

一种基于 CT 影像的肺癌病灶检测新方法

贾 同, 魏 颖, 赵大哲

(东北大学信息科学与工程学院, 辽宁沈阳 110819)

摘 要: 肺癌病灶的检测一直是重要与困难的工作, 本文提出了一种基于三维 CT 影像的肺结节计算机辅助检测新方法. 基于自适应阈值等方法分割肺实质区域; 由于肺血管是肺结节检测的重要干扰, 建立一种形变模型精确分割并过滤肺内血管组织; 基于 Hessian 矩阵特征值构造可选择形状滤波器检测疑似结节, 并进一步过滤剩余的细小血管组织; 提取多个结节特征, 并采用基于规则分类器进行分类. 实验结果表明, 该方法可以有效帮助医生提高肺癌疾病的诊断准确率.

关键词: 计算机辅助诊断; 肺实质分割; 肺血管分割; 肺结节检测; 形变模型

中图分类号: TP391.41 **文献标识码:** A **文章编号:** 0372-2112 (2010) 11-2545-05

A New Lung Cancer Lesions Detection Scheme Based on CT Image

JIA Tong, WEI Ying, ZHAO Da-zhe

(College of Information Science and Engineering, Northeastern University, Shenyang, Liaoning 110819, China)

Abstract: Lung cancer lesions detection has been an important and difficult work. A computer-aided detection (CAD) scheme for detecting lung nodules is proposed in three-dimensional CT images in this paper. The lung parenchyma is segmented from the CT data using adaptive threshold method etc. Pulmonary vascular is the main disturbance for nodules detection, building an active contour model to segment and remove pulmonary vascular accurately in the lung region. Suspicious nodules are detected and omitted renal vascular is filtered using a selective shape filter, which is based on the eigenvalues of a Hessian Matrix. Nodule features are extracted and rule-based classifier is used to distinguish true or false positive nodules. Experiment results indicate that this scheme can help physician improve the diagnosis efficiency.

Key words: computer-aided detection; lung parenchyma segmentation; lung vascular segmentation; lung nodule detection; active contour model

1 引言

肺癌是威胁人类健康的最大杀手, 目前致死率已超过了三种最常见癌症(乳腺癌、前列腺癌和结肠癌)的总和. 由于 CT 影像能够提供高清晰度的影像, 并且为影像中各组织提供很高的对比度, 通常应用于肺部疾病的诊断. 随着多排螺旋 CT 的发展, 医生可以获得更高分辨率的影像(High Resolution CT, HRCT), 通过一次检测获得更多的患者影像信息. 但同时也造成了医生每天阅片负担的增加, 导致漏诊和不确定性检查经常发生. 因此, 肺癌计算机辅助检测(Computer-Aided Detection, CAD)系统得到了广泛的研究, 目的在于辅助医生对肺癌进行早期发现、诊断与治疗^[1].

肺结节是肺癌疾病的重要临床表现, 其二维 CT 影一般近似圆形、三维形状为球形, 直径大约为 4 ~ 30mm

区间. 由于肺 CT 影像包含血管、气管、胸膜、骨骼、胸壁软组织等多种组织信息, 给肺结节检测造成了干扰. 尤其是肺血管组织, 由于其密度、位置、形态都与结节 CT 影较为接近(如图 1 所示), 一般医生阅片时都是通过 CT 序列影像前后层信息的相关性来进行肺血管与真阳性结节的区分, 而通过计算机辅助检测则难以识别.

在对肺结节检测方法的研究中, Fiebich 等人采用了阈值分割与区域增长的方法^[2,3], 由于结节的密度与肺内血管密度相近, 以及受到局部容积效应(Partial Volume Effect, PVE)影响, 很难通过单一的阈值完成检测; Yamamoto 等人采用基于数学形态学滤波器的方法^[4], 对每一个规定范围内的结节尺度进行滤波, 由于肺内血管等 CT 影像的尺度与结节影相近, 给精确检测带来了干扰; Gray 等人采用模式分类的方法^[5,6], 需要基于大量的结节样本数据进行训练, 而且需要提取多个结节特征, 算

