

中图法分类号: TP391 文献标志码: A 文章编号: 1006-8961(2011)02-0227-09

论文索引信息: 魏平, 李均利, 鲁东明, 陈刚. 基于乳腺钼靶 X 线图像等值面特征的乳腺区域阈值分割算法[J]. 中国图象图形学报, 2011, 16(2): 227-235

基于乳腺钼靶 X 线图像等值面特征的乳腺区域阈值分割算法

魏平¹⁾, 李均利²⁾, 鲁东明¹⁾, 陈刚¹⁾

¹⁾(浙江大学计算机科学与技术学院, 杭州 310027) ²⁾(宁波大学信息学院数字技术与应用软件研究所, 宁波 315211)

摘要: 乳腺钼靶 X 线图像中乳腺区域的分割可以帮助对图像进行深入分析和处理, 从而提高乳腺疾病的诊断准确率。提出一种能有效提取乳腺区域的算法。算法分析了乳腺钼靶 X 线图像等值面面积变化不连续的特征并将其用于分割阈值的精确计算。该算法使用基于扫描线的方法来获得含乳腺区域的连通区域, 比种子填充法效率更高。为了获得更纯粹的乳腺区域, 通过一些精细地处理对乳腺区域相连的未曝光图像边框作了剥离。实验结果表明算法在乳腺区域分割的精度和执行效率上都有更好的表现。

关键词: 乳腺 X 线图像; 乳腺区域; 区域生长; 阈值分割; 等值面

Thresholding segmentation method for breast region in mammogram based on isosurface

Wei Ping¹⁾, Li Junli²⁾, Lu Dongming¹⁾, Chen Gang¹⁾

¹⁾ (College of Computer Science and Technology, Zhejiang University, Hangzhou 310027 China)

²⁾ (Institute of DSP and Software Techniques, Ningbo University, Ningbo 315211 China)

Abstract: Breast region extraction of mammogram is a crucial step for further analysis and processing, and can improve diagnosis accuracy of galactophore diseases. An efficient algorithm of automatically extracting breast region is proposed. The discontinuity of the variety of areas of the isosurfaces is explored and used for computing the accurate segmentation threshold. Then the connected region containing the breast region is extracted by using a scan line based method, which is more efficient than seed-filling based methods. In order to obtain more pure breast region, the unexposed borders of the image connected to the breast region are also removed through a few refinement procedures. Experimental results show that the method has good performances both on the accuracy of breast region segmentation and executing speed.

Keywords: mammogram; breast region; region growth; thresholding segmentation; isosurface

0 引言

乳腺癌是最常见的癌症之一, 早发现、早治疗对于挽救癌症患者的生命非常重要。乳腺钼靶 X 线成像是早期检测乳腺癌的主要手段, 有效利用数字

乳腺钼靶 X 线图像提高诊断质量是目前的研究热点。

在数字乳腺钼靶 X 线图像的诊断中, 以下一些应用需要对乳腺钼靶 X 线图像中的乳腺区域进行分割。1) 高清晰的图像显示需要对乳腺钼靶 X 线图像中的乳腺组织像素进行精确分析; 2) 乳腺图像

收稿日期: 2009-04-28; 修回日期: 2009-09-25

基金项目: 国家自然科学基金项目(60672072); 浙江省自然科学基金项目(Y106505); 宁波市自然科学基金项目(2008A610012)。

第一作者简介: 魏平(1978—), 男, 浙江大学计算机科学与技术专业博士, 主要从事医学图像处理技术研究。

E-mail: weiping_zju@gmail.com。

的显示窗口布局需符合挂片协议,对图像中乳腺组织的分析为自动窗口布局提供依据^[1];3)排除背景区域会提高计算机辅助诊断的精确性并降低计算时间;4)乳腺钼靶 X 线图像的数据量很大,去除无诊断信息的背景区域,有利于图像压缩,提高存储和传输效率。

上述应用要求乳腺区域分割算法能够适应不同的图像特征并自动执行,作为软拷贝显示的预处理过程,算法在保证精确性的同时应该有较高的执行效率。

乳腺区域的亮度通常比背景区域的亮度高,阈值分割技术能够比较好地利用这一特征,且阈值分割算法计算复杂度低,因此在乳腺区域提取时普遍采用。一些分割算法,如动态轮廓线^[2]、分水岭方法^[3]等也普遍采用阈值计算结果作为初始分割。阈值选择的优劣影响到整个算法的精确性和收敛速度。在乳腺区域提取中采用的阈值分割算法主要有自适应阈值算法、Otsu 算法、迭代均值法、C 均值聚类法等。

自适应阈值分割技术假设乳腺区域和背景区域在直方图上的灰度信息有不连续性,确定对应乳腺区域和背景区域的直方图峰值,在两峰值之间的区域计算局部不连续性最大的灰度作为阈值^[4]。该方法对图像中部的乳腺区域能够较准确分割,但对图像顶部和底部的乳腺区域分割不准确。

Otsu 算法假设直方图是双峰分布,通过最大化类间方差来获得阈值。但是由于乳腺外围脂肪组织与背景区域的对比度不大,采用 Otsu 分割得到的通常是乳腺实质区域,而不是包含乳腺外围组织在内的整个乳腺区域^[5]。Sun 等人用自适应阈值分割法获得图像中间部分的乳腺轮廓,用 Otsu 算法获得乳腺实质区域。利用乳腺实质轮廓和乳腺外围轮廓距离近似为常数的性质,对图像上部分和下部分的乳腺轮廓线通过外插值估计获得了比较好的效果^[6]。

