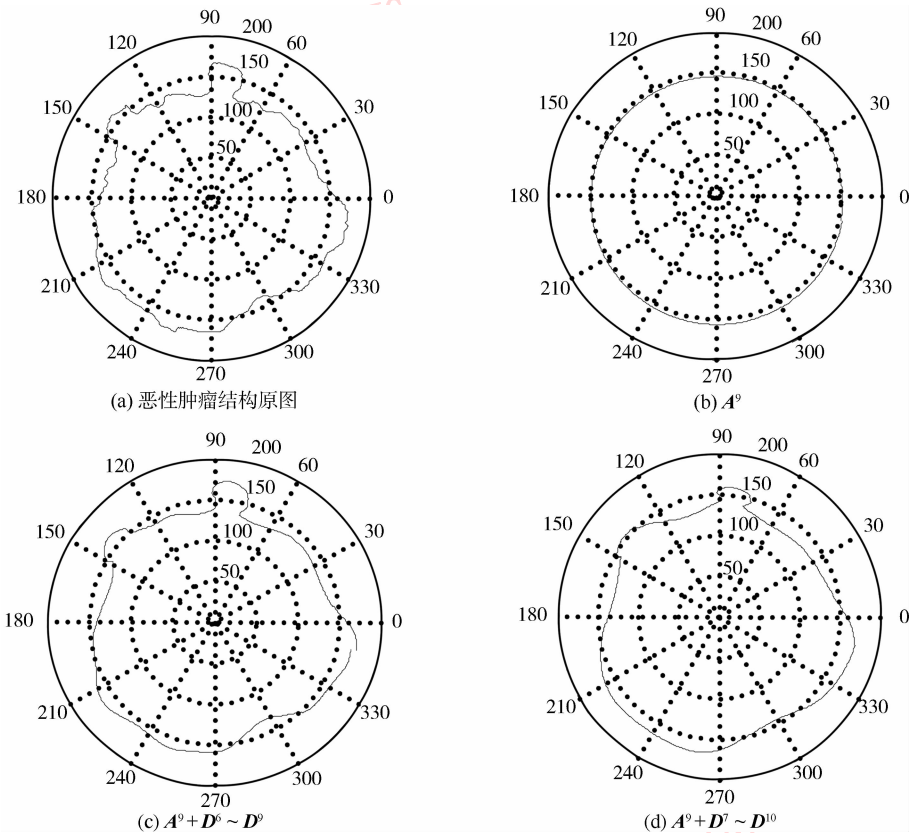


图 6 肿瘤结构分量提取

Fig. 6 Extraction of structure components of a skin lesion

确定了显著小波子带后,轮廓结构部分的重构变得更为方便。即根据 1.2.3 节分析,确定分解层次和分界点( $s = 9, s_i = 6$ )。图 7 显示了良性轮廓和恶性轮廓的提取过程,图 7(a)(d)分别为原始肿

瘤轮廓,图 7(b)(e)为由  $A^9$  和  $D^1$  至  $D^5$  重构得到的轮廓纹理部分( $A^9$  用来将纹理部分放大),图 7(c)(f)为由  $A^9$  和  $D^6$  至  $D^9$  重构得到,描述了原始轮廓的结构部分,该部分含有区分肿瘤的重要信息。