法处理时间较长。

基于以上分析,本文提出了一种新的肺癌病灶检测方法.首先,对原始肺部 HRCT 图像进行肺实质分割;然后,基于形变模型方法分割肺内血管,并构建基于 Hessian 矩阵的形状滤波器,检测疑似结节区域;最后,通过特征提取与分类方法精确检测肺结节。

2 肺癌病灶检测方法

在输入原始肺部 CT 影像后,本文方法主要包括肺实质分割、肺血管分割、疑似结节检测及特征提取与分类等几个步骤,并对最后检测结果进行标定与显示.肺组织示意图如图 1 所示。



图1 肺组织示意图(a)其中A标定为结节区域,形状为“圆”形,B标定为血管区域,形状为“线”形,C标定为肺实质区域

2.1 肺实质分割

肺实质分割是结节检测的重要预处理步骤.主要包括肺实质初步分割、左右肺分离、肺实质边缘修补等。

2.1.1 肺实质初分割

本文采用一种改进的自适应阈值分割方法实现肺实质区域的初步分割,二维图像的最优阈值 T 由图像的平均梯度 G 来确定,如公式(1),(2)所示:

$$G = \frac{\sum_{i,j \in Z} g(i,j)}{\sum_{i,j \in Z} p(i,j)} \quad (1)$$

$$\hat{T} = \arg \max_{T \in Z} [G(T)] \quad (2)$$

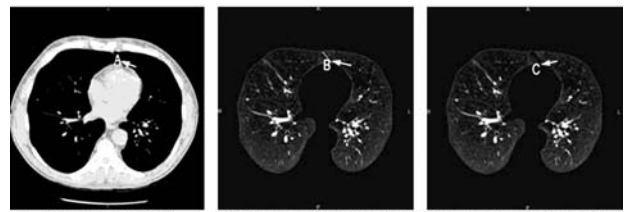
$$p(i,j) = \begin{cases} 1, & g(i,j) > 0 \\ 0, & g(i,j) = 0 \end{cases} \quad (3)$$

其中 $f(i,j)$ 代表被 T 剪切后的图像, $g(i,j)$ 代表 $f(i,j)$ 的梯度图,即对二维离散函数 $f(i,j)$ 求导的过程, $p(i,j)$ 代表图像中非零梯度像素的总和, i,j 分别代表图像的行和列, Z 代表整数,如公式(3)所示.最后, T 即为所计算的最优阈值。

2.1.2 左右肺分离

气管和主支气管区域对于左右肺组织分离是一个

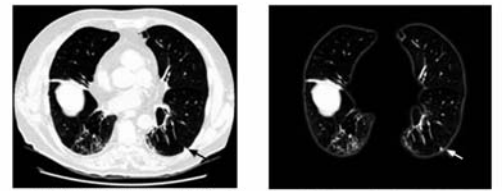
干扰,剔除这部分区域后,左右肺区域在连接线区域是粘连在一起的^[7].由于连接线区域非常狭窄,并且比周围组织有更高的密度,因此,可以先定位出连接线区域,再实现左右肺分离,实验结果如图 2 所示。



(a) 原始肺CT影像, A区域为左右肺位置 (b) 肺实质初分割结果, B区域为左右肺连接线定位区域 (c) 完成左右肺分离, C区域显示左右肺在该连接线区域分离
图2 左右肺分割

2.1.3 肺实质边缘修补

由于肺实质边缘,即靠近肺壁处结节密度和周围组织非常相近,在肺实质初步分割时造成遗漏,因此,采用数学形态学算法在肺实质内包含这部分区域,实验结果如图 3 所示。



(a) 边缘修补前肺CT影像,箭头指向肺实质边缘结节 (b) 修补后,肺边缘结节包含在所分割的肺区域内
图3 肺实质边缘修补

2.2 基于形变模型的血管分割

在肺实质分割后,根据肺内血管三维拓扑结构复杂,且分支较多的特点,本文提出了一种基于几何形变模型的三维血管分割方法,演化曲面可以快速定位目标内外的同质区域,而且可以精确收敛到目标边缘,并采用优化水平集方法^[8]进行求解。