迭代均值法首先采用所有像素值的均值作为阈值初值,分别计算低于该阈值和高于该阈值的像素值均值,两类均值的平均值为新的阈值,直到连续两次迭代获得的阈值足够接近^[7]。Xu 等人用迭代均值法获得初始分割,然后用分水岭方法获得最终分割结果^[8]。

Lou 等人采用 C 均值聚类法获取阈值对图像进行初始分割,由于 C 均值聚类法分割使乳腺轮廓线倾向于乳腺组织内部,作者设计了精细的沿法线方

向的向外插值过程对乳腺轮廓线进行修正^[9]。

上述阈值分割算法都存在对先验知识利用不足的问题,Hu 等人提出了有监督的区域限制阈值分割算法,通过脑部 MR 图像和胸部 CT 图像验证了算法的性能,表明根据图像特征尽量限定阈值的取值范围能有效提高分割效果^[10]。

Shahedi 等人利用乳腺轮廓近似半圆的先验知识提出自适应阈值分割算法,将图像分为与原图像等宽的若干子图像,在每一块子图像上寻找使分割具有最小紧密性的阈值,获得了较准确的结果^[11]。

成像过程中乳腺钼靶 X 线图像边缘不能有效曝光,有些图像中会出现与乳腺区域相连的未曝光边框,如图 1 所示。大部分算法没有对这些并不属于乳腺区域的边框进行处理,Xu 等人对横向边框利用直线拟合与弹性线技术做了剥离,但没有处理与乳腺区域几乎完全相连的纵向边框,影响了乳腺区域的分割精度^[8]。

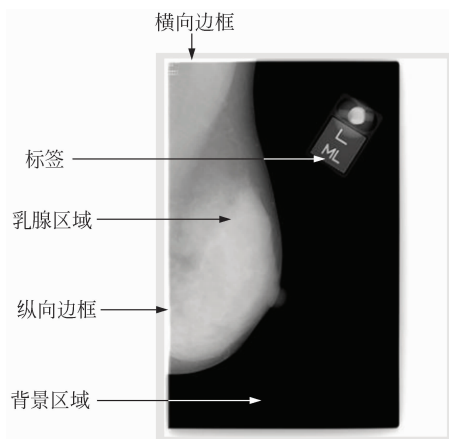


图 1 乳腺 X 线图像组成部分

Fig. 1 Compositions of digital mammogram

现有的乳腺区域阈值分割技术大都基于图像直方图的统计分布假设,没有充分结合乳腺钼靶 X 线图像的特点。通过对乳腺钼靶 X 线图像成像过程的分析,给出了一种提取乳腺区域的阈值分割算法。该算法以背景像素与乳腺组织像素的面积之比为特征,利用先验知识限定求解区域计算最优分割阈值。经过基于线扫描的连通区域分析和对未曝光边框剥离,获取精确的乳腺区域。

1 乳腺区域提取算法

提出的乳腺区域提取算法分 5 个步骤进行:

1) 中值滤波去除脉冲噪声; 2) 乳腺区域方向判定; 3) 分析图像等值面特征并计算分割阈值; 4) 连通性分析获取含乳腺区域在内的连通区域; 5) 剥离未曝光边框区域。

1.1 中值滤波

脉冲噪声对分割的影响比较大, 在对乳腺 X 线图像做处理前, 要对图像进行去噪。中值滤波器对脉冲噪声非常有效, 为了克服滤波效果依赖中值滤波器窗口大小的不足, 选用了变窗口的自适应中值滤波器对图像进行处理。该滤波器包括噪声检测和滤波两个过程, 根据噪声水平自动调节局部滤波窗口, 有很高的执行效率^[12]。

1.2 乳腺区域方向判定

约定胸壁所在一侧为乳腺方向, 如图 1 所示, 乳腺方向为图像左侧。如果将乳腺 X 线图像平均分为左右两幅子图像, 乳腺方向子图像的平均像素灰度会高于另一子图像的平均像素灰度, 这个规律普遍成立^[9]。

设乳腺 X 线图像大小为 $W \times H$, 等间距选择 N 行像素, 行坐标为 $\{n_1, n_2, \dots, n_N\}$, 其中 $n_1 \geq 0.1H$, $n_N \leq 0.9H$ 。这样选择是为了排除图像顶端和底端的未曝光边框影响, 计算左右两侧子图像抽样均值

$$\bar{I}_L = \left(\sum_{i=1}^N \sum_{j=1}^{\lfloor W/2 \rfloor} I(n_i, j) \right) / (N \cdot \lfloor W/2 \rfloor) \quad (1)$$

$$\bar{I}_R = \left(\sum_{i=1}^N \sum_{j=\lfloor W/2 \rfloor+1}^W I(n_i, j) \right) / [N \cdot (W - \lfloor W/2 \rfloor)] \quad (2)$$

式中 $\lfloor \cdot \rfloor$ 为下取整运算, 通过比较 \bar{I}_L, \bar{I}_R 的大小就可以判定乳腺区域的方向。若 $\bar{I}_L > \bar{I}_R$, 则乳腺区域在图像左侧, 否则乳腺区域在图像右侧。假定乳腺区域始终在图像的左侧, 对于乳腺区域在右侧的情况, 在做出判断后做图像左右镜像操作即可。

1.3 等值面特征分析与阈值计算

1.3.1 等值面特征分析

在乳腺钼靶 X 线图像成像时, 挤压板对患者乳腺组织挤压, 乳腺外围组织厚度逐渐降低, X 射线通过乳腺外围组织的衰减率逐步降低, 投影到成像平面上经光电转换后呈现较低的灰度值。

如图 2(a) 所示, X 射线在 A、B、C、D、E 的衰减率逐步降低。图 2(b) 所示图像上像素值与 X 射线的衰减率成正比, 衰减率较小的点呈现较低的像素值。随着像素灰度级降低, 具有相应灰度级水平的像素(位于同一个等值面)所占有的面积倾向于越大^[13], 如图 2(b) 所示的 $S_E > S_D > S_C > S_B > S_A$ 。其

