

doi: 10.12068/j.issn.1005-3026.2018.02.007

基于CT图像3D特征的肺结节检测

王彬, 赵海, 朱宏博, 朴春赫

(东北大学 计算机科学与工程学院, 辽宁 沈阳 110169)

摘要: 为了提高肺结节检测的性能, 提出一种基于中心点连续性的肺结节检测算法. 该算法使用基于简单线性迭代聚类超像素方法分割CT图像, 并根据相似度合并超像素, 进而得到肺部区域及疑似肺结节区域, 降低了疑似肺结节的漏检率. 根据各帧CT图像中疑似肺结节区域的中心点偏移程度评价其中心点连续性, 最终判断出阳性肺结节. 文中的实验数据来自于上海市胸科医院和LIDC数据库. 实验结果表明, 改进后算法的敏感度达到86.36%, 假阳性率为1.76.

关键词: 肺结节; 肺结节检测; 肺结节分割; 超像素; SILC; 中心点连续性

中图分类号: TP 391.41

文献标志码: A

文章编号: 1005-3026(2018)02-0181-05

Pulmonary Nodules Detection Based on 3D Features from CT Images

WANG Bin, ZHAO Hai, ZHU Hong-bo, PAK Chun-hyok

(School of Computer Science & Engineering, Northeastern University, Shenyang 110169, China. Corresponding author: WANG Bin, E-mail: 778999451@qq.com)

Abstract: A detection method based on the center continuity is proposed to improve the performance of pulmonary nodules detection. CT images are segmented by using SLIC (simple linear iterative clustering) superpixel algorithm in the method. Superpixels are merged according to the similarity to extract pulmonary areas and suspected pulmonary nodule areas, which reduces the missing rate of suspected pulmonary nodules. Suspected pulmonary nodules are diagnosed as positive which keeping center continuous in 3D space. All of CT images in experiments are obtained from Shanghai chest hospital and LIDC database. The experimental results of the improved algorithm show that sensitivity is 86.36% and false positive is 1.76.

Key words: pulmonary nodule; pulmonary nodule detection; pulmonary nodule segmentation; superpixel; SLIC (simple linear iterative clustering); center continuity

CT图像是医生诊断肺部病灶的首要依据. 目前, 大部分针对CT图像的计算机辅助检测(CAD)方法均基于2D特征; 但由于肺部的支气管、血管等组织与肺结节的2D特征相似, 对检测产生干扰, 所以基于2D特征的检测效果并不理想^[1-3], 而将3D特征应用于肺结节检测已经成为胸部CT图像研究的主流. 目前有许多已发表的与肺结节3D特征相关的文献^[4-5]. 本文针对肺结节在3D空间具有中心点连续性的特点, 在分割肺部区域及疑似肺结节区域的基础上, 计算

疑似肺结节区域的质量中心点, 并评价其偏移程度, 最终确定该区域是否为阳性肺结节. 本文在实验中使用的CT图像均来自上海市胸科医院和LIDC数据库.

1 基于2D特征的疑似肺结节检测

3D特征分析需要处理相邻的多张图像, 其时间复杂度为 $O(m \cdot n)$, m 表示图像数量, n 表示肺结节数量^[6]. 为了降低计算量, 有时需要利用肺

收稿日期: 2016-08-10

基金项目: 中央高校基本科研业务费专项资金资助项目(N140405004, N161608001); 辽宁省科技厅重点实验室建设基金资助项目(LZ201015).

作者简介: 王彬(1988-), 男, 辽宁沈阳人, 东北大学博士研究生; 赵海(1959-), 男, 辽宁沈阳人, 东北大学教授, 博士生导师.

结节的 2D 特征确定疑似肺结节区域.

1.1 超像素相似度计算

本文使用基于简单线性迭代聚类 (simple linear iterative clustering, SLIC) [7] 方法对 CT 图像分割超像素, 见图 1.

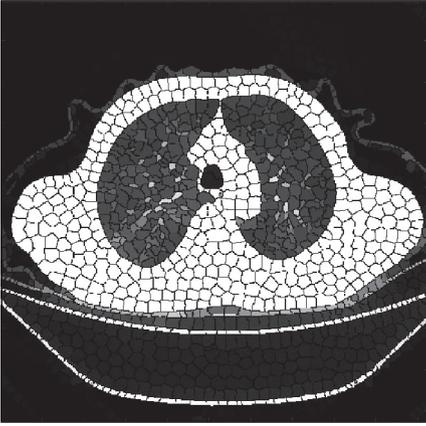


图 1 胸部 CT 影像的 SLIC 超像素分割结果
Fig. 1 Result of superpixel segmentation based on SLIC for chest CT image

超像素在分割之后按照式(1)归一化:

$$\text{sp_value}(i) = \frac{\sum I^{(i)}(x,y)}{N^{(i)}}, i = 1, \dots, N_{\text{sp_total}}. \quad (1)$$

式中: $\text{sp_value}(i)$ 为超像素 i 的像素值; $N^{(i)}$ 为超像素 i 包含像素的数量; $I^{(i)}(x,y)$ 为超像素 i 中位于 (x,y) 处像素点的像素值; $N_{\text{sp_total}}$ 为超像素的数量.

