论 文

基于流向特征熵和测地线距离的粘连血管型肺结节聚类分割

孙申申^{①②}, 郭阳^③, 任会之^④, 范立南^①, 康雁^{②*}

① 沈阳大学信息工程学院, 沈阳, 110044

② 东北大学中荷生物信息与工程学院, 沈阳, 110819

③ 东北大学理学院数学系, 沈阳, 110819

④ 沈阳工业大学机械学院, 沈阳, 110870

* 通信作者. E-mail: kangy@neusoft.com

收稿日期: 2012-10-26; 接受日期: 2012-12-14

国家自然科学基金青年基金 (批准号: 71201105)、辽宁省自然科学基金 (批准号: 20102154) 和辽宁省教育厅科研项目计划 (批准号: L2010376) 资助项目

摘要 肺癌计算机辅助诊断 (lung cancer CAD) 是辅助医生定量判别结节良恶性的新技术. 倍增率是临床上判断结节良恶性的指标, 而精确地分割结节又是计算倍增率的前提. 因为结节和血管的 CT 值相近, 所以难以正确分割粘连血管型结节. 血管里充满着流向同一方向的血液, 使得大部分血管像素梯度的法向量 (流向特征) 都指向同一方向, 流向特征熵值小; 而结节上的像素梯度法向量方向杂乱无章, 流向特征熵值大. 大部分血管像素到结节中心的测地线距离比结节像素到结节中心的距离大. 基于上述血管与结节差异, 文中提出了一种基于流向特征熵和测地线距离的 K 均值聚类算法来分割结节. 针对 132 个临床 CT 影像的肺结节 (104 个孤立型和 28 个粘连血管型), 12 个 LIDC 集合 1 的肺结节 (4 个孤立型和 8 个粘连血管型) 和 182 个 LIDC 集合 2 的肺结节 (25 个孤立型肺结节和 157 个粘连血管型肺结节), 评估实验结果和影像科医生手工绘制的金标准相比较, 孤立型肺结节分割正确率分别为 100/104 (96.2%), 4/4 (100%), 24/25 (96.0%), 粘连血管型分割正确率分别为 26/28 (92.9%), 7/8 (87.5%) 和 149/157 (94.9%). 实验表明, 该方法能在短的时间内正确地分割孤立型结节和粘连血管型结节且具有好的鲁棒性.

关键词 粘连血管型肺结节 CT 影像肺癌 CAD 流向特征熵 测地线距离 聚类分割

1 引言

据世界卫生组织 (WHO) 统计, 肺癌是致死率最高的癌症. 如果肺癌能被早期发现就可以大大提高治愈率. 肺癌早期是以肺结节的形式表现的, 肺结节通常是指直径小于 3 cm 大于 3 mm 的类球形病灶. CT 比 X 线更适合检测早期肺结节. 肺结节的良、恶性的判别历来是影像学诊断的难题, 医生常用的判别指标是结节的形态、大小和倍增率. 其中, 倍增率反映结节的生长速度, 即对比同一病人不同时间 (疑似结节的随访周期大约 3 到 6 个月) 所拍的两幅影像比较, 如果同一个结节的体积有明显增大, 则为恶性结节.

引用格式: 孙申申, 郭阳, 任会之, 等. 基于流向特征熵和测地线距离的粘连血管型肺结节聚类分割. 中国科学: 信息科学, 2013, 43: 1136–1146, doi: 10.1360/112011-1322

医生在 CT 片上手动测量结节的最大直径是不准确且是耗时的. 肺癌计算机辅助诊断 (computer aided detection, CAD) 系统为自动检测早期肺结节和定量测量肺结节的倍增率提供了新的手段, 它能自动地检测出 CT 影像上的肺结节, 并自动、正确地分割肺结节, 测量它的体积和计算一段时间内结节的倍增率 [1].

从肺结节的形态上, 可以分为孤立型、粘连血管型、粘连肺壁型、毛玻璃型和空洞型肺结节 [2]. 在所有类型的肺结节中, 粘连血管型结节的恶性概率最大, 这是因为血管给癌细胞提供了源源不断的养料, 使结节在很短的时间能迅速生长, 吞噬正常的肺组织细胞. 对于血管和结节相连的部分, 即使是影像科医生也很难区分哪些像素是结节, 哪些像素是血管, 因此, 同一患者用来测量倍增率的前后两序列影像应该由同一位医生 (或同一个 CAD 软件) 进行. 这就需要提出一种既能分割所有尺寸的粘连血管型肺结节, 又在速度上满足临床需要的算法.

