

基于 MCA 的乳腺 X 线图像中肿块的 自适应检测方法

王 颖, 李 洁, 高新波

(西安电子科技大学电子工程学院, 陕西西安 710071)

摘 要: 针对肿块通常大小和形状各异, 并且边缘模糊的特点, 提出了一种基于形态学成分分析 (MCA) 和直方图自适应阈值搜索的肿块检测方法. 首先通过引入 MCA 方法有效地抑制了血管和纤维对检测的影响, 在此基础上设计了一种基于直方图的自适应阈值搜索策略, 根据肿瘤的生长特性, 通过自适应阈值和多灰度同心层方法, 有效地检测乳腺 X 线图像中的病变区域. 通过对真实乳腺 X 线图像的测试实验, 其结果表明, 所提出的方法能够检测出不同类型的肿块区域, 并且假阳性区域的数量在可接受的范围内, 能够有效地辅助医生进行诊断.

关键词: 计算机辅助检测; 肿块检测; 形态学成分分析; 自适应阈值搜索; 灰度同心层

中图分类号: TP391.41 **文献标识码:** A **文章编号:** 0372-2112 (2011) 03-0525-06

An Adaptive Mass Detection Method on Mammography Based on MCA

WANG Ying, LI Jie, GAO Xin-bo

(School of Electronic Engineering, Xidian University, Xi'an, Shaanxi 710071, China)

Abstract: For capturing various shapes and blurry margins of tumors, a new mass detection method based on Morphological Component Analysis (MCA) and adaptive histogram threshold searching is proposed in this paper. Firstly, MCA method is introduced to restrain the influence of blood vessels and fibrous structures in mammograms. Then, an adaptive threshold searching method is designed according to the histograms of breast region. Finally, following the Gaussian-like growing feature of masses, the suspicious regions are effectively detected according to the adaptive thresholds and multi-intensity concentric layer methods. The experimental results on mammograms illustrate that the proposed method could effectively detect different types of masses with acceptable false positives, and could be a useful tool for assisting doctors.

Key words: computer-aided detection; mass detection; morphological component analysis; adaptive threshold searching; multiple intensity layers

1 引言

乳腺癌是一种常见于女性的高发恶性肿瘤病. 乳腺 X 线片已广泛应用于乳腺疾病的普查中, 但仍有相当一部分病例被漏检或者误检, 这是由于早期乳腺癌的症像在 X 线片上的特征不够明显, 并且与正常组织相似, 再加上观察者的视觉疲劳等主观因素影响, 致使乳腺癌的早期检出率不高. 因此, 如何采用图像处理方法, 凸显病变区域特征, 从而有效地辅助临床医生提高早期乳腺癌的检出率, 已成为图像处理及医学等跨学科领域的研究热点.

肿块和微钙化点是早期乳腺癌在 X 线图像上的主要表现^[1], 而肿块由于其形状大小各异, 边缘模糊不清, 并且与正常纤维组织极为相似, 一直都是早期乳腺癌检

测的难点. 为了有效地检测出肿块区域, 国内外的众多学者进行了深入的研究和探索, 提出了基于滤波器的方法^[2], 基于形态学和统计分析的方法^[3], 根据肿块的特性建立的一些特殊模型^[4], 以及基于学习的检测系统^[5]. 这些方法通常通过分割得到感兴趣区域, 再提取其灰度、几何以及纹理等特征, 从而得到最终的检测结果. 但它们往往很难在检出率和假阳性率上得到很好的折衷.

Eltonsy 等人通过对肿块生长特性的研究, 提出了一种多同心层的肿块检测方法^[6], 根据肿块随其生长变化, 通常在核心区域具有较大灰度值, 并以此为中心向其边缘处灰度逐渐变暗的特性, 采用不同的灰度阈值进行分割, 来检测肿块区域. 为了去除背景和噪声的影响, Eltonsy 对图像进行了粒度化处理, 但是这种方法使一些

灰度与背景差别较小的微弱肿块区域易漏检,并且容易引起 Mosaic 效应,使提取到的区域形态学特征不准确,对于较亮的血管等组织,还有一定的增强作用.此外,多同心层方法采用的固定阈值,会使一些包含大量灰度信息的灰度级被合并,从而导致细节信息的丢失.