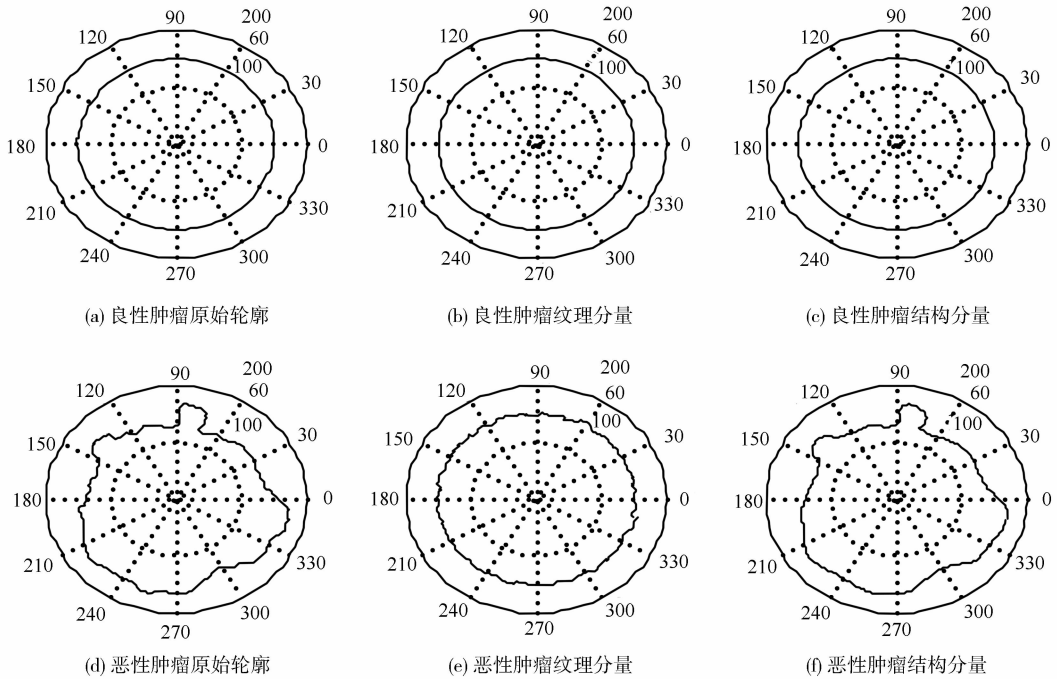


图 7 基于显著小波子带的轮廓重构

Fig. 7 Contour reconstruction based on salient wavelet subbands

表 1 特征相关性选择结果

Tab. 1 Selection results of feature correlativity

层数	特征						
	小波能量均值	小波熵	极限宽	方差	径向偏差	轮廓粗糙度	不规则描述子
6	✓	✓	✓	×	✓	×	✓
7	✓	✓	✓	×	×	✓	
8	✓	✓	×	✓	×	×	
9	✓	✓	✓	✓	×	×	

#### 4.2 特征选择

原始特征向量由 25 个描述不规则性的特征组成,其中 24 个特征分别由 4 个结构子带上提取的 6 个不规则特征描述子表示,一个在结构子带  $s_i = 6$  处的轮廓不规则描述子  $IM$ 。

##### 4.2.1 相关性分析

对提取的特征,计算特征间的相关系数。图 8 给出了特征相关强度的图示。

根据第 3 节的特征选择法,经过相关性分析去除其中 9 个特征。表 1 给出了相关性分析后

特征的选择情况(“✓”表示留下的特征;“×”表示被去除的特征)。从表中可看出,小波能量均值、小波熵和极限宽在小波显著性子带具有较好的独立性。方差在较高子带层(8 和 9)具有较好的独立性,径向偏差度在第 6 层、轮廓粗糙度在第 7 层和不规则描述子在第 6 层具有较好的独立性。

##### 4.2.2 概率密度分析

通过相关性分析,去除冗余特征后得到 16 个相对独立特征。应用 3.2 节的概率密度分布分析,计

算良、恶性肿瘤样本集特征的概率分布。首先计算良性肿瘤各特征的累加概率  $p_b = \sum_{i=1}^k p_i^b$ , 找到合适的  $k$ , 使得所有特征的累加概率  $p_b > T$  (实验中得到当  $k$  为 143 时, 所有  $p_b > 0.75$ , 故分类阈值  $T =$

0.75); 然后再根据所确定的  $k$  值计算对应恶性各特征的累加概率  $p_m = \sum_{i=k+1}^N p_i^m$  进行特征选择。图 9 分别给出了良、恶性特征的累加概率分布。从图 9 (a) 可以看出, 所有 16 个特征均满足  $p_b > T$ 。

