

Top-hat 与 SVM 在乳腺微钙化点检测中的应用

马慧彬, 张海燕, 丁晓迪

MA Hui-bin, ZHANG Hai-yan, DING Xiao-di

佳木斯大学 信息电子技术学院, 黑龙江 佳木斯 154007

College of Information and Electronic Technology, Jiamusi University, Jiamusi, Heilongjiang 154007, China

E-mail: fxue1114@sina.com

MA Hui-bin, ZHANG Hai-yan, DING Xiao-di. Application of Top-hat and SVM in micro-calcification detection of mammograms. Computer Engineering and Applications, 2009, 45(20): 243-245.

Abstract: It is very meaningful for early diagnosis of breast cancer with detecting micro-calcifications in breast cancer. However, used methods of detecting micro-calcifications have shortcomings of high false positive. Micro-calcifications of mammograms are detected with the method of combining wavelet with Top-hat filter. And on this basis, it can farther detect the results of coarse calcifications with SVM, eliminate the false and retain the true. This method can reduce the rate of false positive greatly, on the base of not reducing the rate of true positive. Simulating experiment results show that the rate of detection calcifications achieves 98.46%, and the rate of false positive is less than 3.597%. The method can effectively extract micro-calcifications from complex background.

Key words: micro-calcifications; wavelet; Top-hat; Support Vector Machine(SVM)

摘要: 乳腺 X 线图像中微钙化点的检测对于乳腺癌的早期诊断非常有意义, 然而目前常用的钙化点检测方法普遍存在假阳性高的缺点。采用小波与 Top-hat 算子相结合的方法对乳腺图像进行钙化点粗检测, 并在此基础上, 用 SVM 对钙化点粗检测结果进一步甄别, 去假存真。这样做可以在基本不降低真阳性率的情况下, 大大降低假阳性率。仿真实验证明, 该方法的钙化点检出率达到 98.46%, 错检率仅为 3.597%, 说明该方法能够有效地从复杂背景中提取出微钙化点。

关键词: 微钙化点; 小波; Top-hat; 支持向量机

DOI: 10.3778/j.issn.1002-8331.2009.20.071 文章编号: 1002-8331(2009)20-0243-03 文献标识码: A 中图分类号: TP391.4

1 引言

近年来, 国内乳腺癌发病率较前几年明显增高, 已占妇女恶性肿瘤的首位, 并有逐渐上升的趋势。早期发现、早期诊断、早期治疗能够明显提高患者生存率并降低医疗费用^[1]。细小、颗粒状成簇的微钙化点是早期乳癌的一个重要表征, 约有 30%~50% 的乳腺恶性肿瘤伴有微钙化点。但由于钙化点大小不一、形状各异、分布多变, 可疑病变区域与其周围组织之间的强度差微弱等原因, 对于 X 线图像中微钙化点的检测是非常困难的。目前用于图像小目标识别与特征提取的方法很多, 主要方法有: 基于小波变换的微钙化点提取技术^[2]、基于形态学的目标提取技术^[3-4]、基于支持向量机的目标检测技术^[5-6]等。

但是单一方法进行钙化点检测, 普遍存在假阳性高的缺点, 结合小波与 Top-hat 算子的特点对 X 线图像进行钙化点的粗检测, 并在粗检测基础上, 用 SVM 对钙化点进一步甄别, 去假存真。这样做可以在基本不降低真阳性率的情况下, 大大降低假阳性率。

2 算法基本原理

小波变换用于微钙化点分割的思想源于小波变换的空间

频率特性。由于钙化点在乳腺 X 线影像中多呈细微颗粒状, 它与邻近像素相比具有高灰度值, 近似于函数中的奇异点和信号中的高频成分。当对影像 $f(x, y)$ 进行小波变换时, 若 $f(x, y)$ 在点 n_0 处有奇异性, 即在 n_0 处极大, 则使小波变换系数的振幅在该点急剧增大, 而在光滑点的小波变换系数则较小, 通过调整小波系数使小波的放大倍数(或高度)以及宽度发生变化, 可以与影像信号中不同大小的微钙化点进行匹配和放大。由此可见, 选用适当基函数的小波变换, 能够将数字影像分解成多种尺度成分, 将微钙化点位置相匹配的系数成分反应出来, 从而可以确定微钙化点的空间定位和改善提取准确率。