中等值面 S_E 表示背景区域面积, 背景区域中 X 射线几乎没有衰减, 因此是完全曝光的, 背景区域等值面面积远远大于乳腺区域内的等值面面积, 即 $S_E \gg S_D$ 。

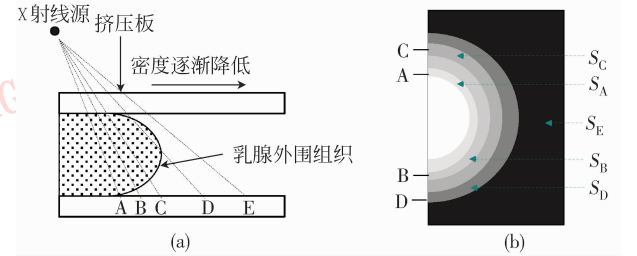


图 2 乳腺组织的不同密度

Fig. 2 Different intensity of breast tissue

若将像素位置视为地理坐标, 每个像素值视为该处的海拔高度值, 则图像本身就是一个 3 维地貌图。高 h 的平面称作值为 h 的等值平面, 像素值为 h 的点就位于这个等值平面。在乳腺钼靶 X 线图像拍摄过程中, 从乳腺实质区域到乳腺组织外围, 图像上等值平面占有的面积增大, 当越过乳腺区域与背景区域的临界区域时, 等值平面的面积会发生一个剧烈的增长, 迅速扩展到整个背景区域。图 2 所示是一个理想状况, 实际由于噪声的影响, 背景区域是由多个等值面组成的, 但从背景到乳腺组织的等值面面积变化的特征仍然存在, 为阈值选择提供了依据。

设 $\mathcal{X}_h = \{(i, j) | I(i, j) \geq h\}$, 即所有像素值不小于 h 的像素坐标集合, 它的补集 $\bar{\mathcal{X}}_h = \{(i, j) | I(i, j) < h\}$, $\mathcal{X}_h \cup \bar{\mathcal{X}}_h = I$, I 为图像中所有像素的坐标集合。令 $|A|$ 表示集合 A 中的元素个数。则 $|\mathcal{X}_h|$ 表示不小于 h 的等值面在图像中所占有的面积, $|\bar{\mathcal{X}}_h| = |I| - |\mathcal{X}_h|$ 。

令 S_h 表示值为 h 的等值面面积, 亦即像素值为 h 的像素个数。若存在阈值 T 将乳腺钼靶 X 线图像准确分为两部分: \mathcal{X}_T 为乳腺区域, $\bar{\mathcal{X}}_T$ 为背景区域。则在 T 的 δ ($\delta > 0$) 邻域中, $S_{T-k} \gg S_{T+k}$ ($1 \leq k \leq \delta$) 普遍成立。

当 T 减小 δ 时, $|\mathcal{X}_{T-\delta}| = |\mathcal{X}_T| + \sum_{k=1}^{\delta} S_{T-k}$, 相应地有 $|\bar{\mathcal{X}}_{T-\delta}| = |\bar{\mathcal{X}}_T| - \sum_{k=1}^{\delta} S_{T-k}$, $|\mathcal{X}_{T-\delta}|$ 比 $|\mathcal{X}_T|$ 有明显增长, $|\bar{\mathcal{X}}_{T-\delta}|$ 比 $|\bar{\mathcal{X}}_T|$ 明显减少。当 T 增加 δ 时, $|\mathcal{X}_{T+\delta}| =$

$|\chi_T| - \sum_{k=1}^{\delta} S_{T+k}, |\bar{\chi}_{T+\delta}| = |\bar{\chi}_T| + \sum_{k=1}^{\delta} S_{T+k}, |\chi_{T+\delta}|$ 只比 $|\chi_T|$ 只有微弱的减少, $|\bar{\chi}_{T+\delta}|$ 只比 $|\bar{\chi}_T|$ 有微弱的增加。于是在 T 的 δ 左邻域 $\{h | T - \delta \leq h \leq T\}$, $|\chi_h|$ 和 $|\bar{\chi}_h|$ 变化比较剧烈; 在 T 的 δ 右邻域 $\{h | T \leq h \leq T + \delta\}$, $|\chi_h|$ 和 $|\bar{\chi}_h|$ 变化相对平缓。

利用 $|\chi_h|$ 和 $|\bar{\chi}_h|$ 此消彼长的关系, 如下定义的函数 $f(h)$ 可更明显地突出上述变化规律。

$$f(h) = |\bar{\chi}_h| / |\chi_h| \quad (3)$$

易知 $f(h)$ 为单调递增函数, 且在 T 的左侧有较大的变化速率, 在 T 的右侧变化速率平缓。图 3(a) 给出了由图 1 的乳腺钼靶 X 线图像获得的 $f(h)$ 曲线。

