

乳腺超声中多普勒特征的提取及应用

刘 研, 黄剑华, 唐降龙, 张英涛, 刘 博

(哈尔滨工业大学 计算机科学与技术学院, 哈尔滨 150001, snacly@163.com)

摘 要: 为弥补以往单纯利用乳腺超声图像中灰度信息进行分类的过程中, 易受噪声和其他因素影响而造成误分的缺点, 提出了一种基于纹理特征和多普勒特征相结合的乳腺超声诊断系统. 针对乳腺超声图像中的动态彩色多普勒信息建立数学模型进行特征提取, 并结合图像中的纹理信息进行肿瘤的良恶性分类. 实验结果表明, 该系统的诊断准确性为 92.30%, 敏感性为 91.42%, 特异性为 93.33%, 假阳性率为 94.11%, 假阴性率为 90.32%, 接收特性 AUC 面积为 0.9523, 性能优于单纯利用纹理特征的分类结果, 从而证明其在提高分类精度方面的有效性.

关键词: 乳腺超声; 纹理特征; 辅助诊断系统; 多普勒特征; 分类

中图分类号: TP391.4

文献标志码: A

文章编号: 0367-6234(2010)09-1424-04

Feature extraction and application of color Doppler in breast ultrasound image

LIU Yan, HUANG Jian-hua, TANG Xiang-long, ZHANG Ying-tao, LIU Bo

(School of Computer Science and Technology, Harbin Institute of Technology,
Harbin 150001, China, snacly@163.com)

Abstract: To reduce the impact of noises and other factors in traditional feature extraction of breast ultrasound based on gray level images, we propose a novel computer aided diagnosis system for breast cancer. Mathematical model was established using both texture features and color Doppler features to distinguish the benign and malignant tumors. The accuracy, sensitivity specificity, false positive rate and false negative rate of the proposed system are 92.30%, 91.42%, 93.33%, 94.11% and 90.32%, respectively. Experimental results show that the proposed system is valuable for radiologists to improve the accuracy of the diagnosis of breast cancer.

Key words: breast ultrasound; texture feature; computer aided diagnosis; color Doppler features; classification

乳腺癌是常见的女性恶性肿瘤^[1], 多普勒超声作为一种无痛、快捷、易操作、在活体上可重复实施、能显示肿瘤全貌以及生理病理特性的检查方法, 可用于术前评估肿瘤血管的生成、抗血管生成和预测预后, 具有广泛的临床意义^[2]. 目前的乳腺超声分类方法主要是对肿瘤的纹理特征以及形态特征进行提取, 进而利用分类器进行分类. 文

献[3-4]采用自相关系数和均方差系数提取纹理特征并用 SVM 分类器对乳腺 X 超声图像进行分类, 其结果证明二维超声图像中的肿瘤的纹理特征的对肿瘤良恶性分类具有较好的效果. 文献[5]采用了灰度自相关矩阵提取纹理特征, 并且应用了逐步回归方法进行分类. 但是, 上述方法只考虑了超声图像中的灰度信息, 因此可能会受到设备仪器的影响而出现噪声干扰产生虚假纹理. 而颜色特征则充分利用了图像的色彩信息, 侧重于图像整体信息的描述, 不受图像旋转和平移变化的影响, 归一化可不受图像尺度变化的影响.

文献[6]阐述了一种基于 HSV 模型来提取血流特征并用 SVM 进行分类的方法, 取得了良好的

收稿日期: 2010-07-06.

基金项目: 国家自然科学基金资助项目(60873142, 30670546);
哈尔滨科技人才创新基金资助项目(2008RFQXS037,
2009RFQXS032); 哈尔滨市科技局优秀学科带头人资
助项目(2009RFXXS211).

作者简介: 刘 研(1981—), 男, 讲师, 博士研究生;
唐降龙(1960—), 男, 教授, 博士生导师.

