

基于特征提取的脑部 MRI 肿瘤自动分割

宣 晓¹, 廖庆敏²

(1. 清华大学电子工程系, 北京 100084; 2. 清华大学深圳研究生院, 深圳 518055)

摘要: 对脑部磁共振图像中肿瘤的自动分割, 有助于了解疾病特征和制定手术方案, 评价治疗效果。该文通过提取基于灰度统计、对称性、纹理等的特征, 结合 AdaBoost 方法, 利用计算机进行自动脑肿瘤分割。该方法综合利用了磁共振(MR)各加权图像的信息和大脑解剖结构的知识, 以及 AdaBoost 算法的特征选择能力。在 20 帧带有肿瘤的 MR 图像上进行实验, 得到了 96.82% 的分类准确率。

关键词: 磁共振图像; 肿瘤分割; 纹理特征; 特征选择; AdaBoost 方法

Automated MRI Brain Tumor Segmentation Based on Feature Extraction

XUAN Xiao¹, LIAO Qing-min²

(1. Department of Electronic Engineering, Tsinghua University, Beijing 100084;

2. Graduate University at Shenzhen, Tsinghua University, Shenzhen 518055)

【Abstract】 Automated MRI brain tumor segmentation provides a powerful tool for diagnosis. In this paper, a tumor segmentation method using multi-feature extraction and AdaBoost feature selection is presented. The method utilizes the information of MR images and the anatomical knowledge, and takes the advantage of the feature selection ability of AdaBoost. Experimental results on 20 slices MR images demonstrate the effectiveness of the feature selection, and achieve an accuracy of 96.82% on tumor segmentation.

【Key words】 Magnetic Resonance Image(MRI); tumor segmentation; texture features; feature selection; AdaBoost

1 概述

由于具有非损伤性和较高的软组织分辨率, 磁共振(Magnetic Resonance, MR)图像已成为诊断脑部肿瘤的重要工具。在利用磁共振图像诊断脑部肿瘤时, 医生必须对图像中的肿瘤进行分割, 从而了解肿瘤的位置和大小, 以及其附近功能组织的分布。现阶段肿瘤分割仍然依靠医生手工完成, 这种方法大量消耗医生的精力和时间。同时, 不同医生或是同一医生在不同时间所做出的分割结果误差很大, 可重复操作性差。如果能够利用计算机对肿瘤进行自动分割将大大提高医生的工作效率, 并且可以提供一个客观的分割结果。然而遗憾的是, 由于肿瘤的形状、位置、灰度分布等的复杂多变性, 现在还没有一种分割方法能够满足临床应用的要求。

现有肿瘤分割方法中大部分是基于模式识别的方法, 将分割问题看作像素的分类问题, 其成功的关键是提取有效的特征区分肿瘤类和各种正常组织。常用的特征有灰度特征和纹理特征等。现有的研究中存在的主要问题是特征提取的同时缺乏对特征的选择, 尤其是纹理特征, 通常维数较高, 但并非每一维都为分割提供了有用的信息。

为了解决以上问题, 本文重点研究了肿瘤分割中的特征提取与选择。通过对 MR 脑部图像中肿瘤性质的分析, 提出了基于灰度、对称性和纹理等的 3 类特征。然后利用 AdaBoost 方法选出对分割最有效的特征, 并进行分类。AdaBoost 方法于 1995 年被提出^[1], 是一种将弱分类器组合成强分类器的方法, 但也可以被用作特征选择。例如文献[2]采用 AdaBoost 算法选择纹理特征, 对乳腺图像中的微小钙化病灶进行检测; 文献[3]用类似的方法对血管超声图像中的血管进行分割; 文

献[4]对 MR 图像中的前列腺肿瘤进行自动检测, 对图像提取特征之后, 根据训练集的知识 and 贝叶斯准则对每个特征建立一幅肿瘤类的似然图, 然后利用 AdaBoost 方法对各似然图进行加权组合得到最终的似然图, 这实际上也是一个特征选择的过程。综上所述, 用 AdaBoost 方法进行特征选择和分类在医学图像处理方面已经取得了一定的成果, 但总的来说应用还不多, 在脑部 MRI 肿瘤分割中, 还没有此类的方法出现。