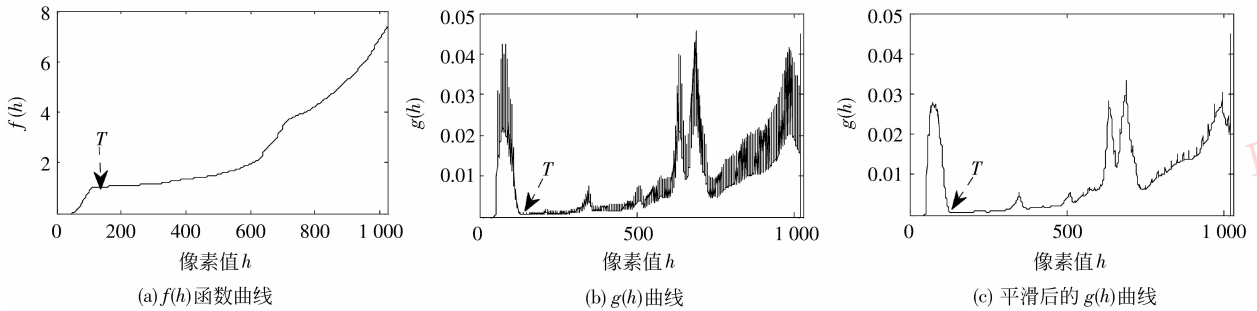


图 3 函数曲线图

Fig. 3 Function curves

函数 $g(h)$ 有多个峰值, 如在乳腺实质和脂肪区域的过度区域, 等值面也会发生较大变化, 加上 $g(h)$ 函数对这种变化的放大作用在高灰度区域表现得更为明显, 因此 $g(h)$ 最高峰值的右侧谷底可能不是乳腺区域与背景区域的分割阈值。我们利用两条先验知识缩小最优阈值 T 的取值区间, 然后在这个区间内计算最优阈值。

先验知识 1 对于乳腺钼靶 X 线图像, 背景区域和乳腺区域在图像中占据的比例有一个合理的范围, 该范围可根据具体的成像设备和操作合理估计 (如背景区域一般不小于整幅图像的 10%, 乳腺区域不大于整幅图像的 90%)。背景区域和目标区域的面积之比 $f(h)$ 应在一个合理的值域 $[\alpha, \beta]$ 之内。 $f(h)$ 是一个单调递增函数, 可以通过反函数计算出对应的定义域区间 $D_1 = [LH_1, RH_1], LH_1 = f^{-1}(\alpha), RH_1 = f^{-1}(\beta)$ 。

先验知识 2 在乳腺钼靶 X 线图像成像过程中, 图像的像素密度范围大部分用于表现乳腺图像

的特征, 以期获得好的组织对比度, 背景像素占用的灰度范围只是整个灰度范围的小部分, 不超过总像素范围的 $1/\gamma$ (如 $1/3$), 背景和目标的分割阈值将落在低灰度范围内 $D_2 = [LH_2, RH_2]$, 令 V_{min} 为图像的最小像素值, V_{max} 为图像的最大像素值, 则 $LH_2 = V_{min}, RH_2 = V_{min} + (V_{max} - V_{min})/\gamma$ 。

$$g(h) = |f'(h)| + |f''(h)| \quad (4)$$

$g(h)$ 在 $h = T$ 的左侧有很高的函数值, 形成一个峰值, 而在 $h = T$ 的右侧函数值就要小得多, 阈值 T 位于 $g(h)$ 峰值右侧的谷底。

1.3.2 分割阈值计算

设图像的概率密度函数 (直方图) 为 $H(i) = S_i, |\chi_h| = \sum_{i=1}^{h-1} H(i), g(h)$ 很容易由式 (3) (4) 计算得到, 其中一、二阶微分采用差分运算进行。直接计算获得的 $g(h)$ 函数曲线有很多的毛刺, 见图 3(b), 采用均值滤波器对其连续性进行改善后的曲线见图 3(c)。

最优阈值所在区间为上述两个区间 D_1, D_2 的交集: $D = D_1 \cap D_2 = [LH, RH]$ 。在区间 D 中寻找 $g(h)$ 的峰值, 记为 L_{peak} , 通过统计方法寻找峰值右侧的谷底作为最优分割阈值 T 。具体做法如下:

1) 对区间 $[L_{peak}, RH]$ 的 $g(h)$ 值的分布进行统计, 选择 $g(h)$ 的概率 p 分位点 h_0 , 使得 $P(g(h) \leq g(h_0)) = p$, 这里取 $p = 0.2$ 。

2) $T = \underset{h}{\operatorname{argmin}} \{g(h) | h \in [h_0, h_0 + len]\}$, 即在 $[h_0, h_0 + len]$ 中寻找使得 $g(h)$ 最小值以获得阈值 T, len 为选定的一个区间步长范围。

上述方法从 $g(h)$ 峰值处向右搜索, 找到接近谷底的一个较低值, 然后在该值右侧的小区间内寻找

的特征, 以期获得好的组织对比度, 背景像素占用的灰度范围只是整个灰度范围的小部分, 不超过总像素范围的 $1/\gamma$ (如 $1/3$), 背景和目标的分割阈值将落在低灰度范围内 $D_2 = [LH_2, RH_2]$, 令 V_{min} 为图像的最小像素值, V_{max} 为图像的最大像素值, 则 $LH_2 = V_{min}, RH_2 = V_{min} + (V_{max} - V_{min})/\gamma$ 。

局部最小值。与直接寻找峰值右侧的最小 $g(h)$ 值相比,这样的方法具有较好的鲁棒性,避免了由于 $g(h)$ 的波动造成的误差。取 $len = 2(V_{max} - V_{min})/256$, 图像显示时通常要将像素值范围投影到 256 个灰度显示区间, 阈值在显示图像中的灰度波动不超过一个单位。

1.4 基于线扫描的连通区域分析

获得阈值后,要对图像进行连通性分析获得乳腺区域。已有算法都采用 4 邻域或 8 邻域种子填充法选择最大连通区域作为乳腺区域,这类方法依赖于种子点的选取,且有较高的计算复杂度。如 4 邻域填充算法采用栈结构存储,出入栈总数为像素个数的 4 倍,需要的栈空间很大,反复的出入栈操作也花费大量计算时间。