2.2.1 区域均值计算

本文依据三维医学图像特点,把二维 Chan-Vese 模型^[9]扩展到三维形变模型.假设定义域为 Ω 的图像 $I(x,y)$ 被闭合边界划分为目标和背景两个同质区域, Chan-Vese 模型具有以下能量函数:

$$F(c^+, c^-, C) = \mu \text{Length}(C) + \lambda^+ \int_{\text{inside}(c)} |I(x,y) - c^+|^2 dx dy + \lambda^- \int_{\text{outside}(c)} |I(x,y) - c^-|^2 dx dy \quad (4)$$

其中 c^+, c^- 分别为目标区域和背景区域的平均灰度, $\mu > 0, \lambda^+, \lambda^- > 0$ 是各个能量项的权重系数。

设原图像 $I(x,y)$ 被活动轮廓线 C 划分为目标和背景两个区域,各个区域的平均灰度值为 C_1 和 C_2 ,对公式(4)进行简化,得到公式(5):

$$E_{REG} = \alpha \int_{\text{inside}(c)} |I(x,y) - c_1|^2 dx dy$$

$$+ \int_{\text{outside}(c)} |I(x, y) - c_2|^2 dx dy] + r \int_c ds \quad (5)$$

其中第一项为图像区域能量拟合项,第二项为几何约束项,要求曲线 C 尽可能短, α, γ 为常数, ds 为弧长微元。

2.2.2 边缘能量计算

本文采用一种使曲线沿着在目标梯度方向的二阶导数进行演化的边缘检测方法,可以在有效抑制噪声的前提下,最大限度的保证边缘的连续性。而且,该方法具有良好的三维扩展特性,适用于采用水平集方法进行求解。

设 $I_{\xi\xi}$ 为在图像梯度方向的二阶导数, $\mathbf{I}_{\eta\eta}$ 为其垂直矢量。由于 $\mathbf{I}_{\xi\xi} = I_{\xi\xi} + \mathbf{I}_{\eta\eta} - \mathbf{I}_{\eta\eta} = \nabla I - \mathbf{I}_{\eta\eta}$, 构建曲线能量函数为:

$$E_{EDGE} = \iint_s \langle \nabla I, \mathbf{n} \rangle da - \iiint_{\Omega_s} \mathbf{H}_I |\nabla I| dx dy dz \quad (6)$$

其中 \mathbf{H}_I 为曲面的曲率,即为三维体数据水平集, S 为演化曲面, da 为曲面表面区域, \mathbf{n} 为曲面单位方向矢量。

2.2.3 总能量函数合成

通过目标边缘能量的计算,可以定位目标边缘,但是当初始曲面远离实际目标时,很难精确定位所有目标边缘。通过与图像目标内外区域信息的集合,可以较好的解决这一问题。因此,总能量函数定义为:

$$E = -E_{EDGE} + E_{REG}$$

通过式(5)、(6),进一步定义能量函数为式(7):

$$E = \alpha \left[\int_{\text{inside}(c)} |I(x, y) - c_1|^2 dx dy + \int_{\text{outside}(c)} |I(x, y) - c_2|^2 dx dy \right] - \left(\iint_s \langle \nabla I, \mathbf{n} \rangle da - \iiint_{\Omega_s} \mathbf{H}_I |\nabla I| dx dy dz \right) + r \int_c ds \quad (7)$$

能量函数(7)由三项构成,第一项为图像区域均值计算,第二项为图像边缘能量计算,第三项为曲面几何约束项。各变量定义同公式(4)、(5)、(6),如 \mathbf{n} 为曲面单位方向矢量。

采用优化水平集方法求解,血管分割结果如图4所示,(a)为原始肺部CT影像,(b)为相应的血管分割三维显示结果,对细小分支得到了准确的分割。

2.3 基于 Hessian 矩阵的疑似结节检测

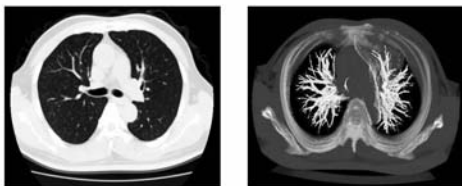


图4 肺血管分割结果

本文基于 Hessian 矩阵构建可选择形状滤波器。肺结节在 CT 影像上表现为圆形结构,三维形态上近似为球形,肺血管三维形态近似为管形。该方法可以对分割遗漏的细小血管组织进行抑制,并着重检测类似“球形”的结节区域。