分类效果. 由于不同类别的多普勒图像序列具有其特有的血流特征,而且提取动态图像序列特征比单一静态特征所提取的信息更为全面,因此本文设计了一种分类系统,针对乳腺超声图像中的灰度信息和彩色信息进行特征提取并将其结合,最后通过 SVM 进行分类从而验证系统的有效性.

1 提取纹理特征

纹理特征的提取首先要确定肿瘤所在位置的感兴趣区域(ROI),并将 ROI 图像从整个图像中进行分离. 本文确定 ROI 区域的方法为人工选取,用一矩形框将肿瘤所在区域进行定位,进而对区域内图像进行纹理特征的提取. 灰度共生矩阵是图像灰度的二阶统计度量,其作用是描述纹理中灰度基元之间的空间联系^[7].

设 $f(x,y)$ 为一幅二维数字图像,其大小为 $M \times N$,灰度级别为 CON ,则满足一定空间关系的灰度共生矩阵为

$$P(i,j) = \#\{(x_1,y_1),(x_2,y_2) \in M \times N | f(x_1,y_1) = i, f(x_2,y_2) = j\}. \tag{1}$$

式中: $\#(x)$ 为集合 x 中的元素个数,若 (x_1,y_1) 与 (x_2,y_2) 间距离为 d ,两者与坐标横轴的夹角为 θ ,则可以

得到各种间距及角度的灰度共生矩 $P(i,j|d,\theta)$.

本文用能量、对比度、熵和逆差矩等作为统计分量. 其中,各种特征信息定义为:

1) 能量(角二阶矩).

$$f_{ASM} = \sum_i \sum_j P(i,j|d,\theta)^2. \tag{2}$$

2) 对比度(或称主对角线惯性矩).

$$f_{CON} = \sum_i \sum_j (i-j)^2 P(i,j|d,\theta). \tag{3}$$

3) 熵.

$$f_{ENT} = - \sum_i \sum_j P(i,j|d,\theta) \log P(i,j|d,\theta). \tag{4}$$

4) 逆差矩.

$$Uniformity = \sum_{i=1}^m \sum_{j=1}^m \frac{r(c_i,c_j)}{1 + |c_i - c_j|}. \tag{5}$$

5) 相关.

$$C = (d,\theta) = \frac{\sum_{i,j} (i - \mu_x)(i - \mu_y) P(i,j|d,\theta)}{\sigma_x \sigma_y}. \tag{6}$$

其中,
$$\mu_x = - \sum_i i \sum_j P(i,j|d,\theta), \tag{7}$$

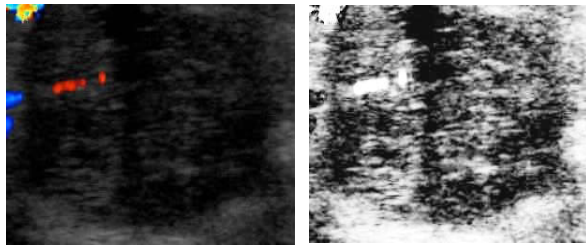
$$\mu_y = - \sum_j j \sum_i P(i,j|d,\theta), \tag{8}$$

$$\sigma_x = \sum_i (i - \mu_x)^2 \sum_j P(i,j|d,\theta), \tag{9}$$

$$\sigma_y = \sum_j (j - \mu_y)^2 \sum_i P(i,j|d,\theta). \tag{10}$$

在实际应用中,由于灰度共生矩阵的计算量

大小是根据图像灰度级和图像大小而确定. 为了减小计算复杂度,本文将对原图像进行灰度级量化,以将原影像的灰度级压缩到较小的范围,从而减少共生矩阵的维数. 但灰度级的压缩会导致图像的清晰度降低,所以首先要对图像进行直方图均衡化处理(图1)且增强对比度,从而最大程度地保留原图像的纹理特征.



(a) 原始乳腺超声图像 (b) 处理后的乳腺超声图像
图1 对原始图像进行直方图均衡化处理

2 提取血流特征

血流特征分为血流动力学特征和形态学特征. 由于二维多普勒超声仪器的限制,拟将对血流的动力学特征进行提取,其过程主要包括对血管位置的检测以及特征提取两个环节.