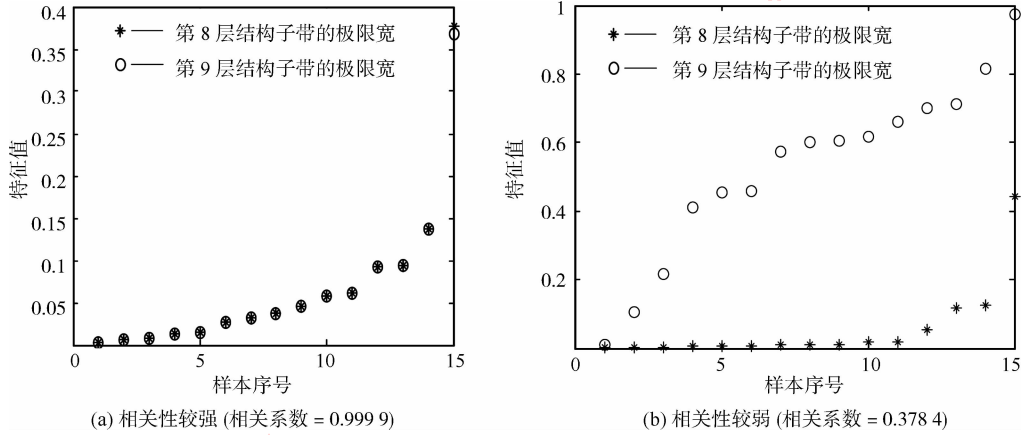


图 8 特征相关性示意图

Fig. 8 Diagram of feature correlation

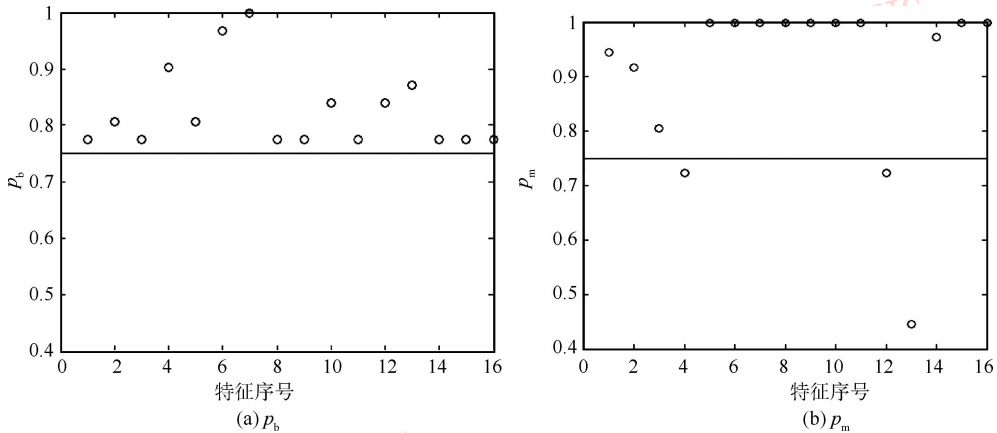


图 9 16 个特征的累加概率  $p_b, p_m$  分布图

Fig. 9 The distribution of cumulative probability  $p_b, p_m$  for 16 features

图 9(b) 中有 3 个特征的累加概率密度  $p_m$  小于阈值  $T = 0.75$ , 分别是 9 层的小波均能, 8 层和 9 层的方差。因此经过基于相关性和概率分析的特征选择后, 一共保留 13 个特征作为后续的分类器输入。从图 9 中可以看出有些特征累加概率  $p_m$  为 1, 这是由于基于此特征的良、恶性两类完全可分。

### 4.3 基于肿瘤轮廓结构不规则性的分类

选择 BP 神经网络作为分类器, 考察原 25 个特征经过选择后保留的 13 个特征的分类性能<sup>[14]</sup>。网络结构为 3 层: 含有 13 个特征的输入层, 具有 3 个节点的隐层和只含 1 个节点的输出层。训练参数设

置如下: 训练误差设置为 0.01, 训练次数设置为 1000。实验中将实验样本 134 幅图像分为训练集和测试集, 训练集由 31 个良性样本和 36 个恶性样本组成, 其余作为测试集。为了检测多尺度特征的分类能力, 实验中首先对特征选择前的单尺度、多尺度特征, 以及基于小样本和多样本的特征构成的分类器分别进行了分类性能研究<sup>[15]</sup>。接着, 对多尺度特征(大样本/小样本)选择前后性能进行比较研究。