形态学滤波器是非线性滤波器, 它通过变换来局部地修改信号的几何特征, 开启和闭合运算可用于对几何特征的定量研究。从消除比背景亮且尺寸比结构元素小的结构角度来看, 开运算有些像非线性低通滤波器。但是开运算与阻止各种高频的低通滤波器不同, 当图像中的大小结构都有较高的空间频域时开运算只允许大结构通过而能除去小的结构。对一幅图像进行开运算可消除图中的孤岛或尖峰等过亮的点。闭运算对较暗结构的功能与开运算对较亮的结构一样, 它可将比背景暗且尺寸比结构元素小的结构除掉。Top-hat 算子是一种形态滤波器, 其

定义如公式(1):

$$OTH_{f,b}(x) = (f \ominus b)(x) \quad (1)$$

公式中 $f(x)$ 与 $b(x)$ 是定义在二维离散空间上的两个离散函数, 其中 $f(x)$ 为灰度图像, $b(x)$ 为结构元素, “ \ominus ” 为形态开运算。

从 Top-hat 定义可以看到, 本文是从原图像中减去开运算结构, 也即在开运算中消除的比背景亮且比结构元素小的结构, 可以在 Top-hat 操作的结果中显现出来, 基于微钙化点通常较周围组织亮且大小约为 0.1~1 mm, 故而可以通过选择适当的结构元素来实现对微钙化点的定位。在图像的形态分析中, 结构元素起着重要的作用。对于灰度图像来说, 结构元素可以是任何的三维结构, 常用的有圆锥、圆柱、半球或抛物体。模板尺寸总是奇数, 这样模板中心正好对应一个像素。

在对乳腺图像进行小波分解后, 可以实现目标信号和背景杂波的初步分离, 即近似图像中主要包括背景杂波, 目标能量主要分布在 3 个细节图像中。因此, 针对各个子带图像的差异, 对各子带图像中所包含的背景成份进行更精确的预测与估计, 可设计相应的结构元素对小波分解各图像分别进行形态学 Top-hat 变换来检测各子带图像中来自钙化点的信号能量, 然后再采用相应的小波逆变换滤波器对所得到的子带目标图像进行滤波和信号重建。

基于统计学习理论的支持向量分类器(SVC, Support Vector Classifier)是一种有监督学习结构风险最小化的二类分类器, 因为其优越的性能成为近年研究的热点。一些学者^[7-8]将 SVC 引入微钙化点检测中并改进了检测效果。在乳腺图像中钙化点主要表现为一系列较周围正常组织亮的小区域, 由于其半径小, 常常淹没于背景组织和噪声中, 这是导致假阳性高的原因。经典的 SVC 分类器总是假定测试样本属于两类之一, 然而在乳腺图像中往往存在许多种类的非目标样本, 并且当非钙化点样本数目不受限制时, 导致训练样本数量庞大; 另外支持向量数目巨大也将导致检测时计算量巨大。

为解决假阳性高与计算量大的问题, 为了让样本空间尽可能的小, 以便让同类样本之间具有更好的一致性, 本文对图像进行粗检测, 得到不能确定是否为真钙化点的粗选目标集, 支持向量机只对这个粗选集合进行处理。这时, 多数非目标样本已经被结合小波的 Top-hat 滤波器滤掉。这样做不仅可以减少用于支持向量机的训练样本和测试样本的数量, 而且可以提高支持向量机的推广性能。