2.1 血管的检测

血管检测的目的是确定血管在图像中所处的位置. 由于在各帧图像中,血管区域显示为具有特定颜色的彩色点(其中,不同的颜色可以表示血流的方向和流速),而彩色和灰度图像的RGB分量间的差值不同,灰度图像的RGB各分量的值比较接近,而彩色图像的RGB各分量的值之间的差距则相对较大,因此可以通过阈值化方法来确定血管区域. 首先计算在各点的RGB分量之间的差值,然后设置一个阈值 T ,对RGB分量之间的差值进行阈值化,从而取得该帧血管位置的图像(T_{gray}). 设置方法为

$$T_{gray} = |R - B| < T \& |G - B| < T \& |R - G| < T. \tag{11}$$

在进行对提取出的像素点进行归一化后,通过提取关键像素的坐标来获得对应的彩色点在整个图像中的坐标值. 图2为血管位置的确定过程.



图2 血管位置确定过程

2.2 血流特征的提取

特征的提取和分析是计算机辅助诊断的关键

环节. 通过提取几个对肿瘤的供血有重要影响的动力学特征对肿瘤进行分析. 动力学特征主要包括血流丰富程度、心脏收缩期流速峰值、舒张期末速率、平均速率、肿瘤内动脉的阻力指数以及搏动指数、最大流速变化率等. 临床诊断表明, 恶性肿瘤的收缩期流速峰值 ($V_{\max}/\text{m} \cdot \text{s}^{-1}$), 肿瘤内动脉的阻力指数 (RI), 搏动指数 (PI) 和平均流速 ($V_{\text{mean}}/\text{m} \cdot \text{s}^{-1}$) 要明显高于良性肿瘤.

在彩色多普勒中, 血流的速度可以用 HSV 模型中的 V 分量来映射. V 值越高表明血流速度越高, 相反 V 值越低表明血流速度越低. 根据这一特性, 首先将 RGB 模型转化为 HSV 模型, 然后求得血管所在区域中各个像素点的 V 值, 并提取血流丰富程度、收缩期流速峰值、肿瘤内动脉的阻力指数以及搏动指数等. 其具体定义为:

1) 收缩期流速峰值 ($V_{\max}/\text{m} \cdot \text{s}^{-1}$): 即一个彩色多普勒序列中 V 值的和最大的帧.

$$V_{\max} = C \cdot \max_{k=2,3,\dots,N-1} \{V_k\}. \quad (12)$$

2) 舒张期末速率 ($V_{\min}/\text{m} \cdot \text{s}^{-1}$).

$$V_{\min} = C \cdot \min_{k=2,3,\dots,N-1} \{V_k\}. \quad (13)$$

3) 平均流速 ($V_{\text{mean}}/\text{m} \cdot \text{s}^{-1}$).

$$V_{\text{mean}} = C \cdot \text{mean}_{k=2,3,\dots,N-1} \{V_k\}. \quad (14)$$

4) 肿瘤内动脉的阻力指数 (RI).

$$RI = \frac{V_{\max} - V_{\min}}{V_{\max}}. \quad (15)$$

5) 搏动指数 (PI).

$$PI = \frac{V_{\max} - V_{\min}}{V_{\text{mean}}}. \quad (16)$$

最大流速变化率 ($Var_{\max}/\text{m} \cdot \text{s}^{-1}$) 反映了当前样本血流变化的大小, 从而反映出动脉供血的多少. 动脉数量多, $Var_{\max}$ 相对则大; 反之, $Var_{\max}$ 相对则小.

$$Var_{\max} = C \cdot \max_{k=2,3,\dots,N-1} \left(\sqrt{\frac{\sum_{k \in N} (V_k - V_{k-1})^2}{\Delta t}} \right). \quad (17)$$

式中: V 为 HSV 模型中每个彩色点的亮度. C 为一常量系数. N 为在一个搏动周期中所代表的帧数. 根据临床所描述的病理意义, 恶性肿瘤中动脉的数量, 血流流速、阻力指数以及搏动指数要相对高于良性肿瘤. 因此, 通过对各样本中血流变化率来判断肿瘤中血流的丰富程度, 可以弥补单纯提取纹理特征所造成的提取信息不全、受噪声等因素干扰的缺点.