如果从胸壁一侧开始,对图像进行列扫描,随着扫描线向乳腺外围方向移动,每一列上属于乳腺区域的像素点将逐渐减少,当没有属于乳腺区域的像素点时,表明扫描线已移出乳腺区域。据此分析,采用基于线扫描的区域增长方法来对连通的乳腺进行提取。

设在一条扫描线上,由大于阈值 T 的连续像素组成的第 i 条线段记为 $Lseg_i$,当前扫描线上属于乳腺区域的线段的集合记为 $CurLS = \{Lseg_1, Lseg_2, \dots, Lseg_n\}$,前一扫描线上所有属于乳腺区域的线段的集合为 $PreLS$ 。 BrA 表示乳腺区域的像素点集合,由各扫描线上的线段集合组成。

程序采用链表结构实现,整个区域生长过程如下:

1) 初始化 $Lseg_1$ 为图像第一列像素, $PreLS = \{Lseg_1\}$, $BrA = \{PreLS\}$,从图像第二列开始扫描 ($ScanIndex = 2$)。

2) 初始化当前线段集合为空 $CurLS = \emptyset$ 。

3) 在当前扫描图像列中从上到下,寻找一条线段 $Lseg_i$,若未找到新的线段,转 5)。

4) 若 $Lseg_i$ 与集合 $PreLS$ 中的某一线段连通,则 $CurLS = CurLS \cup \{Lseg_i\}$,转 3)。

5) 若 $CurLS \neq \emptyset$, $BrA = BrA \cup CurLS$, $ScanIndex = ScanIndex + 1$, $PreLS = CurLS$;若 $CurLS = \emptyset$,程序结束。

6) 若 $ScanIndex \leq W$,转 2);否则 $ScanIndex > W$,程序结束。

第 4) 步中两条线段 $Lseg_1$ 和 $Lseg_2$ 连通是指:两线段像素点的行坐标有重叠部分。设 $Lseg_1$ 的两端

点像素行坐标为 $\{L_1y_1, L_1y_2 \mid L_1y_2 \geq L_1y_1\}$, $Lseg_2$ 的两端点行坐标为 $\{L_2y_1, L_2y_2 \mid L_2y_2 \geq L_2y_1\}$,如果 $L_1y_2 \geq L_2y_1$ 且 $L_2y_2 \geq L_1y_1$,则 $Lseg_1$ 和 $Lseg_2$ 是连通的。

上述方法最多扫描一遍图像,线段的连通性判别采用端点坐标的比较操作,有较高计算效率,有助于提高乳腺区域提取算法的执行效率。

图 4 中 (a)(c) 分别是轴位、侧位的乳腺钼靶 X 线图像, (b)(d) 是对应的经连通性分析后的图像,白色是乳腺区域连通部分,背景灰度保持不变。至此获得的是包括乳腺组织及未曝光边框的连通区域,需要对这些边框进行剥离以提取纯粹的乳腺区域。

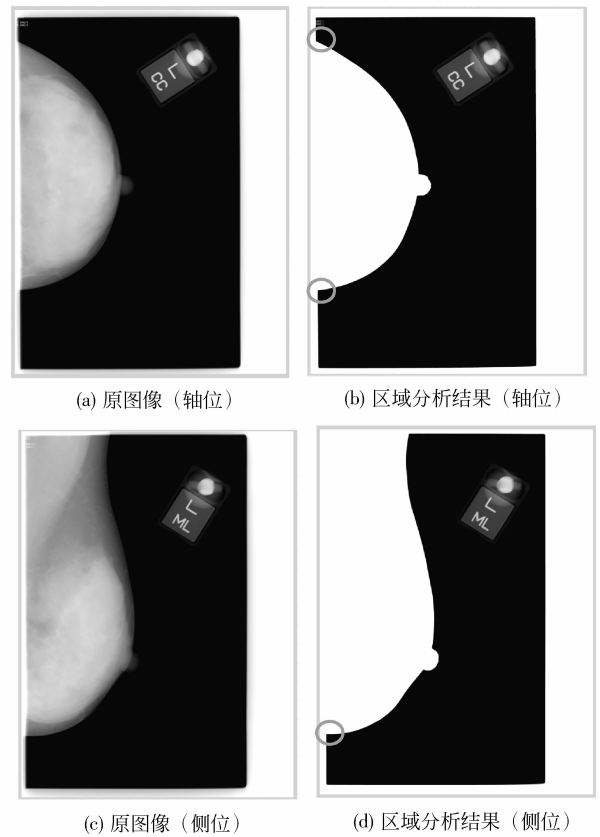


图 4 原图像与区域分析结果

Fig. 4 Original and result after region analysis

1.5 未曝光边框剥离

欲剥离的边框包括图像头部和底部的横向边框、图像左右两侧的纵向边框,对这两类边框分别进行剥离。

1.5.1 横向边框剥离

区域连通性分析过程中,连接乳腺区域的顶部边框位于每条扫描线上第一条线段,底部边框则位

于每条扫描线上最后一条线段。图 5 以图像列坐标为横轴, 扫描线上第一条线段的长度为纵轴。图 5 (a) (b) 分别表示图 4 (b) (d) 中轴位, 侧位图像每条扫描线第一条线段的长度。由于横向边框与两侧纵向边框相连, 图 5 两侧都有与图像高度接近的峰值, 中间最低的部分表示横向边框的长度。首先