3 钙化点检测算法

3.1 基于小波与 Top-hat 算子的钙化点粗检测算法

为了提取图像的微钙化点, 在对乳腺 X 片去噪增强等预处理的基础上, 将图像划分成若干 128×128 的子图像, 对每幅子图像依次进行处理。粗检测的基本做法是: 首先对子图像进行小波分解, 对 3 个高频图像选择合适的结构元素进行 Top-hat 滤波, 去除背景与假目标, 然后做小波重建再进行一次 Top-hat 滤波检测目标区域, 最后做阈值分割检测到目标。在基于小波与形态学算子相结合的微钙化点检测算法中, 检测的效果与小波基的选择和 Top-hat 滤波过程中结构元素的设计有关。

在乳腺 X 片中, 在不影响图像的质量和尖锐边缘的前提下进行钙化点的检测是至关重要的, 因此选用 Daubechies 小波^[9], 它是一种常用的正交小波, 实验数据表明, 钙化点检测效果较好的是 DB4。对图像进行 4 层小波分解, 生成一幅低频子图像

与三幅高频子图像。根据小波分解的三幅细节子图像性质, 定义 Top-hat 算子中的结构元素 b_1, b_1 定义为半径为 1(直径为 3), 高度为 50 的圆柱形结构。小波重构后图像中目标被加强, 而背景被抑制, 根据处理后图像性质, 定义 Top-hat 算子中的结构元素 b_2, b_2 定义为半径为 3(直径为 7), 高度为 65 的圆柱形结构。

用于目标检测的阈值分割算法中的门限 Y 取值如公式(2)所示:

$$Y = p + K * q \quad (2)$$

其中, p 为图像灰度均值; q 为标准差; K 是调整因子, 可用它来控制虚警率, 实际应用中 K 的取值范围为 $[-1, 1]$ 。

经实验发现, 按上述操作检测, 如果虚警因子 K 取值较高, 将大大影响钙化点的真阳性率, 如果虚警因子 K 取值较低, 则检测后的目标中, 有许多单独一个像素的假目标。为了解决这个问题, 采用二次滤波的方法, 在取较低 K 值的情况下, 再滤掉那些 8 邻域内无其他目标的单像素假目标。

小波与 Top-hat 相结合的乳腺 X 线图像钙化点检测算法过程如下:

- (1) 读入 128×128 子图像;
- (2) 对图像做 DB4 的 4 层小波分解;
- (3) 选择略大于钙化点的圆柱形结构元素 b_1 , 对小波分解后的高频图像 LH、HL、HH 做开 Top-hat 运算;
- (4) 对低频图像 LL 乘以系数 K_1 ($0 < K_1 < 1$), 并进行小波重构, 可以得到目标与背景差异非常大的图像;
- (5) 选择略大于钙化点的圆柱形结构元素 b_2 , 对重构的小波图像做开 Top-hat 运算, 从而去除平缓的背景;
- (6) 用阈值分割方法获得粗选目标;
- (7) 滤掉单独一个像素的假目标;
- (8) 输出粗检测目标集合 CS。

3.2 基于 SVM 的钙化点识别算法

一般认为, 特征越多, 描述图像越充分, 但是, 获取特征数据不仅需要大量的测量设备与时间, 样本存储占用的空间和数据处理的数量也急剧增长, 而且, 还有一些对分类来说是冗余的甚至是不相关的特征会影响分类的速度与分类器的泛化能力。如何在低维特征空间中描述图像是特征提取需要解决的问题。

对待分类的样本选取适当的特征, 假定特征维数为 M 。使用这些特征进行 SVM 训练, 在训练的过程中进行特征选择: 每次去掉 M 维特征中的一维, 然后看支持向量的数目, 若支持向量明显减少, 说明这个特征比较重要; 若无明显变化, 说明该特征是次要特征, 可有可无; 若明显增多, 则是无用特征, 应去掉。最后选择使支持向量最少的特征组合来训练 SVM^[10]。按照上述方法, 经过实验, 在钙化点识别算法中选取六种钙化点特征, 它们是: 区域对比度、面积、圆弧度、直方图均值、直方图方差、熵。微钙化点在图像中表现为灰度值较高的亮点, 区域对比度可以描述钙化点的这种局部灰度特性; 面积与圆弧度描述了钙化点的几何特性; 直方图均值表示微钙化点区域的平均幅度; 直方图方差表示测量幅度的局部活动性; 熵表示微钙化点区域图像灰度信息量的量度。