3 实验结果分析

本文应用了支持向量机分类器^[8], 在有监督

学习的模式下对 65 个多普勒序列 (30 个良性, 35 个恶性) 样本进行分类实验. 为了克服训练样本数量较小这一缺点, 这里采用 leave-one-out 方法对分类效果进行训练和评估.

在本实验中, 所有的图片均来源于哈尔滨医科大学第二附属医院专家诊断, 且这些病例经过活体切片的确认. 所用超声诊断仪为 GE VIVID-7, 探头为线阵, 工作频率为 5 ~ 14 MHz. 该样本集中包括 30 个良性病例以及 35 个恶性病例. 每个样本均被分离成若干个 ROI 序列, 并对每个样本提取了相同的帧数进行特征提取. 由于血流信号不稳定, 造成部分相邻帧彩色区域的缺失. 本文将彩色信号出现频率 > 30% 的区域设为血流区域.

该分类系统中设置了下列重要参数: 多项式核函数参数 (degree), 算法终止条件 (eps), 多项式, 径向基函数和 Sigmoid 函数的参数 (gamma), 多项式和 Sigmoid 函数的参数 (coef0), 惩罚性因子 (C), 为了能使系统达到最佳分类状态, 采用交叉项验证方法来调整参数.

这里采取的训练方式为:

1) 从 k 个良性样本和恶性样本中各选取一个样本, 并用其余 $k - 1$ 个样本进行训练.

2) 从其余 $k - 1$ 个样本中选取一个良性或者恶性样本, 进行分类.

3) 从其余 $k - 1$ 个样本中再选取一组良性和恶性的样本, 并重复步骤 1.

由于所提取的血流特征中各特征之间数量级差距较大, 为达到良好的分类效果, 分类前将体征进行归一化处理. 这里参数设置为 degree = 3.0, eps = 0.01, gamma = 1, coef0 = 1, $C = 256$ 时, 可使分类效果达到最佳状态. 为评价诊断的系统性能, 采用精确性、敏感性、特殊性、错误正比率、错误负比率和接受特性曲线等 6 个指标来进行评估.

$$\text{精确性} = \frac{TP + TN}{TP + TN + FP + FN};$$

$$\text{敏感性} = \frac{TP}{TP + FN};$$

$$\text{特殊性} = \frac{TN}{TN + FP};$$

$$\text{错误正比率} = \frac{FP}{TP + FP};$$

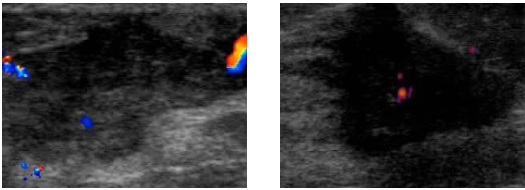
$$\text{错误负比率} = \frac{FN}{TN + FN};$$

式中: TP 为被正确分类的恶性肿瘤的数量. TN 为被正确分类的良性肿瘤的数量. FP 为被错误分类的良性肿瘤的数量. FN 为被错误分类的恶性肿瘤的数量. 实验结果如表 1 所示.

表1 本系统与文献[3]、文献[7]系统仿真结果

分类结果	良性	恶性	精确性/%	敏感性/%	特殊性/%	错误正比率/%	错误负比率/%
本系统正确数目	28	32	92.30	91.42	92.33	94.11	90.32
文献[3]正确数目	22	28	76.92	80.00	73.33	77.78	75.86
文献[6]正确数目	26	31	87.69	88.57	86.67	87.57	86.67

在对于纹理间粗细和疏密相差较小的图片进行分类时,本系统的分类效果相对于文献[3]中系统的分类效果得到了明显的提高,相对于文献[7]中系统的效果也有所提高.图3所示为单用纹理特征而错分的样本,而利用本文所提出的系统则可以正确地识别.