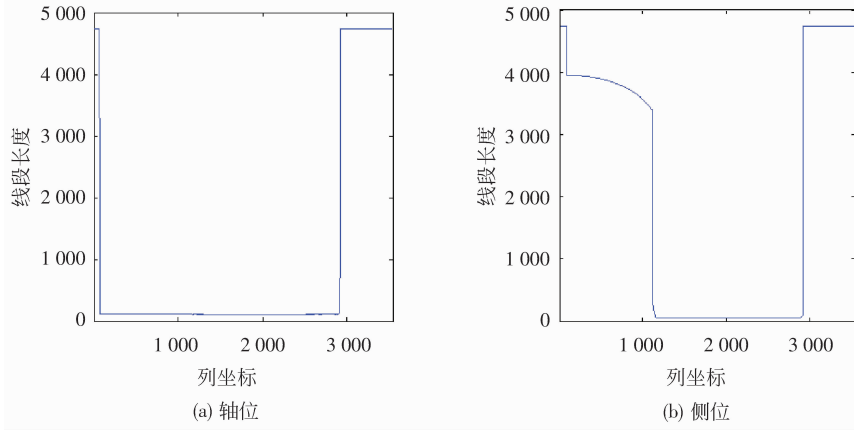


图 5 第一条线段长度

Fig. 5 Length of the first line segment

实际上, $\{l_i\}$ 只有有限个数值, 只要选择整数阈值, 使得两类的算术平均值相差最大即可完成分类, 即选择阈值:

$$T_l = \underset{t}{\operatorname{argmax}} \{ \operatorname{mean}\{l_i \mid l_i > t\} - \operatorname{mean}\{l_j \mid l_j \leq t\} \} \quad (5)$$

通过对 $\{l_i\}$ 的有限个数值的统计, 只需要数十次加法和乘法就可以完成聚类。图像底部的边框位于每条扫描线上最后一条线段, 也通过这样的聚类过程获得属于横向边框的线段。

从 **BrA** 中剔除属于横向边框的线段, 因为剔除线段会对区域的连通性造成影响, 因此需要对 **BrA** 的连通性进行考察, 剔除那些不与前面扫描线中任一线段相连接的线段, 即需要对 **BrA** 中每一条线段进行连通性检查。

通过这个过程, 绝大部分的横向边框及乳腺区域相反一侧的纵向边框被剔除, 但部分残余的横向边框连接了纵向边框, 只有断开图像左上角及左下角乳腺区域与纵向边框的连接才能彻底剥离 (见图 4 (b) (d) 中圆形虚线区域)。

左上角和左下角的乳腺轮廓线呈单调曲线, 分析发现指数函数 $y = e^{ax+b}$ 能较好地近似。其中 y 表示图像行坐标, x 表示列坐标。进行对数变换后得到 $\ln(y) = ax + b$, 采用最小二乘法对参数 a, b 进行

通过聚类获得属于横向边框的线段, 然后从 **BrA** 中剔除。

记扫描线第一条线段的长度序列为 $\{l_1, l_2, \dots, l_w\}$, l_i 为整数, i 为扫描线序号, w 为图像宽度。采用聚类算法将线段长度分为两类, 其中数值较低的一类是横向边框区域, 应该予以剔除。

估计, 完成对左上角和左下角的向外插值计算。

图 6 (a) (b) 用二值图像给出了图 4 (b) (d) 的处理结果, 该图中背景区域用纯黑色显示。可看到轴位图像中的横向边框都已被剥离, 在侧位图像中, 与乳腺区域直接相连的横向边框仍有残余 (图 6 (b) 中虚线标出)。采用最小二乘法拟合横向边框的分界线, 使边框与乳腺区域分离^[8]。

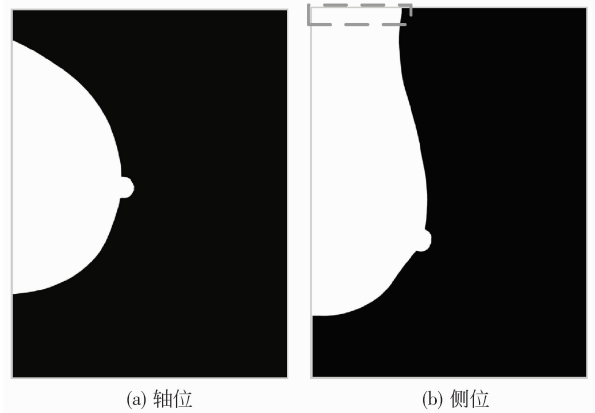


图 6 剥离横向边框

Fig. 6 Peeling off the horizontal border

1.5.2 纵向边框剥离

未曝光的边框通常有很高的亮度, 左侧边框区域与右侧乳腺区域之间的均值有比较大的差异。

图 7(a)是部分纵向边框的示意图, (b)为 (a)上虚线表示的行像素值, 边框区域亮度比乳腺区域高, 在边框与乳腺区域的交接处, 像素值有比较明显的台阶。设边框的宽度不超过 L 像素(可统计得出或根据经验设定), 在原图像数据上采用最大类间均差法获得边框的分界点。即遍历第 i 行中列坐标位于 $[0, L]$ 范围的每个像素点, 使得左右两侧像素均值

之差为最大者即为分割点 P_i 。

$$P_i = \arg \max_{1 \leq k \leq L} \left\{ \left| \left(\sum_{j=1}^k I(i, j) / k \right) - \left(\sum_{j=k+1}^L I(i, j) / (L - k) \right) \right| \right\} \quad (6)$$