为了解决上述问题,本文提出了一种新的肿块检测方法,首先引入了形态学成分分析 (Morphological Component Analysis, MCA) 来得到原图的平滑分量;在此基础上,根据图像直方图的分布特性,设计了自适应阈值搜索方法;最后结合各个灰度层中区域的形态学特征,得到检测结果.与 Eltonsy 等人的方法相比,新方法具有以下优势:(1)MCA 方法能将原始图像中的平滑分量和纹理分量分离,有效保留了图像的主要灰度分布特性;(2)基于 MCA 方法得到的平滑图像不存在 Mosaic 效应,较好的保留了原始图像中各区域的边界信息;(3)MCA 方法能够将图像中的细节纹理去除掉,从而抑制了图像噪声以及血管等组织的影响,降低了假阳性率;(4)自适应阈值搜索方法能够自适应的确定灰度的分层阈值,避免了固定阈值可能造成的细节信息丢失.通过对乳腺 X 线图像的分析 and 对比实验结果表明,提出的方法具有良好的检测性能.

2 MCA 方法简介

Starck 等人在 2004 年提出了 MCA 方法^[7],若假定数据 s 是 K 个子元素的线性组合, $s = \sum_{k=1}^K s_k$, 其中每一个 s_k 都表示一种不同类型的子信号,则 MCA 模型可表示为在条件 $s = \sum_{k=1}^K \Phi_k \alpha_k$ 下求解:

$$\{\alpha_1^{opt}, \dots, \alpha_K^{opt}\} = \text{Arg} \min_{\{\alpha_1, \dots, \alpha_K\}} \sum_{k=1}^K \|\alpha_k\|_1 \quad (1)$$

这里 Φ_k 是一组基函数的集合或称为字典,不同的 Φ_k 在表示不同类型的子信号时能够分别得到最稀疏的解.

对于上述问题的求解,可以通过放宽式(1)的限制条件来得到近似解,即

$$\{\alpha_1^{opt}, \dots, \alpha_K^{opt}\} = \text{Arg} \min_{\{\alpha_1, \dots, \alpha_K\}} \sum_{k=1}^K \|\alpha_k\|_1 + \lambda \left\| s - \sum_{k=1}^K \Phi_k \alpha_k \right\|_2^2 \quad (2)$$

实际中,为了降低运算代价,MCA 方法只考虑具有快速变换 $T_k (\alpha_k = T_k s_k)$ 和重构算法 $R_k (s_k = R_k \alpha_k)$ 的字典 Φ_k . 那么式(2)就可以重新改写为关于信号 s_k 和其快速变换的形式:

$$\{s_1^{opt}, \dots, s_K^{opt}\} = \text{Arg} \min_{\{s_1, \dots, s_K\}} \sum_{k=1}^K \|T_k s_k\|_1 + \lambda \left\| s - \sum_{k=1}^K R_k s_k \right\|_2^2 \quad (3)$$

3 基于 MCA 的自适应检测系统

3.1 基于 MCA 的乳腺 X 线图像分解

乳腺 X 线图像可以看成是由平滑的灰度图像层与包括血管、噪声以及纤维等周期或细小局部纹理的图像层的近似线性组合,而后者会给检测造成很大的干扰.因此,将 MCA 引入乳腺 X 线图像的检测中,通过选择能够将这两个图像层区分开来的子字典,提取出检测所需要的平滑灰度图像层^[8].平稳小波变换 (Stationary Wavelet Transform, SWT) 和局部离散余弦变换 (Local Discrete Cosine Transform, LDCT),被分别选择作为表示平滑图像层和包含血管及背景噪声等图像层的子字典.这样就能够将乳腺 X 线图像中较易干扰肿块检测的血管及纤维等噪声去除,仅在包含了图像主要灰度分布的平滑图像层中进行检测.