2 脑部磁共振图像基本特征

MR 图像根据成像参数的不同, 可以得到不同的加权像, 本文实验所用的加权像有 T1, T2 和 FLAIR 3 种。正常的大脑组织主要包括灰质、白质、脑脊髓液等, 它们和肿瘤在各加权像中的灰度特性均有所不同, 如图 1 所示。

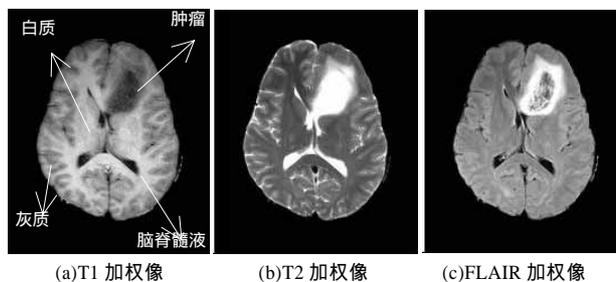


图 1 MR 图像

基金项目: 国家自然科学基金资助项目(60372023)

作者简介: 宣 晓(1982 -), 女, 硕士研究生, 主研方向: 计算机图像处理, 模式识别等; 廖庆敏, 教授、博士生导师

收稿日期: 2007-06-04 **E-mail:** xuanxiao@mails.tsinghua.edu.cn

在 T1 像中,肿瘤呈暗信号,在 T2 像中,肿瘤呈亮信号,在 FLAIR 像中,肿瘤呈亮信号。但是肿瘤内部的灰度分布常常是不均匀的,例如图 1(c)的 FLAIR 加权像中,大部分肿瘤都呈亮信号,但中心部分却呈现暗信号,原因是肿瘤内部有囊变。

3 特征提取

将图像中的像素及其所在矩形邻域作为一个整体(简称小块)进行特征提取。本文选取小块尺寸的基本原则是尽量确保落在灰质和白质中的小块内部包含脑沟,因为灰质和白质是脑组织中最大的组成部分,脑沟的纹理特征对于判断灰质和白质具有重要的意义。实验表明,当小块尺寸为 25×25 像素时,包含脑沟的小块最多。

3.1 基于灰度的统计特征

对每个小块提取 10 维基于灰度的统计特征,包括灰度的均值、方差、标准差、最大值、最小值、灰度范围、中心像素值、中值、偏度及峰度等。灰度信息直接反映了图像中各组织的特征,但是由于不同组织的灰度分布有所重叠,因此仅凭灰度特征来进行分割是不够的。

3.2 基于对称性的特征

一幅正常的脑部 MR 图像的重要特点是左右脑半球的对称性,由于肿瘤的存在会占据一部分正常脑组织的空间,因此会破坏这种对称性。最简单的检测对称性的方法是将对称轴两边的对称像素相减,差的绝对值越大说明越不对称。但是人体的对称性是对宏观而言的,具体到每一个像素总是存在微小的偏差,所以本文提出了基于模糊对称映射的特征,用以克服这种微小偏差造成的不对称。对于每一帧图像 I 按式(1)作映射得到模糊对称映射图 S 。

$$S(i, j) = \min_{(k, l) \in N(i', j')} |I(i, j) - I(k, l)| \quad (1)$$

其中 (i', j') 是像素 (i, j) 关于对称轴 L 的对称点; $N(i', j')$ 是 (i', j') 的一个圆形小邻域,定义如式(2), ε 为邻域 N 的半径,依照经验选取一个比较小的正数。

$$N(i', j') = \{(k, l) | \|(k, l), (i', j')\| < \varepsilon\} \quad (2)$$

基于对称性的特征即定义为小块中心像素对应的模糊对称映射 S 值。

3.3 基于纹理的特征

正常大脑中的各组织相对位置是基本固定的,因此同一组织内部和不同组织之间会形成一些相对固定的纹理样式,例如灰质和白质之间的脑沟等。于是可以采用某种纹理特征来区分正常组织和肿瘤。