#### 4.3.1 特征选择前分类器性能研究

实验中 大样本训练集由 31 幅良性痣和 36 幅恶性黑色素瘤图像构成, 小样本训练集由大样本

训练集中分别选择 9 个良性样本和 9 个恶性样本构成。在大样本训练集中随机选择小样本的组合数为  $C_{36}^9 \times C_{31}^9$ ,为了平衡小样本集的随机性和泛化性的矛盾,选择了其中 15 个组合,按照该 15 种样本集训练下的神经网络分类 ROC 曲线的均值作为小样本训练集的分类结果。特征选择前的分类实验由 4 种方案构成:1)大样本单尺度神经网络分类器;2)小样本单尺度神经网络分类器;3)大样本多尺度神经网络分类器;4)小样本多尺度神经网络分类器。其中单尺度是指在  $s_1$  层的 7 个特征,多尺度由显著小波子带的 25 个特征组成(特征选择前)。图 10 对比了单尺度/多尺度特征分类 ROC 曲线。从图中可以明显地看出,多尺度特征比单尺度特征的分类能力更强。可以看出,大样本多尺度分类器的 ROC 曲线覆盖的面积最大,最接近坐标系的(0,1)位置,小样本单尺度分类器最差,接近 45° 直线。另外,小样本多尺度分类性能与大样本多尺度接近。由此可以看出,取自显著小波子带的多尺度特征凸显了肿瘤轮廓不规则性的类间差异,提高了分类的准确性。

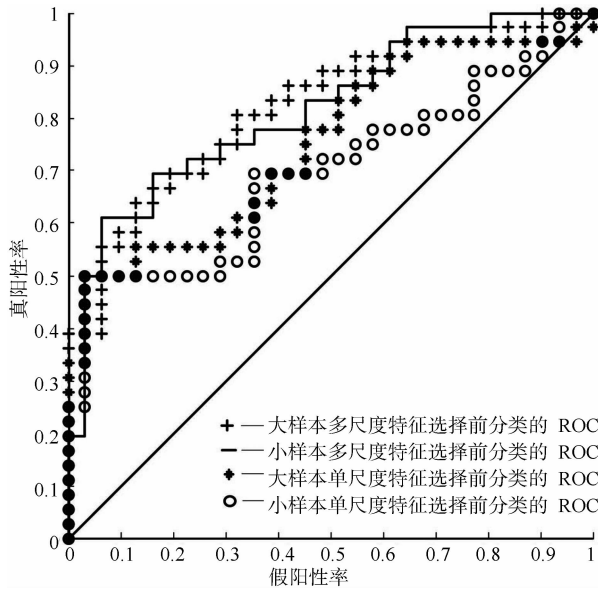


图 10 单尺度/多尺度分类器 ROC 曲线比较

Fig. 10 Performance comparison of single-scale/multi-scale classifiers

### 4.3.2 特征选择对多尺度分类器性能影响

由上节实验结果可知,单尺度分类器的性能较差。由于实验中没有对单尺度方案进行特征选择,因此特征选择仅在多尺度方案中进行。对于多尺度特

征来说,原始特征为 25 个,特征选择后为 13 个特征。实验方案包括多尺度大样本特征选择前后的神经网络分类器以及多尺度小样本特征选择前后的神经网络分类器,神经网络训练用的样本集与上节一致,并且小样本多尺度神经网络分类器的 ROC 曲线取 15 组结果的平均值。图 11 给出了多尺度特征分类器的性能比较。从图中 4 条 ROC 曲线可以看出:经特征选择后的小样本多尺度神经网络分类器的性能最好,依次为特征选择后大样本多尺度、特征选择前大样本多尺度和小样本多尺度神经网络分类器。表 2 给出了各种分类器 ROC 曲线所覆盖的面积,面积值表征分类性能的优劣。除此之外,表 2 还列出了在图 10 和图 11 的任一方案(共 6 种)中,基于肿瘤轮廓结构不规则特征的分类器所能达到的最佳灵敏度和特异度。其中经特征选择后的小样本多尺度神经网络分类器能达到的特异度和灵敏度值为 0.903, 0.833, 优于特征选择后的大样本多尺度神经网络分类器的对应值为 0.742, 0.833。另外,特征选择前的小样本/大样本多尺度分类器性能比特特征选择后的差(难以同时达到较高的特异度和灵敏度),小样本单尺度的分类能力最差。