如果用 S_{WT} 表示 SWT,用 L_{DCT} 表示 LDCT,相应的子图像表示为 s_{WT} 和 l_{DCT} ,根据 MCA 方法的表示形式,可以将本文的问题表示成为:

$$\{s_{WT}^{opt}, l_{DCT}^{opt}\} = \text{Arg} \min_{\{s_{WT}, l_{DCT}\}} (\|S_{WT} s_{WT}\|_1 + \|L_{DCT} l_{DCT}\|_1 + \lambda \|s - s_{WT} - l_{DCT}\|_2^2) \quad (4)$$

根据 Block-Coordinate Relaxation 方法^[7],就能够快速的计算出经过 MCA 分解后所得到的最优子图像 s_{WT} 和 l_{DCT} .

3.2 自适应阈值搜索

肿块是一种特殊的病变区域,早期的肿块面积较小,X 线图像上难以为肉眼所见.随着其不断扩散和生长,中心处由于 X 线难以透过表现为局部最亮的灰度,在其中心周围,则表现为由中心向其边缘处灰度逐渐减弱.由于肿块区域可近似表示为类椭圆区域,则其生长特性从 X 线图像上来看,接近一个二维的高斯区域,如图 1 所示.

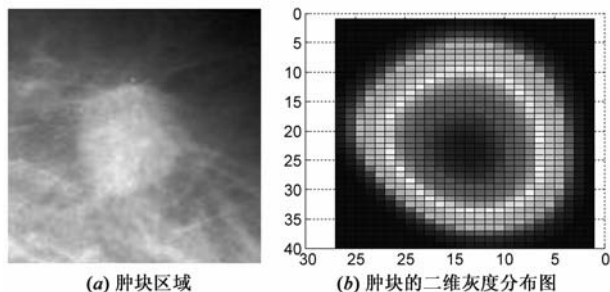


图1 肿块区域灰度分布特性

多同心层方法^[6]采用了训练得到的固定阈值,以灰度直方图的最大值为起始值,灰度每减小 0.05,对图像进行一次二值化,但在局部直方图像素数较大的灰度值处,可能会遗漏一些细节信息,如低对比度区域的同心层,从而导致了漏检和误检.

在直方图中像素数目较少的灰度值上,希望当前

阈值与下一灰度层的阈值间隔稍大一些,从而避免由于步长过小而分层过细,引入过多的假阳性区域;而在直方图中像素数目较大的灰度值上,希望间隔小一些,避免区域在下次阈值化后扩张过度,而使得肿块的多同心层特性消失,导致检测灵敏度下降.为此首先将有效的直方图段(即乳腺区域的直方图)进行局部归一化,再取直方图函数的反函数作为自适应间隔函数:

$$\Delta t = \frac{1}{\epsilon} \cdot e^{-(1+h_i)} \quad (5)$$

式中 h_i 为原图中有效区域的直方图, ϵ 为一常数,用来调整灰度间隔值在期望的范围中,文中取为 10.

图 2 为一幅乳腺 X 线图像的有效直方图,以及对应的自适应间隔函数曲线.可以看出,自适应间隔函数曲线能够很好的描述与原直方图曲线相反的走势,符合设计的初衷.这样就可以依据第 i 层的灰度阈值 T_i ,推算出第 $i+1$ 层的灰度阈值,即:

$$T_{i+1} = T_i - \Delta t \quad (6)$$

然而大部分乳腺 X 线图像的直方图都非常复杂,得到的灰度间隔函数曲线有多个波峰波谷,其中的波峰意味着对应的直方图灰度值较低,需要较大的灰度间隔来避免分层过细.因此,如果前一阈值与根据灰度间隔曲线得到的下一阈值之间含有波峰时,可以忽略不计.若两个阈值之间相对应的灰度阈值间隔曲线有波谷存在,就需要在波谷处设置一个阈值.