逐行计算的分界点组成的分割边界线不连续, 需进行简单均值滤波处理, 图 7(c)为处理结果。

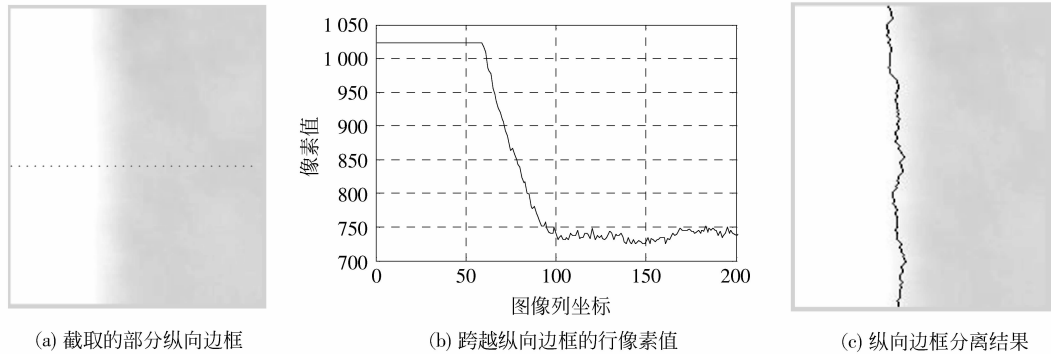
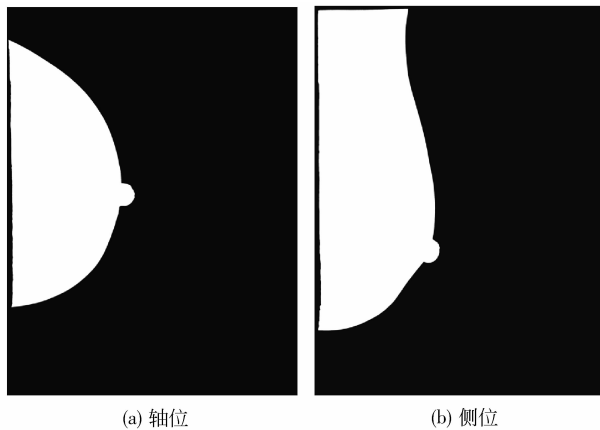


图 7 部分纵向边框

Fig. 7 One segment of vertical border

图 8 给出图 4(a)(c)中两幅原始图像的最终乳腺区域提取结果, 乳腺区域为白色区域, 背景为黑色区域。

床诊断需要, 轴位图像和侧位图像各占一半。



(a) 轴位 (b) 侧位

图 8 钼靶图像中的乳腺区域
Fig. 8 Breast region of mammogram

在医生指导下人工提取每幅图像的乳腺区域, 分割误差为自动分割和人工提取的乳腺区域的不匹配像素数目, 不匹配像素分为漏报像素和误报像素两种情况。

像素点在手工抽取的乳腺区域内, 不在自动抽取的乳腺区域内, 称为漏报像素 (MRP), 漏报像素占手工分割乳腺区域像素的比例称为漏报率 (MR)。不在手工抽取的乳腺区域内, 但在自动抽取的乳腺区域内的像素称为误报像素 (FRP), 误报像素数目占实际背景区域像素数目的比例称为误报率 (FR)。单纯的漏报率和误报率都不能说明算法的准确性, 分割误差定义为两种错误的和^[9]。

图像所有像素点数目记为 TP, 手工分割的乳腺区域像素点数目为 TBP, 则有

$$MR = \frac{MRP}{TBP} \cdot 100\%, FR = \frac{FRP}{TP - TBP} \cdot 100\% \quad (7)$$

$$Err = MR + FR \quad (8)$$

本文算法和最近 Shahedi 等人提出的自适应阈值分割算法作了比较。该方法将图像分为与原图像等宽的若干子图像, 利用乳腺轮廓近似半圆的性质, 认为最恰当的阈值分割得到的乳腺轮廓线具有最小的紧密性 (compactness)。紧密性定义为 2 维封闭

2 实验结果与评价

实验采用 20 幅乳腺钼靶 X 线图像来自浙江省邵逸夫医院, 由富士钼靶 X 线机采集。图像分辨率为 $3\,540 \times 4\,740$, 像素值为 10 bit 深度, 每像素 $50 \mu\text{m}$, 采用 CR 技术进行摄片数字化, 清晰度较好, 满足临

图形周长的平方与其所围面积的比值,2 维图形中,圆具有最小的紧密性。算法在每一块子图像上反复进行阈值选择和区域生长过程,选择具有最小紧密性的分割为最终结果,迭代计算,计算复杂度较高^[11]。

图 9 给出了两种乳腺区域提取算法的误差。表 1 给出了两种算法的平均分割误差与计算时间。

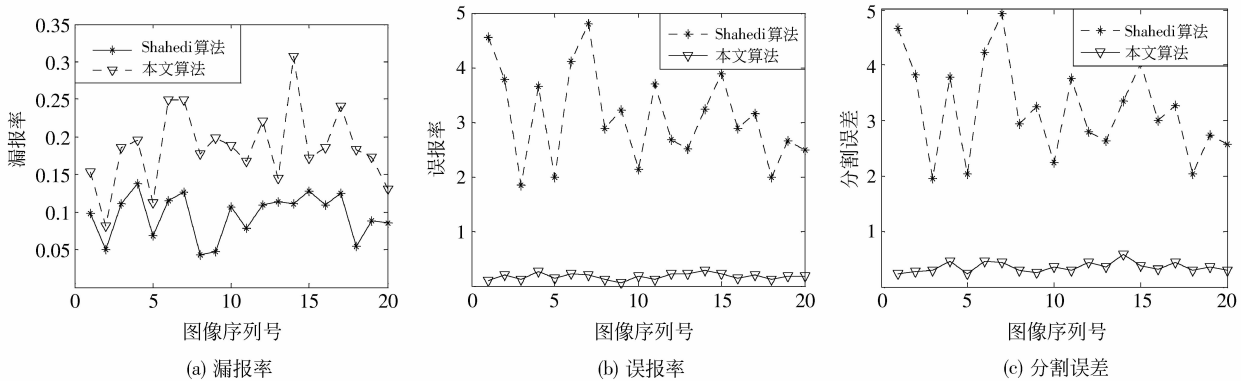


图 9 乳腺区域提取算法误差

Fig. 9 Errors of breast region segmentation

表 1 平均分割误差和平均执行时间

Tab. 1 Average error and executing time

	MR/%	FR/%	Err/%	每幅图像平均执行时间/s
本文算法	0.18	0.169 6	0.355 2	1.26
Shahedi 算法	0.095	3.107 5	3.202 5	4.72