采用高斯函数分别建立 CT 影像中结节、血管的数学模型,表示如下:

$$\begin{cases} d(x, y, z) = \exp\{- (x^2 + y^2 + z^2)/2\delta^2\} \\ l(x, y, z) = \exp\{- (x^2 + y^2)/2\delta^2\} \end{cases} \quad (8)$$

这里 $d(x, y, z)$ 是一个由三维高斯函数构成的球形表达式, $l(x, y, z)$ 是一个二维高斯函数表达式,参数 σ 在高斯函数中表示球的尺度。

设 CT 序列影像表示为 $f(x, y, z)$, 根据它的二阶导数构建 $f(x, y, z)$ 的 Hessian 矩阵,如公式(9)所示:

$$\mathbf{H} = \begin{bmatrix} f_{xx} & f_{xy} & f_{xz} \\ f_{yx} & f_{yy} & f_{yz} \\ f_{zx} & f_{zy} & f_{zz} \end{bmatrix} \quad (9)$$

对于球的理想形状,六个二阶偏导数 $f_{xy}, f_{xz}, f_{yz}, f_{yx}, f_{zx}$ 与 f_{zy} 相等为零。因此构建的滤波器只依赖于 f_{xx}, f_{yy} 与 f_{zz} , 且为 Hessian 矩阵的三个特征值,分别用 λ_1, λ_2 和 λ_3 表示。

通过对形状相似性与尺度的分析,构建的球形滤波器 Z_{dot} 如式(10)所示:

$$Z_{dot}(\lambda_1, \lambda_2, \lambda_3) = |\lambda_3|^2 / |\lambda_1| \quad \lambda_1 < 0, \lambda_2 < 0, \lambda_3 < 0$$

$$Z_{dot}(\lambda_1, \lambda_2, \lambda_3) = 0, \text{ otherwise}$$

(10)

在肺结节检测中,由于结节大小的不一致,需要在不同尺度空间下对结节进行滤波检测,通过调节尺度因子 σ ,可以得到不同尺度的滤波函数。具体实现步骤如下所示:

(1)根据结节直径大小范围,指定 N 个平滑尺度, N 为正常数。对于每一个平滑尺度 σ_s ,利用高斯函数平滑原始影像。

(2)对于影像中的每一个像素,重复 3-5 步。

(3)利用式(9)计算 Hessian 矩阵和它的特征值 λ_1, λ_2 和 λ_3 。

(4)利用式(10)计算球形滤波器。

(5)结束对每个像素的循环。

(6)对于滤波后影像的每个像素乘以 σ_s^2 ,对于 N 个尺度,选取最大值作为滤波后影像。

该方法不仅可以检测类球形疑似结节区域,而且可以进一步过滤肺内血管等管状组织的干扰。疑似结节滤波结果如图5所示。

2.4 结节特征提取与分类

2.4.1 特征提取

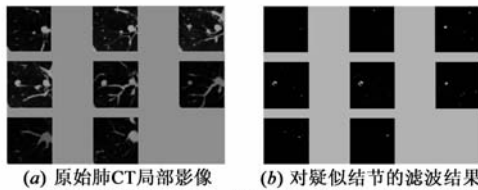


图5 过滤血管信息后,基于形状滤波器的滤波显示结果

得到疑似结节区域后,提取有效的特征对真假阳性结节的识别非常重要.在对真阳性肺结节病理特征研究的基础上,本文提取多个结节三维特征,如 CT 最大值、CT 最小值、CT 均值、最大直径、体积、球形度及空间分布密度等.

对于提取的多个特征,本文基于多组样本数据进行训练,利用概率分布的可分性进行了特征选择.在图6中,各图的横坐标表示该特征的值,纵坐标表示概率密度,实线表示真阳性结节样本分布,虚线表示假阳性结节样本分布.通过比较可以发现,结节直径、体积特征曲线交叠较多,表明这些特征对分类的贡献较小;而结节球形度、空间分布密度特征曲线交叠较少,表明它们对分类贡献较大.