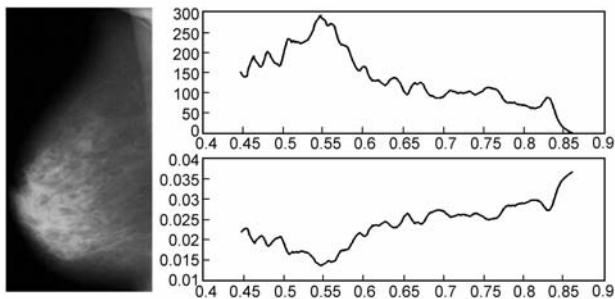


图2 乳腺X线图像及其有效直方图和自适应间隔函数曲线

为了探测到间隔曲线中的波谷位置,可以通过计算两个阈值间的多尺度距离差分值得来确认两个阈值间是否存在波谷.从最大灰度值开始,如果以上一阈值 T_a 为起始点,分别与它和下一阈值 T_b 之间的所有灰度间隔曲线上的值做差分,可以得到一组多尺度距离差分集合:

$$\{dt_i | dt_i = \Delta t_a - \Delta t_i, a \leq i \leq b\} \quad (7)$$

根据灰度间隔曲线的特性,不论前一阈值所对应的灰度间隔位于曲线的哪个位置,以此阈值为起始点对所有位于两阈值间的灰度间隔所做的差分,只会在波谷处取到最大值.因此,对上述集合进行排序,就可以得到需要额外增加的位于灰度间隔曲线波谷处的阈值:

$$T_{add} = \text{Arg max}_i (dt_i) \quad (8)$$

则最终的灰度分层阈值序列就由上述的阈值集合 $\{T_i, T_{add}\}$ 组成.

3.3 扩展的多同心层检测方法

根据 3.1 节中经 MCA 分解得到的平滑图像,以及 3.2 节中自适应阈值方法得到的各个灰度层的阈值序列,通过结合多同心层检测方法,就建立起了整个肿块检测系统(如图 3 所示).

形态学特征已广泛应用于医学图像中的目标检测^[9].因此,通过对各个灰度阈值下所得到的二值化图像中的独立区域提取形态学特征,来选择需要进行进一步同心层分析的区域^[6].根据肿块区域占有一定的面积、较为接近椭圆区域、以及与其最大外接矩形有一定的占空比这三个特征,可以设定如下的规则,初步筛选出各个灰度层上需要进一步分析的区域 R_i :

$$\{R_i | (F_{R_i}^{AR} \geq Th_{AR}) \& (F_{R_i}^{EC} \geq Th_{EC}) \& (F_{R_i}^{EX} \geq Th_{EX})\} \quad (9)$$

$$i \in S_n, 1 \leq n \leq N$$

其中 Th_{AR} 为区域面积所对应的阈值条件,根据肿块区域通常的面积范围,取为 50 至 9000 像素之间. Th_{EC} 为与区域最接近的椭圆对应的长短轴之比,取为 0.4. Th_{EX} 为占空比阈值,设为 0.26. 而 S_n 代表第 n 个灰度层, N 则表示总灰度层数.

肿块区域的灰度并不均匀,如果仅仅通过判断相邻灰度层间的灰度重心距离来确定其同心层的关系,无法确定有效的距离阈值.为此,将多同心层方法扩展为:若下一灰度层中的区域完全包含上一灰度层中的某个区域,并且其灰度重心距离不超过此区域的最大半径,则认为它们为同心层区域.

如果在这样相邻关系的灰度层中,能够找到满足包含关系的区域,就可以在此基础上,再向其相邻的阈值较低的灰度层中搜索能够包含此区域的区域,一直到在阈值较低的相邻层中无法获取满足一定形态学条件的同心层区域或没有满足包含关系的区域为止.对于满足第一轮筛选的各灰度层上的独立区域,如果第 m 层上的第 i 个区域 R_i^m ,包含于第 $m+1$ 层上相对应的同心区域 R_i^{m+1} 中,则有:

$$\{R_i^m | \forall R_i^{m+1}, \text{and } R_i^m \subset R_i^{m+1}, \text{where } R_i^m \in S_m, R_i^{m+1} \in S_{m+1}\} \quad (10)$$