本文利用 MPEG-7 中的同构型纹理图像描述子来描述一个小块的纹理特征,步骤如下所示^[5]:

(1)将输入的小块经过快速傅立叶变换变换到频域中,并以极坐标的方式表示为 $F(\omega, \theta)$ 。

(2)以一组不同方向、不同频率的 Gabor 滤波器处理 $F(\omega, \theta)$, 得到 $H_i(\omega, \theta) = G_{s,r}(\omega, \theta) \cdot F(\omega, \theta)$ 。 i 是频道的索引, $G_{s,r}(\omega, \theta)$ 是第 s 个半径分割和第 r 个方向上的 Gabor 函数:

$$G_{s,r}(\omega, \theta) = \exp\left[-\frac{(\omega - \omega_s)^2}{2\sigma_{\omega_s}^2}\right] \cdot \exp\left[-\frac{(\theta - \theta_r)^2}{2\sigma_{\theta_r}^2}\right] \quad (3)$$

其中, $\{\omega_s = \omega_0 \cdot 2^{-s}, s = 0, 1, 2, 3, 4\}$ 为滤波器在其相应方向 θ_r 上的中心频率, ω_0 为中心频率的最大值 $3/4$; 相应的带宽 $\{B_s = B_0 \cdot 2^{-s}, s = 0, 1, 2, 3, 4\}$, B_0 为最大带宽 $1/2$; $\{\theta_r = 30^\circ \times r, r = 0, 1, 2, 3, 4, 5\}$ 为滤波器的中心角度; $\sigma_{\omega_s} = B_s / (2\sqrt{2\ln 2})$;

$$\sigma_{\theta_r} = 30^\circ / (2\sqrt{2\ln 2})。$$

(3)计算滤波结果的平均能量 e_i 和标准差 d_i

$$e_i = 1b[1 + p_i], \text{ 其中 } p_i = \frac{\sum_{\omega, \theta} H_i^2(\omega, \theta)}{\sum_{\omega, \theta} H_i^2(\omega, \theta) - p_i} \quad (4)$$

$$d_i = 1b[1 + q_i], \text{ 其中 } q_i = \sqrt{\frac{\sum_{\omega, \theta} [H_i^2(\omega, \theta) - p_i]^2}{\sum_{\omega, \theta} [H_i^2(\omega, \theta) - p_i]^2}} \quad (5)$$

依照以上 3 个步骤,对每一个小块可以得到 30 个能量平均值和 30 个能量标准差,将这 60 个数值构成的向量作为小块的纹理特征。

4 基于 AdaBoost 的特征选择和分类

AdaBoost 是一种分类方法,它通过叠加大量能力一般的弱分类器,构成分类能力很强的强分类器。只要令算法中每个弱分类器只对应一个特征并根据特征值的大小来进行分类判断,则 AdaBoost 挑选弱分类器的过程也就是挑选特征的过程。参考文献[1]并做修改,采用如下的 AdaBoost 算法:

实验训练集表示为 $(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)$ 。其中, x_i 属于特征空间 X , m 为样本总数;类别标签 y_i 属于标签集 $Y = \{-1, +1\}$, -1 代表正常组织, $+1$ 代表肿瘤类;循环次数 $t = 1, 2, \dots, T$, 每一次循环样本 i 对应的权值分布为 $D_t(i)$ 。

(1)初始化样本权重 $D_1(i) = 1/m$ 。

(2)对于 $t = 1, 2, \dots, T$:

对特征向量中的每个特征 f_j 设计简单分类器

$$h_j(x_i) = \begin{cases} 1 & p_j f_j(x_i) < p_j \theta_j \\ -1 & \text{其他} \end{cases} \quad (6)$$