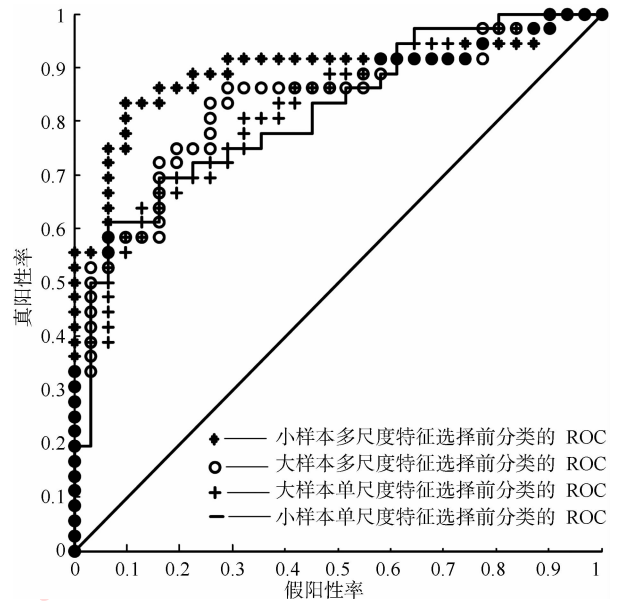


图 11 特征选择前后分类器 ROC 曲线比较

Fig. 11 Performance comparison of classifiers before and after feature selection

综合以上分类器实验结果,可以看出:1)多尺度特征的轮廓不规则性表达优于单尺度特征。这是因为轮廓结构不规则性并不体现在单一小波子带,

而是由多个显著性子带共同描述。2) 特征选择(去相关和概率可分性分析)可以提高分类器的性能。3) 在特征选择后,统计实验表明,小样本多尺度分类器性能优于大样本多尺度分类器。同时注意到特征选择前小样本多尺度分类器可以达到与大样本多尺度分类器相同的性能。笔者认为这是由于 BP 网络训练算法易于陷入局部最小造成的。当样本数较少时,神经网络易于较快收敛,使得小样本分类器具有较好的泛化能力;而大样本分类器数据量大,数据噪声使分类复杂性增加,神经网络收敛慢,陷入局部极小,分类能力变差。

表 2 不同分类器下的分类性能比较表

Tab. 2 Classification performances under different schemes

实验方案	特异度/%	灵敏度/%	ROC 面积
小样本单尺度	0.645 66	0.694 44	0.695 34
大样本单尺度	0.879 03	0.555 56	0.751 79
小样本多尺度特征选择前	0.838 71	0.694 44	0.815 41
大样本多尺度特征选择前	0.806 45	0.694 44	0.819 89
小样本多尺度特征选择后	0.903 23	0.833 33	0.890 68
大样本多尺度特征选择后	0.741 94	0.833 33	0.833 33

## 5 结 论

探讨了小波多分辨率分析在皮肤肿瘤轮廓结构不规则性度量以及基于多尺度特征的分类器设计问题。针对肿瘤原始轮廓中包含的纹理不规则性,不利于正确分类这一难点,提出以小波子带能量为特征,寻找能够较好区分良恶性肿瘤的小波子带——显著性小波子带。进而利用显著性小波子带确定出肿瘤轮廓结构和纹理分量在频域的分界点,重构出肿瘤轮廓的结构分量。对肿瘤轮廓在显著小波子带上所呈现的多尺度信号进行不规则性特征度量,提出基于统计的特征和几何特征的不规则性描述子,以及由这些特征描述子构成多尺度特征向量。讨论了基于特征相关性和概率分布的特征选择方法,并对单尺度/多尺度、小样本/大样本,以及特征选择前后神经网络分类器性能进行了实验研究。

肿瘤轮廓结构不规则性分析和分类的研究工作有以下几个特点:1) 提出利用肿瘤轮廓小波子带能量的 HD 分布寻找显著性小波子带的方法,进而构建肿瘤轮廓结构分量模型;2) 给出轮廓结构分量不

规则特征的多尺度描述——显著性小波子带信号不规则性描述簇,并在此基础上进行了特征选择;3) 针对小样本分类器问题,对单尺度、多尺度特征、特征选择前后所构建的神经网络分类器进行了实验研究。实验结果表明:特征选择后的小样本多尺度特征分类器的灵敏度、特异度和 ROC 面积指标优于大样本下多尺度特征分类器,以及单尺度特征分类器。研究结果提供了在临床缺少病理确认样本情况下,利用小波多分辨率分析构建基于显著子带域的多尺度特征向量,为设计出具有较高特异度和灵敏度的良恶性肿瘤分类器提供了有效途径。