上式为只有一个灰度同心层时的规则,依此类推,包含同心层的区域 R_i^m 可以以其下一层的同心区域 R_i^{m+1} 为基准,继续向下一灰度层寻找能够包含此区域的区域,直到没有同心区域存在为止:

$$\{R_i^m | \forall R_i^{m+n}, \text{and } R_i^{m+n-1} \subset R_i^{m+n}, \text{where } R_i^{m+n} \in S_{m+n}, n = 0, 1, \dots, N-1\} \quad (11)$$

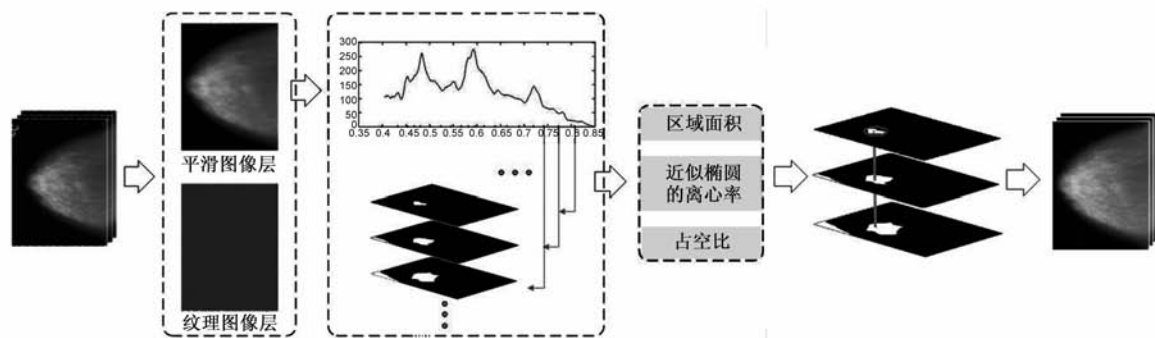


图3 肿块检测系统

则对于区域 R_i^n 来说,能够包含它的 n 个区域就代表着其同心层个数,当 $n \geq 1$ 时,就认为区域 R_i^n 可能为肿块区域,以此得到最终的检测结果。

4 实验结果分析

为了验证所提出的肿块检测系统的有效性,对来自南佛罗里达大学 DDSM (Digital Database for Screening Mammography) 数据库的 210 幅乳腺 X 线图像进行了测试实验^[10]。这些图像均为大小在 5000×3000 左右的 12 位灰度图像,并已由专家标记了病变区域的位置,因此可据此来测试检测系统的性能。本文所采用的 210 幅图像中包含 156 个恶性肿瘤及 60 个良性肿瘤。在进行实验之前,由于乳腺 X 线图像尺寸过大,在不影响检测的前提下,对原始图像进行了下 5 采样。

图 4 为一幅乳腺 X 线图像从 MCA 分析到最终检测

结果的处理过程。图 4(a)中曲线所包围的区域为专家所标记的肿块区域。在 MCA 处理后的平滑图像中,首先根据自适应搜索到的阈值序列,得到从高到低的各个灰度层,再根据提取到的符合一定形态学条件的相邻层独立区域间的包含关系及多同心层特性,得到了最终的检测结果(图 4(l))。

图 5 所示为两幅含有恶性肿瘤的乳腺 X 线图像。可以看出,在经过 MCA 处理之后,图像中的噪声和一些细节的纹理,如纤维和血管等都已经有效的去除或者弱化了,更有利于肿块的检测。图 5(c)两幅图像中的三角形标记即为本文的方法检测所得的可疑区域,可以看出在第一幅图像的检测结果中有一个假阳性区域,而第二幅图像检测结果最理想,仅检测到了肿块区域。图 5(d)所示为肿块区域的多个灰度同心层同时显示在图像中的结果,近似反应了肿块的生长特性。