肺 CAD 的工作流程是: 首先, 导入肺部 CT 序列影像; 然后, 一键式分割肺实质; 接下来, 一键式检测肺结节, 并把检测出的结节的信息显示到结节信息列表中; 再由医生逐一判断结节的类型; 最后, 程序根据结节类型自动调用不同的分割算法. 而本文算法重点解决了粘连血管型结节的分割问题. 文献 [3] 中的肺结节检测算法能提供结节的中心点, 分割算法是在以检测算法提供的结节中心点为中心的包围核 (volume of interest, VOI) 中执行, 且以插件的形式集成在现有的肺癌计算机辅助诊断工作站上 LungCAD.

传统的最大轴直径测量 (RECIST) 是不可重复且费时的一种手工方法. 早期的肺结节分割方法是 2D 的, 在肺结节的每一层上, 医生手工绘制初始轮廓, 再用动态规划的方法拟合出结节的轮廓 [4]. 随着多层螺旋 CT 的出现, 以近似各向同性的三维分辨率扫描胸部成为可能. 近年来国内外在三维结节分割方向上做了广泛的研究, 尤其是粘连血管型结节和粘连肺壁型结节的分割.

粘连血管型结节的 CT 值通常与血管上的 CT 值接近, 仅考虑灰度的区域生长算法无法正确分割. 基于形态学分割方法的主要缺点是该方法对结构元素的尺寸敏感, 很难找到一个固定的尺寸适合于所有的结节. 如果结构元素的半径太小就不能把血管和结节分开; 如果结构元素的半径太大在把两者分开的同时也把结节的边缘毛刺腐蚀掉了, 而结节的边缘毛刺是判断结节良、恶性的又一重要指标, 毛刺越多说明结节的恶性可能性就越大. 为了弥补这一不足, 文献 [2] 采用递减结构元素尺寸的迭代膨胀方法重新生长出结节边缘毛刺, 但在一定程度上把前一步骤去掉的血管又生长出来了. 文献 [5] 把结节 VOI 转换为 3D 欧氏距离变换图, 并求得结节的中心和半径, 把结节半径作为结构元素的大小, 用形态学开操作分割结节.

3D 主动轮廓的方法被应用在文献 [6] 中进行结节分割, 在 2D 主动轮廓方法的基础上添加了 3D 能量成分, 如引导主动轮廓寻找物体表面的 3D 梯度、3D 曲率和掩模. 文献 [7] 利用测地线距离来表达多阈值灰度图像. 定义基于灰度值和形状的混合区域生长方法进行结节分割, 并寻找平均梯度最大的轮廓线作为结节的边缘. 文献 [8] 利用基于模糊连接度的区域生长方法来分割肺结节, 根据模糊连接度来决定像素是属于结节还是属于血管.

文献 [9] 首先评估了部分容积效应对分割小结节的影响, 然后, 基于多尺度理论的高斯拉普拉斯 (Laplacian of Gaussian, LOG) 模板匹配检测结节的位置和估计结节的尺寸, 从而确定 VOI, 再用文献 [4] 的方法去除血管, 用基于规则的迭代方法去除肺壁, 最后用刚性配准使两序列影像对齐并计算它们的差从而得到倍增率. 文献 [10] 首先把结节的形状拟合成服从各向异性高斯分布的椭球, 然后根据 Jensen-Shannon 分散准则求出最稳定状态下的模型尺度参数. 均值漂移方法估计出每个尺度下的均值和协方差矩阵, 分别代表结节的中心点和各个轴方向上的跨度.

由于血管中有流向同一方向的血液, 所以大部分血管像素的流向特征 (像素梯度的法向量方向) 基本一致, 而结节像素的流向特征杂乱无章^[11]. 这样就使得血管像素的流向特征熵值小而结节像素的流向特征熵值大. 大部分的血管像素到结节中心点的测地线距离都比结节像素到结节中心点大. 基于上述粘连血管型结节的属性, 本文提出了一种基于流向特征熵和测地线距离的 k 均值聚类分割粘连血管型肺结节的方法.

文章的组织结构如下. 第 2 节是方法的描述部分, 详细地介绍了基于流向特征熵和测地线距离聚类分割粘连血管型结节的方法. 之后第 3 节评估该算法的正确性、可重复性、鲁棒性和速度. 第 4 节是结果和讨论部分. 第 5 节是结论.