SVM 分类的基本过程是: 计算粗选目标的上述六种特征值并构造特征向量 Y ; 按照医生对真实钙化点的标记, 对样本进行分类, 实际钙化点为正样本, 对应输出值为 1, 假目标为负

样本, 对应输出值为-1, 构造输出向量 X ; 选取部分元素做为训练样本, 训练 SVC, 生成支持向量并记录该支持向量; 其余样本为测试样本, 可以对 SVM 分类性能进行测试, 从而得到最后确认的钙化点集合。

实验表明, 按照上述方法进行钙化点识别, 效果理想, 达到预期目的。图像 mdb245 中一幅含有钙化点部分的 128×128 子图像测试结果如图 1 所示, 其中原图像的实际钙化点已做标记。

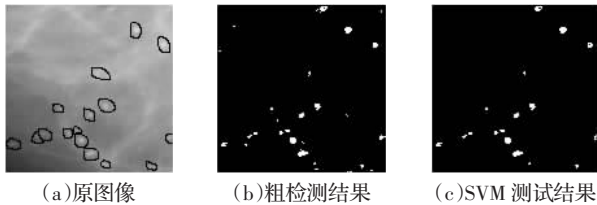


图 1 图像 mdb245 钙化点区域子图像处理结果

4 实验结果与分析

为了验证本文提出的钙化点检测方法能够较好地识别出钙化点, 进行了仿真实验。选取来自 MIAS 提供的小型数据库 Mini-MIAS 中的含有钙化点病灶的图像 25 幅、不含钙化点病灶的图像 200 幅作为实验图像。原图像大小均为 1024×1024 , 为方便处理, 每幅图像分割成若干 128×128 的子图像。

钙化点的粗检测工作要求检出率尽量高, 这样才能减少后面 SVM 分类结果的影响。同时, 在保证一定检出率的同时, 错检率越高, 后面的 SVM 分类的工作量越大。本文将小波变换与 Top-hat 滤波相结合的方法与 Robert 算子、Sobel 算子、子带小波滤波、单独的 Top-hat 滤波结果进行对比, 对比结果如表 1 所示。实验表明, 本文方法对于钙化点的粗检测结果好于其他普通方法, 不仅具有较高的检出率, 同时具有相对较低的错检率。

表 1 钙化点粗选算法检出率对比

指标	Robert	Sobel	子带小波	Top-hat	本文方法
检出率/(%)	98.13	97.36	95.12	98.04	99.08
错检率/(%)	33.59	29.46	18.77	24.43	13.63

对全部 128×128 的子图像按小波与 Top-hat 相结合的方法进行钙化点粗检测, 得到可疑钙化点集合。根据有经验的专家医生对钙化点的标记, 对 SVM 分类的测试结果进行统计。统计结果表明, 在 SVM 分类过程中, 当使用多项式核函数时, 真阳性率达到 99.37% 时, 假阳性率为 26.39%。

根据上述数据, 可计算出本文算法的总检出率达到 98.46% (=

99.08% \times 99.37%), 总错检率为 3.597% (=13.63% \times 26.39%)。

5 结束语

为了能够精确地检测出图像中的钙化点目标, 提出先进行钙化点粗检测, 再利用 SVM 对粗选结果进行分类的方法。粗检测应尽量多地检测到钙化灶, 本文根据小波的基本特性与 Top-hat 滤波器的形态学特性, 采用小波与 Top-hat 算子相结合的方法进行钙化点粗检测。这种方法在保证高检出率的情况下, 具有相对较低的错检率。为了能够提高 SVM 分类器性能, 根据钙化点的特点, 选取了六种钙化点特征参数, 它们分别代表了图像的灰度信息、空间分布信息和直方图信息。仿真实验的结果说明, SVM 分类器在基本不降低检出率的同时, 大大地降低了错检率。