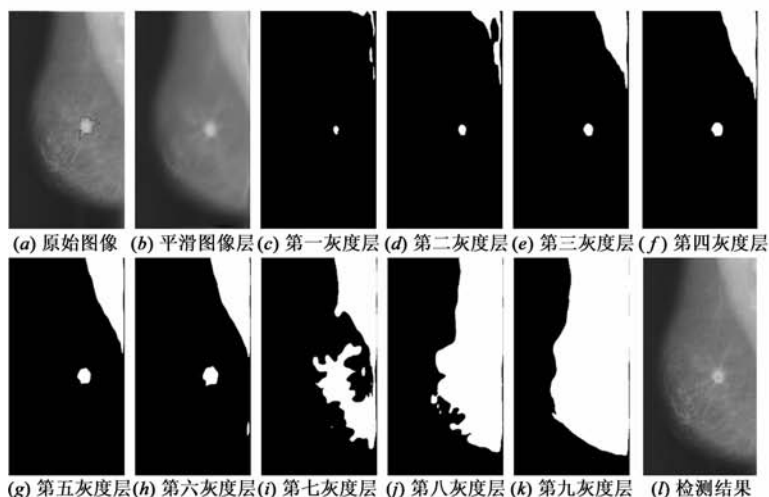


图4 肿块检测系统的检测过程

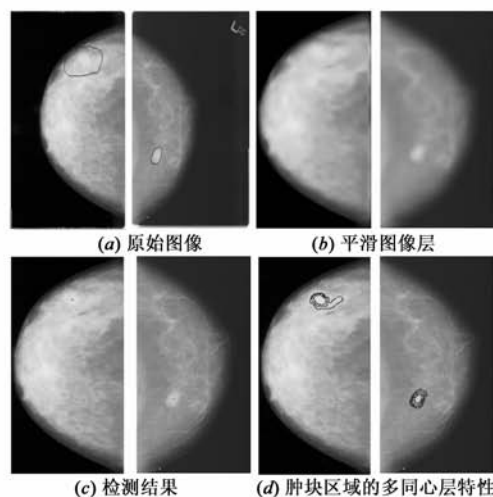


图5 乳腺X线图像的多同心层特性示例

图 6 则显示了图 5 中的两幅乳腺 X 线图像的有效区域,即除去黑色背景的乳腺区域的灰度直方图,图中 x 轴上的星型坐标点,代表自适应搜索方法得到的阈值。可以看出,本文提出的自适应阈值搜索方法在局部像素数分布较少的灰度值上,阈值间隔较大,而在局部像素数分布较多的灰度值上,阈值间隔较小,能够更为有效地分析到包含像素信息量较大的灰度值。

图 7 所示为一组乳腺 X 线图像采用不同检测方法所得到的检测结果,前两行为恶性肿瘤的检测结果,后两行为良性肿瘤的检测结果。其中,第一列为带有专家标记的原始图像,第二列则是 Duke 大学的 Eltonsy 等人提出的多同心层方法^[7]所得到的检测结果,虽然假阳性区域个数很少,但是存在严重的漏检现象。考虑到数据库对检测结果的影响,为了得到有意义的检测结果,

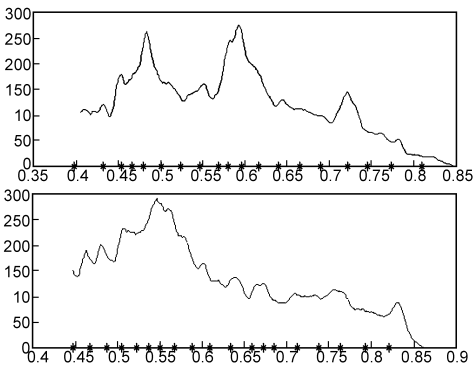


图6 两幅乳腺X线图像有效区域直方图及自适应阈值

对原始的同心层方法进行了参数调整,选取了与本文相同的形态学阈值和同心层数判断条件.第三列即为参数调整后的多同心层方法得到的检测结果,由图可见,对于一些内部灰度差异较小的肿块区域,无法检测到真正的肿块区域,而对于能检测到肿块区域的图像来说,其假阳性区域数量过大.第四列为引入了 MCA 方法但没有采用自适应搜索的检测结果,虽然假阳性区域的数目得到了抑制,但检出率也随之大大降低.最后一列是本文方法所得到的检测结果,在采用了自适应阈值搜索方法后,内部灰度变化较小的肿块区域也能够有效的检测到,并且假阳性区域个数并没有大幅增