2 粘连血管型结节体积测量

2.1 肺结节分类

肺结节在 CT 影像上的表象是紧凑且不透明的类球形病灶, 肺结节大部分像素的 CT 值高于它周围的肺实质 CT 值. 肺结节分为以下五类:

(1) 孤立型: 孤立型肺结节是指不与其它肺内解剖结构相连接的, 通常远离肺门和肺壁的结节.

(2) 粘连血管型: 粘连血管型结节是指长在血管末端、血管中间或两血管交叉处的肺结节, 其生长所需要的养料由血管提供, 通常这类结节恶性可能性很大.

(3) 粘连肺壁型: 结节与肺壁相连接, 形态或以细小的尾巴结构, 或以类半球形贴在肺壁表面上. 以上三类结节其共同的 CT 表象是高亮度实心类球状物体.

(4) 毛玻璃型: 呈现模糊稀薄影, 不但是医生最易漏的一类结节同时也是早期筛查检测意义最大的一类结节.

(5) 空洞型: 由于癌细胞已经吞噬掉了正常的肺组织, 在 CT 表象上结节核的中央呈现黑色的空洞, 说明该类结节已经是晚期, 没有早期筛查的必要.

2.2 步骤

算法流程图如图 1 所示. 基于圆点滤波器所得到的结节候选点集合, 自适应地计算包围核的尺寸; 基于大津法阈值的区域生长分割 VOI 得到前景区 (即只含有血管和结节部分); 基于高斯梯度相关矩阵提取流向特征; 并求得流向特征熵图像; 流向特征熵最大的像素为结节中心点; 基于六邻域的区域生长建立前景区的无向图, 并用 Dijkstra 算法求图中各结点 (即前景区各像素) 到结节中心点的最短距离 (即最短路径所经过的结点个数) 组成测地线距离图像. 基于流向特征熵和测地线距离特征的 K -均值聚类分割出粘连血管型肺结节.

2.3 预处理

图 2 是一个粘连血管型肺结节横断面序列影像. 预处理包括用圆点滤波器提取结节候选点集合和自适应地计算 VOI 尺寸. 文献 [11] 把结节看作是球状结构, 把血管看作是管状结构. 具有球形结构的像素, 其 Hessian 矩阵的三个特征值近似相等; 而具有管状结构的像素, 在血管方向上的 Hessian 矩阵特征值远远大于另外两个. 用给定阈值二值化被圆点滤波器增强后的影像, 大于给定阈值的像素被标

图 1 分割粘连血管型结节流程图

Figure 1 Flowing chart for Juxta-vascular nodule segmentation

图 2 一个粘连血管型肺结节横断面序列影像

Figure 2 A CT sequence images of a Juxta-vascular nodule in VOI

记为红色, 它们被看作是结节候选点, 如图 3(a) 所示. 经过“去假阳”步骤后得到结节像素集合. 结节像素集合坐标的平均值为 VOI 的中心.

如果 CT 影像的 x, y 方向像素间距为 x mm, z 方向影像间的层厚为 z mm, 则插值系数 l 为

$$l = \lfloor z/x \rfloor. \quad (1)$$

将候选结节像素集合的中心看作是 VOI 的中心, VOI 的初始尺寸为 $3 * 3 * \lfloor 3/l \rfloor$, 逐渐扩大 VOI 的尺寸直到 VOI 中候选像素个数占整个 VOI 的像素数目的比率 r_c^k 小于给定阈值. 设 W_c^k 为第 k 次迭代以 c 为中心的包围核, H 为结节候选点集合. (2) 式的分子为在包围核内候选结节像素个数. 一个粘连血管型结节 VOI 的最大密度投影 (max intensity projection, MIP) 如图 3(b) 所示. 在 VOI 中用大津法 *ostu* 自适应得到初始阈值, 以 VOI 中心点为种子点区域生长分割得到高亮前景区 (即只包含血管和结节部分). 前景区的最大密度投影示图为图 3(c).

$$r_c^k = \frac{\#\{x \in W_c^k \ \& \ x \in H\}}{(k+2)^2 * \lfloor (k+2)/l \rfloor}. \quad (2)$$

图 3 基于 Hessian 矩阵结节候选点和 MIP 结节 VOI

Figure 3 Candidate points detected by Hessian and MIP nodule VOI

2.4 提取流向特征

用高斯管状模型模拟血管, 血管方向可用血管上的各个像素梯度的法向量来表示. 像素梯度的法向量 v_j 是与该像素的邻域像素梯度都近似垂直的向量.