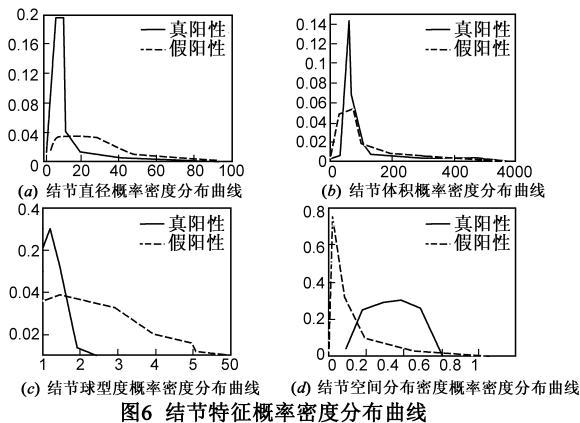


图6 结节特征概率密度分布曲线

2.4.2 特征分类

对于提取的特征参数,本文采用基于规则分类的方(Rule Based Classification, RBC)进行真假阳性结节的区分^[10].每一个特征与一个规则相联系,一个规则的输出作为下一个规则的输入. RBC 与神经网络、K 近邻等分类方法相比,是一种最为高效的分类方法,符合各类型 CAD 系统对实时性、准确性要求.

对于具体规则的确定,本文基于医学诊断规则并对大量样本数据反复训练获得一系列对应各个特征的阈值范围,并在测试样本中的指标是否符合系统要求为最终确定依据.特征分类后,本文对检测到的真阳性结节区域进行轮廓跟踪、标定与显示,并与采用文献[2]与文献[4]方法的检测结果进行了对比,结果如图7所示.

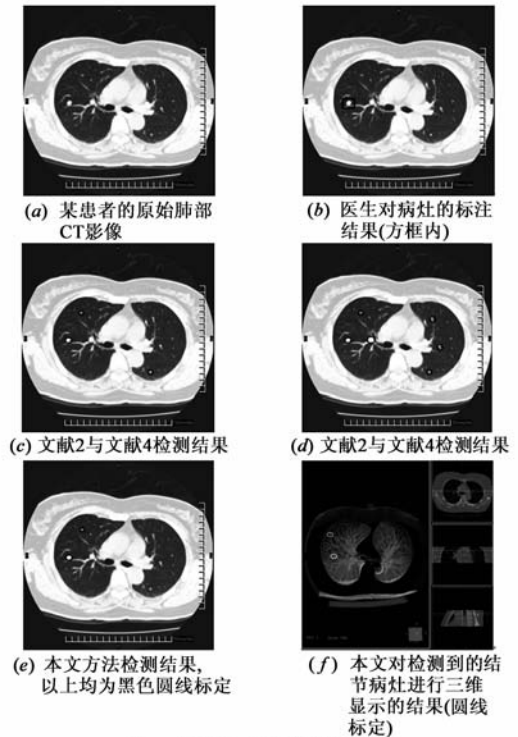


图7 肺结节病灶检测结果

3 实验结果与分析

本文实验数据集都来自医院临床数据,共包含 90 组肺癌患者 CT 影像,有经医生确诊结节(TP)232 个,其中 45 组作为训练数据集(TP = 124),另 45 组作为测试数据集(TP = 108).每层 CT 影像均为 512×512 像素,层厚 $1 \sim 2\text{mm}$,像素间距 0.6835mm .

对于 CAD 系统的评价,本文采用敏感性与每组数据所检测出的假阳性结节个数(FPs/scan)两个指标^[15]作为评价依据,其定义如式(11)所示.

$$\begin{cases} Sen = TP / (TP + FN) \\ FPs / scan = FP / N \end{cases} \quad (11)$$

其中,TP 为检测出的真阳性结节个数,FP 为检测出的假阳性结节个数,FN 为没有检测出的真阳性结节即假阴性结节个数,N 为测试样本中的数据组数(本文 $N = 45$).

测试数据集中共有经医生确诊的真阳性结节 108 个,结节分布包括肺实质边缘、孤立性肺结节、与血管相邻三个位置,覆盖了所有临床病例中肺结节分布情况,其中与血管粘连带结节占总样本的 55%,符合临床病例比例.最终检测结果为 $Sen = 90\%$, $FPs / scan = 1$.由于本文方法先对肺血管进行了较为准确的分割与过滤,较好的提高了对与血管粘连带结节的检测准确率,并显著提高了总体检测敏感性.