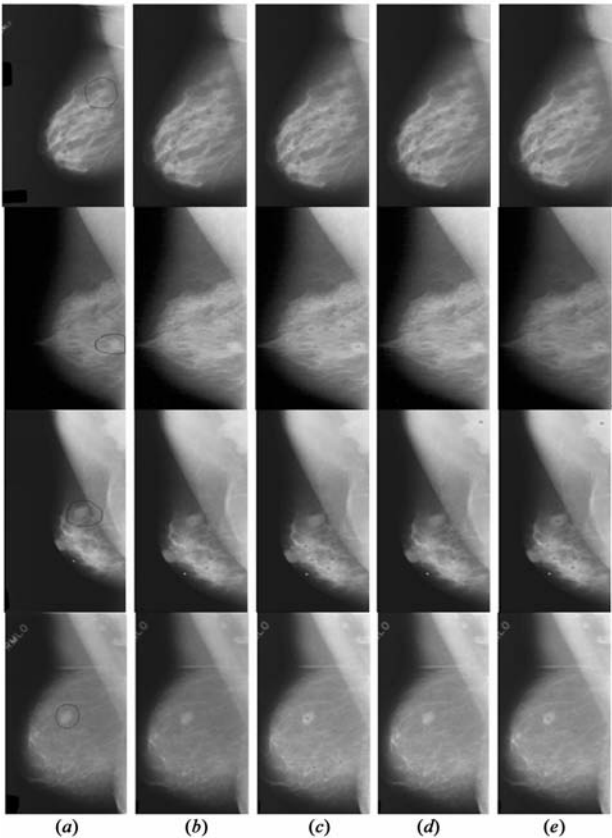


图7 乳腺X线图像的检测结果对比

(a) 原始图像; (b) 基于原始多同心层方法的检测结果; (c) 基于参数调整后的多同心层方法的检测结果; (d) 基于MCA的检测结果; (e) 基于MCA和自适应阈值搜索的检测结果

加.

为了得到具有一定统计意义的检测结果,对数据库中的 210 幅乳腺 X 线图像进行了测试实验.表 1 中分别给出了对应于图 7 中的四种方法所得到的最终测试结果,其中 TP (True positive) 代表检出率,而 FPI (False positives per image) 代表平均每幅图像中的假阳性区域个数.经过对比实验可以看出,本文提出的肿块检测系统在保证了检测灵敏度也即检出率的同时,假阳性区域的个数也得到了了一定的控制,使得改进后系统的检测性能有了较大的提高,能够更好地辅助临床医生进行诊断分析.

表 1 四种方法的检测结果比较

检测方法	恶性 (TP/FPI)	良性 (TP/FPI)	所有图像 (TP/FPI)
多同心层方法检测	51.9%/1.6	46.7%/1.6	50.5%/1.6
调整参数后的多同心层方法检测	92.5%/7.9	85.0%/7.3	90.7%/7.7
MCA 检测	72.4%/0.9	61.7%/1.1	69.4%/1.0
本文方法	95.5%/2.2	88.3%/2.7	94.8%/2.4

5 总结

为了能够有效地保留乳腺 X 线图像中病变区域的灰度分布以及形态学特征,并且同时抑制图像中的纤维、血管等组织对检测造成的影响,引入了 MCA 方法来对乳腺 X 线图像进行预处理,分解得到了图像的平滑部分.并在此基础上,分析了肿块区域的生长特性,提出一种基于直方图分布特性的自适应阈值搜索方法,自动地检测出了各个灰度阈值下符合一定形态学特征的独立区域.再根据相邻灰度层上独立区域间的包含关系,得到最终的检测结果.实验结果表明,本文的方法能够有效地检测出乳腺 X 线图像中的肿块区域,并且抑制了图像中大量假阳性区域的产生,有良好的检测性能.