以像素 j 为中心的 $3 \times 3 \times 3$ 的邻域包围核中, 所有像素的 3 维高斯梯度向量组成集合, 记作 $\zeta_j = \{g_i\}_{i=1}^n$, 其中 $n = 3^3$, $g_i = \{g_{ix}, g_{iy}, g_{iz}\}$. g_{ix} 等于序列图像卷积 x 方向高斯梯度模板 $\frac{-ix}{2\sqrt{2}\pi^{3/2}\sigma^5} \times e^{\frac{-(ix^2+iy^2+iz^2)}{2\sigma^2}}$, g_{iy} 等于序列图像卷积 y 方向高斯梯度模板 $\frac{-iy}{2\sqrt{2}\pi^{3/2}\sigma^5} \times e^{\frac{-(ix^2+iy^2+iz^2)}{2\sigma^2}}$, g_{iz} 等于序列图像卷积 z 方向高斯梯度模板 $\frac{-iz}{2\sqrt{2}\pi^{3/2}\sigma^5} \times e^{\frac{-(ix^2+iy^2+iz^2)}{2\sigma^2}}$. VOI 中所有像素的梯度向量 ζ_j 如图 4(a) 所示.

垂直于 ζ_j 的向量 v_j 被定义为投影到 ζ_j 中所有向量的平方和最小的向量,

$$E(v_j) = \frac{1}{n} \sum_{i=1}^n ((g_i)^T v_j)^2 = v_j^T G G^T v_j,$$

其中, $G = \frac{1}{\sqrt{n}}[g_1, \dots, g_n]$, $G G^T$ 是 3×3 的关系矩阵. 基于最小二乘法求得使 $E(v_j)$ 最小时, v_j 的取值为 $(\sqrt{\lambda_2 + \lambda_3})e_1$, 其中 $\lambda_1 \leq \lambda_2 \leq \lambda_3$, e_1, e_2, e_3 为关系矩阵 $G G^T$ 的特征值和归一化后的特征向量. 虽然特征向量有唯一的方向, 但特征向量是不唯一的. 为了保证血管流向一致, 需要把一些向量的方向转换到其相反的方向. 对于前景区的每个像素都计算向量 v_j 与向量 $v_0 = \{1, 1, 1\}$ 之间的余弦夹角 $\theta_j = \arccos[\frac{v_j \cdot v_0}{|v_j| |v_0|}]$, 组成夹角集合 Θ , 求出 Θ 中出现最多的角度 θ_{most} , 如果 θ_j 与 θ_{most} 之差的绝对值大于 $\pi/2$, 就把 v_j 变化到其相反的方向. 像素 j 梯度的法向量 v'_j , 也就是像素 j 上血流流动向量为

$$v'_j = \begin{cases} -v_j, & |\theta_j - \theta_{\text{most}}| > \pi/2, \\ v_j, & \text{otherwise.} \end{cases}$$

图 4(b) 为一致化后的像素 j 梯度的法向量图. 像素 j 的流向特征定义为 v'_j 与向量 $v_0 = \{1, 1, 1\}$ 之间的余弦夹角 $\theta'_j = \arccos[\frac{v'_j \cdot v_0}{|v'_j| |v_0|}]$.

2.5 计算各像素流向特征熵及结节中心点

像素 j 的流向特征熵被定义为以像素 j 为中心的邻域像素所组成的流向特征集合的熵 ent_j :

$$\text{ent}_j = \text{Entropy}[\{\theta_i\}_{i=1}^n], \quad n \leq (\text{ls})^3,$$

其中 ls 为以像素 j 为中心的邻域尺寸, θ_i 为邻域像素 i 的流向特征. 流向特征熵的直方图 H (采样间隔为 0.05). 最大流向特征熵值所对应的像素坐标 (或坐标的平均值) 为结节中心坐标, 在图 5 中黑色

图 4 VOI 像素梯度向量图及梯度的法向量图

Figure 4 The vector fields of the foreground region of VOI

图 5 流向特征熵图像

Figure 5 The flowing entropy feature images of the foreground region in VOI and the nodule center is marked by center black point

图 6 粘连血管型结节分割结果

Figure 6 Segmentation result of Juxta-vascular nodule by the proposed method

点标示. 生成的流向特征熵图像如图 5. 可以看出, 在结节上像素的流向特征远大于在血管上像素的流向特征.