本文算法大部分计算复杂度低,计算复杂度较高的区域生长过程采用了基于扫描线的连通性分析方法,主要的运算量转换为线段端点的比较判定。Shahedi 算法需迭代进行区域生长及最大连通区域选择过程。实验在 PC 机上完成,配置为 Intel Core 2 Duo E4500 CPU, 2.20 GHz。在 VC++ 6.0 环境下编译程序,本文算法每幅图像平均执行时间为 1.26 s, Shahedi 算法的平均执行时间为 4.72 s。乳腺区域提取通常作为预处理过程,本文算法能更好地适应实时应用需求。

3 结 论

乳腺钼靶 X 线图像在乳腺疾病的诊断中有重要的作用,提取乳腺区域是计算机辅助诊断(CAD)的重要预处理过程。然而,准确地提取出乳腺区域

Shahedi 算法倾向于将更多的像素划分到乳腺区域中,对边框区域没有剥离,因此漏报率较小。但不属于乳腺区域的一些像素点被误判定,出现较高误报率。从图 7 可看到,边框与乳腺区域的分界比较模糊,使得人工划分的分界线和自动计算的分界线较容易出现偏差,造成本文算法漏报率增高。从图 9(c)的综合误差来看,本文算法有更高的分割精度。

并不是一个简单的工作,涉及比较繁琐的技术处理。本文提出了一种快速自动提取乳腺钼靶 X 线图像中乳腺区域的阈值分割算法。该算法利用乳腺钼靶 X 线图像的等值面特征计算分割阈值,通过基于扫描线的连通性分析获得乳腺连通区域,并对未曝光边框进行了剥离,算法有较高的精度。虽然在乳腺区域的提取过程中,涉及较多计算步骤,但每个步骤的计算都具有较低的计算复杂度,因此算法也满足实时应用需求。

图像质量对分割精度产生很大的影响,在分割前采用适当的滤波技术是有帮助的,一些针对乳腺图像的增强处理技术,如增强外围组织对比度的乳腺图像组织均衡技术,有助于提高乳腺区域提取的精度。本文实验采用的图像有较好的清晰度和对比度。但一些老型号的 CR 机器拍摄的乳腺 X 线图像质量不高,特别是对于背景区域有很大噪声的图像,阈值技术可能在某些图像中的分割效果不理想,这些是需要进一步研究的。

参考文献 (References)

- [1] Hemminge B M. Soft copy display requirements for digital mammography[J]. Journal of Digital Imaging, 2003, 16(3):

- 292-305.
- [2] Wirth M A, Stapinski A. Segmentation of the breast region in mammograms using snakes [C] //Proceedings of the 1st Canadian Conference on Computer and Robot Vision. Ontario, Canada: Canadian Image Processing and Pattern Recognition Society, 2004: 385-392.
- [3] Kang Wei, Wang Guangzhi, Ding Hui. Segmentation of the breast region in mammograms using watershed transformation[C] //The 27th Annual International Conference of IEEE Engineering in Medicine and Biology Society Engineering. Washington, US: IEEE CS Press, 2005: 6500- 6503.
- [4] Ojala T, Näppi J, Nevalainen O. Accurate segmentation of the breast region from digitized mammograms [J]. Computerized Medical Imaging and Graphics, 2001, 25(1): 47-59.
- [5] McLoughlin K J, Bones P J. Segmentation of the breast air boundary for a digital mammogram image [C] //Proceedings of Image Vision Computing. Hamilton, New Zealand: University of Auckland, 2000: 228-233.
- [6] Sun Y J, Suri J S, Desautels J E, et al. A new approach for breast skin-line estimation in mammograms[J]. Pattern Analysis & Applications, 2006, 9(1):34-47.
- [7] Ridler T W, Calvard S. Picture thresholding using an iterative selection method[J]. IEEE Transactions on Systems, Man and Cybernetics, 1978, 8(8):630-632.
- [8] Xu Weidong, Li Lihua, Xu Ping. Estimation of breast skin-line using least square estimation and watershed segmentation[C] // The 1st International Conference on Bioinformatics and Biomedical Engineering. Piscataway, US: IEEE Engineering in Medicine and Biology Society, 2007: 913-915.
- [9] Lou S L, Lin H D, Lin K P, et al. Automatic breast region extraction from digital mammograms for PACS and telemammography applications [J]. Computerized Medical Imaging and Graphics, 2000, 24(4):205-220.
- [10] Hu Q, Hou Z, Nowinski W L. Supervised range-constrained thresholding[J]. IEEE Transactions on Image Processing, 2006, 15 (1): 228-24.
- [11] Shahedi B K, Amirfattahi R, Azar F T, et al. Accurate breast region detection in digital mammograms using a local adaptive thresholding method [C] //The 8th International Workshop on Image Analysis for Multimedia Interactive Services. Santorini, Greece: The Institution of Engineering and Technology, 2007: 26-30.
- [12] Wei Ping, Li Junli, Lu Dongming, et al. A Fast and reliable switching median filter for highly corrupted images by impulse noise [C] //IEEE International Symposium on Circuits and Systems. Piscataway, US: IEEE CAS, 2007: 3427-3430.
- [13] Padayachee J, Alport M J, Rae W I. Identification of the breast edge using areas enclosed by iso-intensity contours [J]. Computerized Medical Imaging and Graphics, 2007, 31(6): 390-400.