4 结论

本文提出了一种新的肺结节病灶检测方法. 主要分为以下几个步骤: 首先, 从原始肺部 CT 影像中分割出肺实质区域; 然后, 根据血管拓扑结构特点, 提出一种形变模型方法分割并过滤肺内血管组织; 再次, 基于 Hessian 矩阵特征值构建可选择形状滤波器, 精确检测“球形”疑似结节区域, 并进一步过滤“管形”组织的干扰; 最后, 提取多个结节三维特征, 采用按规则分类的方法进行真假阳性结节的区分, 并对分割结果进行三维显示.

通过多组肺癌 CT 影像测试以及与其它相关方法的对比, 本文方法对于肺结节检测敏感性较高, 基本满足了医生对于肺癌 CAD 系统准确性、易操作性及实时性的需求. 肺内血管分割在临床上具有重要意义, 可以辅助医生对多种疾病进行诊断, 具有多种应用目的及更为精确的血管分割和建树是我们下一步将要研究的内容.

参考文献:

- [1] K DOI. Current status and future potential of computer-aided diagnosis in medical imaging [J]. *The British Journal of Radiology*, 2005, 78(3): 3 - 19.
- [2] Armato SG III, Li F, Giger ML, MacMahon H, Sone S, Doi K. Lung cancer: performance of automated lung nodule detection applied to cancers missed in a CT screening program [J]. *Radiology*, 2002, (225): 685 - 692.
- [3] Wormanns D, Fiebich M, Saidi M, Diederich S, Heindel W. Automatic detection of pulmonary nodules at spiral CT: clinical application of a computer-aided diagnosis system [J]. *Euro Radiol*, 2002, (12): 1052 - 1057.
- [4] S Yamamoto, H Takizawa, H Jiang, T Nakagawa, T Matsumoto, Y Tateno, T Iinuma, et al. A CAD system for lung cancer screening test by X-ray CT [A]. *Proceedings of the International Conference on Computer Assisted Radiography and Surgery CARS'01*[C]. Berlin: Elsevier, 2001. 605 - 610.
- [5] Y Lee, T Hara, H Fujita, S Itoh, T Ishigaki. Nodule detection on chest helical CT scans by using a genetic algorithm [J]. *Proc Intel Inf Systems*, 1999, 3(5): 67 - 70.

- [6] Mcnitt Gray M F, Sayre J W, Huang HK. Pattern classification approach to segmentation of digital chest radiographs and chest CT image slices [J]. *Proc SPIE*, 1994, 2(1): 465 - 476.
- [7] Leader JK, Zheng B, Rogers RM. Automated lung segmentation in X-ray computed tomography: Development and evaluation of a heuristic threshold-based scheme [J]. *Acad Radiol*, 2003, 10: 1224 - 1236.
- [8] 陈志彬, 邱天爽, SU Ruan. 一种基于 FCM 和 Level Set 的 MRI 医学图像分割方法 [J]. *电子学报*, 2008, 36(9): 1733 - 1736.
Chen Zhibin, Qiu Tianshuang, Su Ruan. FCM and level set based segmentation methods for brain MR images [J]. *ACTA Electronica Sinica*. 2008, 36(9): 1733 - 1736. (in Chinese)
- [9] T Chan, L Vese. Active contours without edges [J]. *IEEE Trans of Image Processing*, 2001, 10(2): 266 - 277.
- [10] Qilang Li, Kunio Doi. Analysis and minization of overtraining effect in rule-based classifiers for computer-aided diagnosis [J]. *Med Phys*, 2006, 33(2): 320 - 328.

作者简介:



贾 同 男, 1975 年生于辽宁沈阳. 东北大学信息科学与工程学院讲师. 研究方向为医学图像处理与模式识别.

E-mail: jiatong@ise.neu.edu.cn



魏 颖 女, 1968 年生于辽宁本溪. 东北大学信息科学与工程学院副教授. 研究方向为医学图像处理与模式识别, 自动检测技术.

E-mail: weiyiing@ise.neu.edu.cn