2.6 测地线距离图像

以结节中心为种子点, 做六邻域区域生长, 把前景区图像转换成无向图 $G(V, E)$. 前景区图像中每个非零像素都看作是图中的结点, 像素 i 与像素 j 间有四邻域关系的所对应的结点 i 与结点 j 之间由边连接, 该无向图的所有边权重为 1.

$$E(V_i, V_j) = \begin{cases} 1, & i \text{ 与 } j \text{ 为四邻域} \\ 0, & \text{otherwise.} \end{cases}$$

用 Dijkstra 算法求出图 G 中各结点间的最短路径, 从而得到前景区各非零像素所对应的结点到结节中心所对应的结点的最短距离. 该距离是测地线的, 因为最短路径所经历的所有像素 (结点) 都在前景区内.

基于流向特征熵特征和测地线距离特征对前景区像素进行 K 均值聚类分割, $K = 2$, 需对流向特征熵特征和测地线距离特征归一化, 结果如图 6 所示.

图 7 人造数据分割结果

Figure 7 A synthetic Juxta-vascular nodule. (a)–(d) the sequence images; (e) its MIP(maximum intensity projection); (f) its flowing direction image

3 实验仿真

3.1 环境及数据

算法是利用 WindowsXP 平台上安装的 VC6.0 和 Mathematica7.0 实现的, 算法运行在 CPU 为 Intel(R) Core(TM)2 Duo CPU T5670 @1.80GHz 1.79GHz, 1G RMA 存储器的 PC 机上.

用两个数据集进行测试: 一个是从合作医院临床获取的含有肺结节 CT 影像所组成的数据库, 是 2007 年扫描 64 个肺 CT 影像序列, 层厚为 2 mm 到 3 mm, 共有 104 个孤立型结节和 28 个粘连血管型肺结节, 由一位医生手工绘制肺结节边缘作为金标准; 另一个是从 lung image database consortium (LIDC) 数据库得到的, 其中, 一部分是 2005 年扫描 23 个肺 CT 影像序列, 层厚为 6.05 mm 到 7 mm, 共有 4 个孤立型结节和 8 个粘连血管型肺结节; 另一部分是 2007 年扫描 86 个肺 CT 影像序列, 层厚为 1~3 mm, 共有 25 个孤立型结节和 157 个粘连血管型肺结节, 在 LIDC 数据库中含有结节坐标金标准的文件, 文件格式为.xml^[12].

第一个数据集的影像是从东芝螺旋 CT 上获取的, 具体的拍摄参数是峰值电压 (peak voltage) 120 kVp、管电流 (tube current) 110 mAs; LIDC 数据集中的影像是从西门子 16 排 CT 上获取的, 具体的拍摄参数是峰值电压 120~140 kVp、管电流 40~388 mAs、校准系数 1.25~3 mm.

3.2 实验仿真分割结果

人造数据集: 用灰度服从高斯分布的椭球模拟结节; 用灰度服从高斯分布的圆柱模拟血管. 把多个圆柱体和一个椭圆体叠加来模拟粘连血管型结节. 图 7 (a) ~ (d) 按上述方案人造的粘连血管型结节断层图像, 图 7 (e) 为最大密度投影 (maximum intensity projection, MIP) 的结果. 图 7 (f) 为流向特征图.

3.3 真实数据分割结果

图 8 列出了不同结节 (包括孤立型和粘连血管型) 正确的分割效果, 呈现了结节各层的 2D 形态

图 8 结节分割结果

Figure 8 Nodule segmentation result

和 3D 可视化效果. 分割结果与医生的金标准相比较, 当重叠度 $\text{overlap} > 90\%$ 时, 认为该结节分割结果正确. R_{gold} 和 R_{new} 分别指金标准的像素集和分割算法的像素集.

$$\text{Overlap} = \frac{R_{\text{gold}} \cap R_{\text{new}}}{R_{\text{gold}} \cup R_{\text{new}}} * 100\%.$$

对于 132 个孤立型肺结节, 用所提出的方法分割的正确率是 100%; 对于 194 个粘连血管型肺结节, 用所提出的方法分割结果的正确率是 92.2%. 平均结节分割时间是 20 s, 结节半径越大, 算法执行时间就越长.