实验数据表明, 结合了小波与 Top-hat 滤波粗检测方法的 SVM 钙化点识别算法在具有较高检出率的同时, 也具有很低的错检率, 说明这种算法对乳腺 X 线图像钙化点识别是有效的。

参考文献:

- [1] 张宏艳, 刘端祺, 范萍, 等. 乳腺普查工作初步报道及资料分析[J]. 解放军医学杂志, 2007, 32(6): 654.
- [2] 李树楠, 万柏坤, 马振鹤, 等. 基于小波变换的乳腺 X 线影像微钙化点感兴趣区域提取新技术[J]. 生物医学工程学杂志, 2005, 22(2): 360-362.
- [3] 董妍, 高新波, 王颖. 一种基于 Top-Hat 的乳腺图像中钙化点的检测方法[J]. 中国图象图形学报, 2006, 11(12): 1839-1843.
- [4] 王明泉, 杨静, 李志刚, 等. 薄壁焊缝 X 射线图像缺陷的自动提取与分割[J]. 计算机工程与应用, 2007, 43(33): 237-239, 245.
- [5] 胡正平, 吴燕, 张晔. 基于迭代顺序滤波子的空间约束的可拒识支持向量机微钙化点检测[J]. 电子学报, 2006, 34(2): 312-316.
- [6] 侯惠芳, 白莉媛, 刘素华. 基于支持向量机的图像边缘检测研究[J]. 计算机工程与应用, 2007, 43(18): 32-33, 58.
- [7] Wan Bai-kun, Wang Rui-ping. Principles of SVM and its application in micro-calcifications detection in mammogram[J]. Acta Electronica Sinica, 2004, 32(4): 587-590.
- [8] El-Naqa Issam, Yang Yong-yi. A support vector machine approach for detection of microcalcifications[J]. IEEE Transactions on Medical Imaging, 2002, 21(12): 1552-1563.
- [9] Daubechies. The wavelet transform, time-frequency Localization and signal analysis[J]. IEEE Transactions on Information theory, 1990, 36(5): 961-1005.
- [10] 邹斌, 张腊梅, 裴彩红, 等. 基于 SVM 的 POL-SAR 图像分类研究[J]. 遥感技术与应用, 2007, 22(10): 633-636.

(上接 207 页)

参考文献:

- [1] Höhne K H, Hanson W A. Interactive 3D segmentation of MRI and CT volumes using morphological operations[J]. Comp Assisted Tomogr, 1992, 16(2): 285-294.
- [2] 刘宾全. 基于遗传算法的肝 CT 序列图像的分割应用研究[D]. 长沙: 中南大学, 2005.
- [3] 陆仁枝, 宋志坚, 唐厚君. CT 序列图像分割的实现及分割结果的重建[J]. 计算机工程, 2003, 13(29): 152-154.
- [4] Wan S Y, Higgins W E. Symmetric region growing [J]. Image

Processing, 2003, 12(9): 1007-1015.

- [5] Chang Y L, Li X B. Adaptive image region-growing[J]. IEEE2 IP, 1994, 3(6): 868-872.
- [6] Adams R, Bischof L. Seeded region growing[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1994, 16(6): 641-647.
- [7] Mehnert A, Jackway P. An improved seeded region growing algorithm[J]. Pattern Recognition Letters, 1997, 18: 1065-1071.
- [8] Poh le R, Toennies K D. A new approach for model-based adaptive region growing in medical image analysis[C]//Proceedings of the 9th International Conference on Computer Analysis and Patterns, Warsaw, 2001: 238-246.