(a) 被文献[3]系统误分的良性样本 (b) 被文献[3]系统误分的恶性样本

图3 文献[3]系统错分样本

对类似图3中的被错分纹理间粗细和疏密相差较小或受噪声干扰的样本,本系统均能正确识别.整体对比实验数据如表2所示.

表2 试验对比结果

分类结果	良性	恶性
文献[3]错分数目	6	4
本系统错分数目	1	0

接受特性曲线 ROC 法可用于描述相对于特殊性余数(1-特殊性)的敏感性,被广泛用于医学诊断中,图4为文献[3],[7]中系统与本文所提出系统的 ROC 曲线.对于曲线的曲线下面积(AUC)分别计算可得,文献[3]中 $AUC=0.8226$,文献[7]中 $AUC=0.9113$,而本文中的 $AUC=0.9523$.

从实验可以看出,应用多普勒特征与纹理特征相结合进行肿瘤的良恶性分类,可弥补超声图像所带来的干扰,从而得到更加的分类效果.

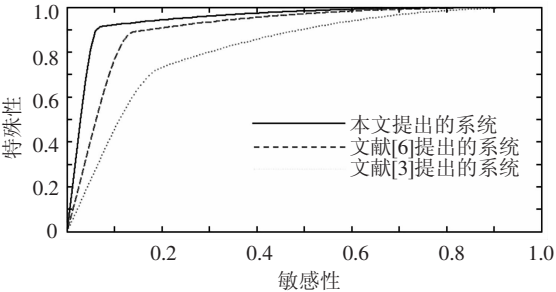


图4 系统的分类效果比较

征相结合,可以弥补单纯靠纹理特征分类而受到的干扰,进而提高分类精度.

2)实验结果表明,加入血流特征后系统的分类精度要比单纯提取特征的分类精度提高15.38%,敏感性提高11.42%,特殊性提高19%,错误正比率提高16.33%,错误负比率提高14.46%.可以看出本系统可以明显弥补单纯提取纹理特征的不足.

参考文献:

[1] 陈菲,吴凤霞. 超声在乳腺癌中的应用[J]. 医学综述, 2009, 15(2):279-281.

[2] FANELLI M, LOCOPO N, GATTUSO D, *et al.* Assessment of tumor vascularization: Immunohistochemical and non-invasive methods [J]. Int J Biol Markers, 1999, 14(4): 218-231.

[3] HUANG Yu-Len, CHEN Dar-Ren. Support vector machines in sonography application to decision making in the diagnosis of breast cancer [J]. Journal of Clinical Imaging, 2005, 29(3):179-184.

[4] SIVARAMAKRISHNA R, POWELL K A, LIEBER M L, *et al.* Texture analysis of lesions in breast ultrasound images [J]. Computerized Medical Imaging and Graphics, 2002, 26(5): 303-307.

[5] CHANG R F, WU W J, WOO K M, *et al.* Improvement in breast tumor discrimination by support vector machines and speckle-emphasis texture analysis [J]. Ultrasound in Med & Biol, 2003, 29(5): 679-686.

[6] ZHANG Ying-tao, CHENG H D, GUO Yan-hui, *et al.* Classification of breast masses using color doppler flow imaging [C]//Proceedings of the 11th Joint Conference on Information Sciences. Atlantis; Published by Atlantis Press, 2008: 554-558.

[7] HARALICK R M, SHANMUGAM K, DINSTEN I. Textural features for image classification [J]. IEEE Transactions on Systems Man and Cybernetics, 1973, 3(6): 610-621.

[8] STAELIN Carl. Parameters Selection for Support Vector Machines. HPL2002-354 (R. 1) [R/OL]. Development Company, 2002[2003-11-10]. <http://www.hpl.hp.Com/techreports/2002/HPL-2002-354R1.pdf>.

(编辑 张 红)

4 结 论

1)本系统将多普勒血流特征与肿瘤纹理特