3.4 参数讨论

1) 对于初始阈值敏感性.

本文算法的前提是前景区中只含有结节和血管, 而且结节是完整的. 当分割前景区所用的初始阈值过大或过小, 是违背这个前提的. 图 9 (a2) 为初始阈值过小, 前景区没有包含完整的结节. 图 9 (d2)

图 9 不同初始阈值对应前景区域和最终分割结果

Figure 9 Corresponding to the foreground regions by the different initial thresholds and the final segmentation result

图 10 图 8(c) 中, 当 l_s 取不同值时, 结节的分割结果

Figure 10 When l_s is taking different values, nodule segmentation result of Figure 8(c)

为初始阈值过大, 前景区把除结节和血管外的背景区域也包含进来了. 这两种情况都不符合前提条件, 不在本文考虑的范畴内. 取不同的初始阈值 (-572 Hu, -588 Hu), 最终分割结果相同, 如图 9 (b2) 和图 9 (c2).

2) 对于流向特征熵的邻域尺度敏感性.

评估流向特征熵的邻域尺度 l_s 对分割结果的影响. 图 10 为图 8(c) 在不同流向特征熵的邻域尺度 l_s 下的分割结果, 图 10 (a) ~ (g) 分别对应 $l_s = 1, \dots, 7$. 当 $l_s = 1, 5, 6, 7$ 时, 血管被分割到结节区域内.

3) 噪声对分割结果影响.

向粘连血管型结节前景区域影像不断地加入以尺度为 σ 的高斯噪声后, 再用所提出的算法进行结节分割 [13]. 除非 σ 增大到血管被腐蚀掉的程度, 否则该算法的分割结果均在可接受的范围内.

4 结果与讨论

从图 9 的断层图像和三维可视化图像可以看出, 本文所提出的结节分割算法可以分割孤立型结节和粘连血管型结节, 尤其能把血管和结节分割开. 从图 10 可以看出流向特征熵的邻域尺度 l_s 取太大或者太小都会保留一部分血管; 且 l_s 取值和结节体积有关; 结节体积大的, l_s 的取值也应该大; 结节体

表 1 结节体积与 σ 的关系Table 1 The relationship between nodule volume and σ

σ	0	0.25	0.5	0.75	0.9
Volume	543	477	609	595	643

积小的, ls 的取值也应该小. 一般 ls 取 2 或 3, 能得到比较好的分割结果. 从表 1 可以看出, 该结节分割算法对高斯噪声不敏感, 这是因为所提出的分割算法不是用灰度特征进行血管和结节的区分.

5 结论

本文提出了基于流向特征熵和测地线距离特征的聚类方法来分割孤立型结节和粘连血管型结节. 该算法能在短的时间内, 把结节和血管分开, 且该算法有好的鲁棒性.

致谢 感谢 Bart M.ter Haar Romeny 教授在技术上的指导.

参考文献

- 1 Sluimer I, Schilham A, Prokop M, et al. Computer analysis of computer tomography scans of the lung: a survey. *IEEE Trans Med Imaging*, 2006, 25: 385–405
- 2 Kostis W J, Reeves A P, Yankelevitz D F, et al. Three-dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical CT images. *IEEE Trans Med Imaging*, 2003, 22: 1259–1273
- 3 Sun S S, Ren H Z, Kang Y, et al. Lung nodule detection by GA and SVM. *J Syst Simul*, 2011, 23: 497–502 [孙申申, 任会之, 康雁, 等. 基于遗传算法和支持向量机的肺结节检测. *系统仿真学报*, 2011, 23: 497–502]
- 4 Xu N, Ahuja N, Bansal R. Automated lung nodule segmentation using dynamic programming and EM based classification. *P SPIE*, 2002, 4684: 666–676
- 5 Kuhnigk J-M, Dicken V, Bornemann L, et al. Morphological segmentation and partial volume analysis for volumetry of solid pulmonary lesions in thoracic CT scans. *IEEE Trans Med Imaging*, 2006, 25: 417–434
- 6 Way T, Hadjilski L, Sahiner B, et al. Computer-aided diagnosis of pulmonary nodules on CT scans: segmentation and classification using 3D active contours. *Med Phys*, 2006, 33: 2323–2337
- 7 Diciotti S, Picozzi G, Falchini M, et al. Natale villari guido valli, 3-D segmentation algorithm of small lung nodules in spiral CT images. *IEEE Trans Med Imaging*, 2008, 12: 7–19
- 8 Dehmshki J, Amin H, Valdivieso M, et al. Segmentation of pulmonary nodules in thoracic CT scans: a region growing approach. *IEEE Trans Med Imaging*, 2008, 27: 467–479
- 9 Diciotti S, Lombardo S, Coppini G, et al. The LOG characteristic scale: a consistent measurement of lung nodule size in CT imaging. *IEEE Trans Med Imaging*, 2010, 29: 397–409
- 10 Okada K, Comaniciu D, Krishnan A. Robust anisotropic Gaussian fitting for volumetric characterization of pulmonary nodules in multislice CT. *IEEE Trans Med Imaging*, 2005, 24: 409–423
- 11 Agam G, Armato Samuel G III, Wu C H. Vessel tree reconstruction in thoracic CT scans with application to nodule detection. *IEEE Trans Med Imaging*, 2005, 24: 486–499
- 12 Armato S G 3rd, McLennan G, Bidaut L, et al. The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. *Med Phys*, 2011, 38: 915–931
- 13 Fu S J, Zhang C M, Tai X C. Image denoising and deblurring: non-convex regularization, inverse diffusion and shock filter. *Sci China Inf Sci*, 2011, 54: 1184–1199