调整阈值参数 θ_j 以及方向参数 p_j ($p_j = 1, -1$), 使得分类加权误差 $\varepsilon_j = \sum_i \omega_i |h_j(x_i) - y_i|$ 最小。在所有分类器中, 选择出加权误差最小的分类器 h_r , 其加权误差为 ε_r , 则 f_r 即为本轮循环所选出的特征。对所有样本, 令 $f_r = 0$, 记 $h_t = h_r$ 。

更新权重:

$$D_{t+1}(i) = [D_t(i) \exp(-\alpha_t y_i h_t(x_i))] / Z_t \quad (7)$$

其中, $\alpha_t = (1/2) \ln[(1 - \varepsilon_t) / \varepsilon_t]$; Z_t 为归一化因子使得所有样本权重 D_{t+1} 的和为 1。

(3)输出最终的分类器

$$H(x) = \text{sign}\left(\sum_{t=1}^T \alpha_t h_t(x)\right) \quad (8)$$

5 实验结果及分析

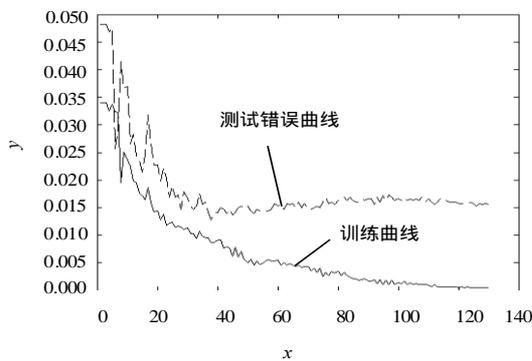
所用的实验数据为 T1, T2 和 FLAIR 3 类加权像, 共 20 个肿瘤病例, 随机挑选其中 10 个作为训练集, 其余 10 个作为测试集。由有经验的医生描绘的肿瘤边界在实验中被作为真值(Ground Truth)。

按第 3 节的方法提取基于灰度、对称性、纹理的 3 类特征, 由于采用了 3 种加权像, 因此总的特征维数为 $(10+1+60) \times 3$ 共 213 维。

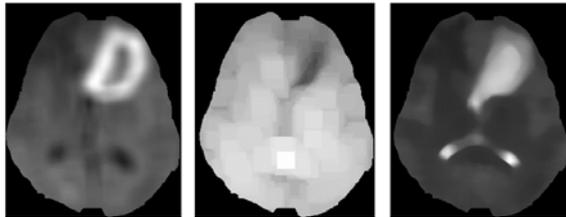
对于训练集的所有图像, 一共提取出 6 000 个小块, 其中 3 000 个正常组织小块, 3 000 个肿瘤小块。利用 AdaBoost 算法在这些数据上训练最优分类器, 然后在测试集上测试分类器的有效性。测试集包括从测试图像中提取的 2 000 个正常组织小块和 2 000 个肿瘤小块。AdaBoost 分类器的训练和测试错误曲线如图 2(a)所示, 曲线描述了每次循环分类器在训练集和测试集上的分类错误率。图中横轴为循环次数, 纵轴为分类错误率。由曲线可知, 随着循环次数的增加, 训练的错误率逐渐趋近于 0。而当循环次数增加到 40 左右之后, 测试的错误率就基本不变了, 这说明剩下的特征对于分类基

本没有起作用,属于无用特征。因此,在对一幅新的图像进行处理时,只需提取 40 维特征而无需 213 维,这样就大大减少了运算复杂度。

AdaBoost 算法每循环一次都会挑选出一个在当前样本权值分布下最有效的一维特征,前 3 维挑选出来的特征记为 F1,F2,F3,分别为:FLAIR 加权像的纹理特征(Gabor 滤波器参数 $\omega_s = 3/4, \theta_r = 90^\circ$), T1 加权像的最大值特征和 T2 加权像的中值特征,如图 2(b)所示,左起:F1-FLAIR 的纹理特征(Gabor 滤波器参数 $\omega_s = 3/4, \theta_r = 90^\circ$); F2-T1 的最大值特征; F3-T2 加权像的中值特征。在 F1 中,肿瘤呈现高亮度,内部的囊变和脑脊髓液都呈现低亮度。在 F2 中,囊变呈现低亮度而其他部分呈现高亮度。在 F3 中,肿瘤、囊变和脑脊髓液均呈现高亮度。因此通过 F1 可以大致确定肿瘤的位置,而 F2 和 F3 又分别补充了囊变和脑脊髓液的信息,可见这 3 维特征对于肿瘤分割提供了不相关且具有互补性的信息。