但是对于个别纤维结构复杂的乳腺 X 线图像,还是存在假阳性区域较多的情况,而良性肿瘤的检出率也有待进一步提高.因此我们将继续深入研究更为有效的假阳性区域去除方法和良性肿瘤的特性,以进一步提高系统的检测性能.

参考文献

[1] 万柏坤,王瑞平,朱欣,等.SVM 算法及其在乳腺 X 片微钙化点自动检测中的应用 [J].电子学报.2004,32(4):587-590.

Wan B K,Wang R P,Zhu X,et al.Principles of SVM and Its application in micro-calcifications detection in mammogram [J].Acta Electronica Sinica,2004,32(4):587-590.(in Chinese)

- [2] H Kobatake, M Murakami, H Takeo, et al. Computerized detection of malignant tumors on digital mammograms [J]. IEEE Trans Med Imaging, 1999, 18(5): 369 – 378.
- [3] H Georgiou, M Mavroforakis, N Dimitropoulos, et al. Multi-scale morphological features for the characterization of mammographic masses using statistical classification schemes [J]. Artif Intell Med, 2007, 41(1): 39 – 55.
- [4] D Guliato, R M Rangayyan, J D Carvalho, et al. Polygonal modeling of contours of breast tumors with the preservation of spicules [J]. IEEE Trans Biomed Eng, 2008, 55(1): 14 – 20.
- [5] Y Wang, X B Gao. Mass detection algorithm based on support vector machine and relevance feedback [J]. Frontiers of Electrical and Electronic Engineering in China, 2008, 3(3): 267 – 273.
- [6] N H Eltonsy, G D Tourassi, A S Elmaghraby. A concentric morphology model for the detection of masses in mammography [J]. IEEE Trans Med Imaging, 2007, 26(6): 880 – 889.
- [7] J L Starck, M Elad, D L Donoho. Redundant multiscale transforms and their application for morphological component analysis [J]. Adv Imag Elect Phys, 2004, 132: 287 – 348.
- [8] J L Starck, M Elad, D L Donoho. Image decomposition via the combination of sparse representations and a variational approach [J]. IEEE Trans Image Process, 2005, 14(10): 1570 – 1582.
- [9] 刘海华, 陈心浩, 高智勇, 等. 基于形态学操作和模糊聚类技术的超声图像分割 [J]. 电子学报, 2007, 35(7): 1306 – 1312.
Liu H H, Chen X H, Gao Z Y, et al. Segmentation of ultrasound image based on morphological operation and fuzzy clustering technique [J]. Acta Electronica Sinica, 2007, 35(7): 1306 – 1312. (in Chinese)

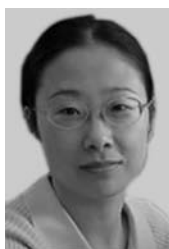
- [10] University of South Florida. Digital Database for Screening Mammography (DDSM). [DB/OL] <http://marathon.csee.usf.edu/Mammography/Database.html>, 2004.

作者简介



王 颖 女, 1981 年出生于陕西户县, 讲师. 于 2006 在西安电子科技大学获得硕士学位, 2010 年获博士学位. 目前主要从事模式识别以及影像处理的研究.

E-mail: ying_w81@gmail.com



李 洁 女, 1972 年出生于陕西西安, 工学博士, 西安电子科技大学教授, 分别于 1995 年、1998 年和 2005 年在西安电子科技大学获得学士、硕士和博士学位. 主要从事于计算智能、机器学习及图像处理等方面的研究.



高新波 男, 1972 年生于山东莱芜. 分别于 1994 年、1997 年和 1999 年在西安电子科技大学获得学士、硕士和博士学位. 1997 – 1998 年赴日本静冈大学交流学习, 2000 – 2001 年赴香港中文大学客座研究. 现任西安电子科技大学教授、博士生导师、学科带头人, IEEE 高级会员、中国图象图形学会常务理事, 陕西省图象图形学会副理事长. 目前主要从事影像处理、分析和理解、模式识别和机器学习等领域的研究.