## 参考文献 (References)

- [1] Grana C, Giovanni P, Cucchiara R, et al. A new algorithm for border description of polarized light surface microscopic images of pigmented skin lesions [J]. IEEE Trans. on Medical Imaging, 2003, 22(8):959-964.
- [2] Lee T K, McLean D I, Atkins M S. Irregularity index: a new border irregularity measure for cutaneous lesions [J]. Medical Image Analysis, 2003, 7(1):47-64.
- [3] Jayalalitha A, Uthayakumar R. Estimating the skin cancer using fractals [C]//Proceeding of the International Conference on Computational Intelligence and Multimedia Applications. Sivakasi, India: IEEE, 2007:307-312.
- [4] Liao Y C, Hung K C, Ku C T, et al. Wavelet octave energy for breast tumor classification on sonography: A new shape feature [C]// International Conference on Networking, Sensing and Control. Okayama, Japan: IEEE Press, 2009:388-392.
- [5] Patwardhan S V, Dhawan A P, Relue P A. Classification of melanoma using tree structured wavelet transforms [J]. Computer Methods and Programs in Biomedicine, 2003, 72(3):223-239.
- [6] Clawson K M, Morrow P, Scotney B, et al. Analysis of pigmented skin lesion border irregularity using the harmonic wavelet transform [C]//Proceedings of the 13th International Conference on Machine Vision and Image Processing. Dublin, Ireland: IEEE Press, 2009:18-23.
- [7] Chiem A, Al-Jumaily A, Khushaba R N. A novel hybrid system for skin lesion detection [C]//Proceedings of the 3rd International Conference on Intelligent Sensors, Sensor Networks and Information Processing. Melbourne, Australia: IEEE, 2007:567-572.
- [8] Lin K H, Guo B F, Lam K M, et al. Human face recognition using a spatially weighted modified hausdorff distance [C]// Proceedings of the International Symposium on Intelligent Multimedia, Video and Speech Processing. H. K., China: IEEE Press, 2001:477-480.

- [ 9 ] Aribisala B S, Claridge E. A border irregularity measure using a modified conditional entropy method as a malignant melanoma predictor[J]. *Image Analysis and Recognition*, 2005, 3656: 914-921.
- [10] Ma Li. Description of boundary irregularity on melanomas using multiscale local fractal Dimension [J]. *Journal of Image and Graphics*, 2010, 15(5): 736-741. [马莉. 基于多尺度局部分形维的黑色素瘤轮廓不规则性描述方法[J]. *中国图象图形学报*, 2010, 15(5): 736-741. ]
- [11] Li Yun, Wu Zhongfu, Liu Jiamin, et al. Efficient feature selection for high-dimensional data using two-level filter[C] // *Proceedings of the 3rd International conference on Machine Learning and Cybernetics*. Shanghai: IEEE Press, 2004. 1711-1716.
- [12] Zhang Li, Sun Gang, Guo Jun. Unsupervised feature selection method Based on K-means clustering[J]. *Application Research of Computers*, 2005, 22(3): 23-25. [张莉, 孙钢, 郭军. 基于 K-均值聚类的无监督的特征选择方法[J]. *计算机应用研究*, 2005, 22(3): 23-25. ]
- [13] Wei Ying, Guo Wei, Sun Yuefang, et al. Feature selection and classification algorithm for region of interest in lung cancer CAD system[J]. *Information and Control*, 2008, 37(4): 445-458. [魏颖, 郭薇, 孙月芳等. 面向肺癌 CAD 系统的感兴趣区域特征选择与分类算法[J]. *信息与控制*, 2008, 37(4): 445-458. ]
- [14] Lin D T, Yan C R, Chen W T. Autonomous detection of pulmonary nodules on CT images with a neural network-based fuzzy system[J]. *Computerized Medical Imaging and Graphics*, 2005, 29(6): 447-458.
- [15] Wang Kena, Zou Bei, Huang Wenmei. A distorted Image correction method based on Neural networks [J]. *Journal of Image and Graphics*, 2005, 10(5): 603-607. [王珂娜, 邹北骥, 黄文梅. 一种基于神经网络的畸变图像校正方法[J]. *中国图象图形学报*, 2005, 10(5): 603-607. ]