(a)AdaBoost 算法的训练和测试错误曲线



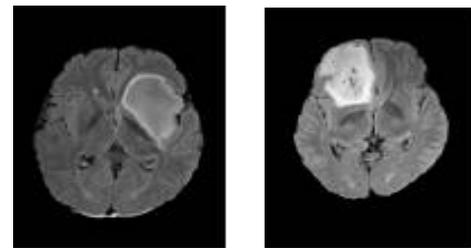
(b)AdaBoost 算法选出的前 3 维特征

图 2 基于 AdaBoost 的特征选择

在选择出的全部 40 维特征中,包含了 16 维基于灰度的统计特征,3 维基于对称性的特征和 21 维基于纹理的特征。这一结果证明了本文中提取的 3 方面特征信息对于肿瘤分割均有意义。

此外,这 40 维特征在 T1、T2 以及 FLAIR 加权像上的分布分别为 11,14,15,基本相等,这说明不同的加权像都为分割提供了信息。不过,T2 和 FLAIR 对于肿瘤的区别度略高于 T1,这与影像学的结论是相一致的。利用 40 维挑选出来的特征和训练得到的分类器对训练集中的图像进行肿瘤分割,平均分割准确率为 96.82%,其中,假阳性率 1.3%(被错分为肿瘤的正常像素数/正常像素总数),假阴性率 3.65%(被错分为正常的肿瘤像素数/肿瘤像素总数)。

部分分割结果如图 3 所示,可以看出本文的分割结果和医生手绘的结果非常接近,证明了该方法对于肿瘤分割的有效性。



(a) FLAIR 原图



(b)Ground Truth



(c)分割结果

图 3 肿瘤分割结果

6 结束语

本文提出了一种脑部 MR 图像的自动肿瘤分割方法,通过对脑部 MR 图像中各组织灰度特性及脑部解剖结构的分析,提出了基于灰度统计、对称性和纹理的 3 类特征,并且利用 AdaBoost 算法对特征进行了选择,得到 40 维对分类最有效的特征。实验证明 AdaBoost 算法选出的特征具有互补性和不相关性,且本文提出的 3 类特征对于肿瘤分割均提供了有效的信息。此外,实验所用的 3 类磁共振加权图像均对分割肿瘤有意义,但 T1 加权像提供的有效信息比 T2, FLAIR 略少。本文的方法在测试集上的肿瘤分割准确率达到 96.82%,证明了该方法的有效性。

参考文献

- [1] Freund Y, Schapire F E. A Short Introduction to Boosting[J]. Journal of Japanese Society for Artificial Intelligence, 1999, 14(5): 771-780.
- [2] Yuan Xin, Cheng Shipeng. Microcalcification Detection Based on Localized Texture Comparison[C]//Proceedings of International Conference on Image Processing. [S. l.]: IEEE Press, 2004: 2953-2956.
- [3] Pujol O, Rosales M. Intravascular Ultrasound Images Vessel Characterization Using AdaBoost[M]. Berlin: Springer, 2003: 242-251.
- [4] Madabhushi A, Feldman M D. Automated Detection of Prostatic Adenocarcinoma from High-resolution *Ex Vivo* MRI[J]. IEEE Transactions on Medical Imaging, 2005, 24(12): 1611-1625.
- [5] Ro Yongman, Kim M, Ho Kyung Kang, et al. MPEG-7 Homogeneous Texture Descriptor[J]. ETRI Journal, 2001, 23(2): 41-51.