Juxta-vascular nodule segmentation based on the flowing entropy and geodesic distance feature

SUN ShenShen^{1,2}, GUO Yang³, REN HuiZhi⁴, FAN LiNan¹ & KANG Yan^{2*}

1 College of Information Engineering, Shenyang University, Shenyang 110044, China;

2 Sino-dutch Biomedical and Information Engineering School, Northeastern University, Shenyang 110819, China;

3 Department of mathematics, Faculty of Science, Northeastern University, Shenyang 110819, China;

4 Institute of mechanical engineering, Shenyang University of Technology, Shenyang 110870, China

*E-mail: kangy@neusoft.com

Abstract Computed aided diagnosis (CAD) of lung CT is a new quantitative analysis imaging technique to distinguish malignant nodules from benign ones. Nodule growth rate is a key indicator for judgment of benign or malignant nodule. Accurate nodule segmentation is the premise condition of calculating nodule growth rate. However, it is difficult to segment Juxta-vascular nodules, due to similar gray levels between nodule and vessel. To distinguish the nodule region from the adjacent vessel region, a flowing direction feature, referred to as the direction of normal vector for a pixel, is introduced. Since all blood has the same flowing direction through a vessel, the normal vectors of pixels in the vessel region typically point to similar orientations while the directions of those in the nodule region can be viewed as disorganized. So the entropy value of flowing direction features in a neighbor region for a vessel pixel is bigger than that for a nodule pixel. And vessel pixels typically have a larger geodesic distance to the nodule center than nodule pixels. Based on k-Means clustering method, the flowing entropy feature, combined with geodesic distance feature, is proposed to solve the segmentation problem of the vessel attachment nodule. The validation of the proposed segmentation algorithm was carried out on Juxta-vascular nodules (104 solid nodule and 28 Juxta-vascular nodule), identified in the Chinalung-CT screening trail and on the lung image database consortium (LIDC) dataset. Among them, there are 12 nodules in the first LIDC database (4 solid nodules and 8 Juxta-vascular nodules) and 182 nodules in the second LIDC database (25 solid nodules and 157 Juxta-vascular nodules). Comparison is done between the gold standard and experimental results. The correct segmentations on solid nodule are 100/104(96.2%), 4/4(100%) and 24/25(96.0%), respectively, while the correct segmentations on Juxta-vascular nodule are 26/28(92.9%), 7/8(87.5%) and 149/157(94.9%), respectively, showing that the proposed method has low time complexity and high accurate rate.

Keywords vessel attachment nodule, CT image lung cancer CAD, flow feature, geodesic distance feature, cluster segmentation

SUN ShenShen was born in 1980. She received her Ph. D. degree in Computer Science from the Northeastern University, Shenyang, in 2009. Currently, she is a lecturer at Shenyang University. Her research interest includes medical image processing.

KANG Yan was born in 1978. He received his Ph. D. degree in Medical Imaging Industrial Engineering from the Erlangen-Nuernberg University, Shenyang, in 2002. Currently, he is dean of Sino-Dutch Biomedical and Information Engineering School at Northeastern University. His research interest includes medical image processing.