超像素 i 和 j 之间的相似度定义为

$$\text{similarity}'(i,j) = \frac{\min(\text{sp_value}(i), \text{sp_value}(j))}{\max(\text{sp_value}(i), \text{sp_value}(j))}. \quad (2)$$

将超像素 i 和 j 之间的拓扑关系与式(2)结合, 得到相似度的定义如下:

$$\text{similarity}(i,j) = \begin{cases} \text{similarity}'(i,j), & \text{neighbor}(i,j) = \text{true}; \\ 0, & \text{neighbor}(i,j) = \text{false}. \end{cases} \quad (3)$$

式中: $\text{neighbor}(i,j) = \text{true}$ 表示超像素 i 和 j 相邻, $\text{neighbor}(i,j) = \text{false}$ 表示超像素 i 和 j 不相邻.

1.2 基于超像素相似度合并的肺分割

通常相邻超像素在空间和颜色方面具有一定的相似性, 合并相似超像素能够突出 CT 图像中的感兴趣区域. 相似的像素点合并, 而差异较大的则会分裂, 超像素的合并与分裂由边界线上的像素点控制. 经过多次迭代合并就会提取出肺部区域, 提取的结果如图 2 所示.



图 2 胸部图像分割结果
Fig. 2 Result of chest CT image segmentation

由于肺部轮廓提取时仅考虑超像素的相似度, 最终得到的轮廓可能是非连接或者弱连接的. 为了解决这个问题, 合并超像素时需要利用边界能量, 这样可以强化存在连接的图像轮廓. 把相邻超像素分界线上所有点的平均分割能量作为对应超像素之间的分割能量, 并按照这样的规则合并所有的超像素. 随着合并的进行, 能量值低的分割线逐渐淡化, 能量值高的分割线逐渐突出, 最终可以形成连续的图像轮廓.

分割能量的计算公式为

$$\text{energy}^{(i,j)} = \frac{\sum_{n=1}^{N^{(i,j)}} E_n^{(i,j)}}{N^{(i,j)}}, \quad (4)$$

$$E_n^{(i,j)} = \left| \sum_{(x,y)_n \in i} I(x,y) - \sum_{(x,y)_n \in j} I(x,y) \right|. \quad (5)$$

式中: $\text{energy}^{(i,j)}$ 是两个超像素区域 i 和 j 之间的分界线能量; $N^{(i,j)}$ 是超像素 i 和 j 的分界线像素点总数. $E_n^{(i,j)}$ 的定义如式(5)所示, 表示超像素 i 和 j 分界线上的第 n 个像素点的分割能量, 其中 $I(x,y)$ 表示位置在 (x,y) 处的像素点的像素值; $(x,y)_n$ 表示像素点 (x,y) 包含于超像素分界线上第 n 个像素点的邻域. 超像素之间分离线的分割能量更新过程如式(6):

$$\text{energy}^{(k,j')} = \frac{\sum_n E_n^{(k,i)} + \sum_n E_n^{(i,j)}}{N^{(k,i)} + N^{(i,j)}}. \quad (6)$$

式中: 区域 j' 为区域 i 和区域 j 合并后新产生的区域; $\text{energy}^{(k,j')}$ 为区域 k 和区域 j' 之间的分割能量. 反复迭代此合并过程直到所有超像素边界的分割能量都大于设定阈值, 则会得到完整分割区域. 图 3a 为上述过程的分割结果, 生成的分割轮廓既连接又简单. 图 3b 为最终的肺分割结果.

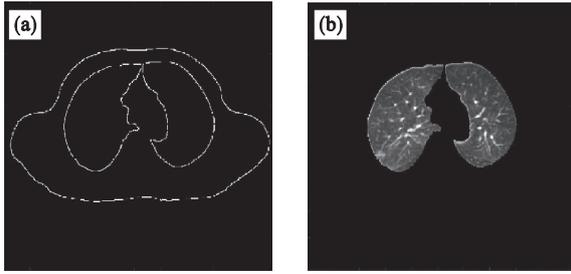


图3 基于边界能量的分割结果

Fig. 3 Segmentation result based on boundary energy
(a)—分割轮廓; (b)—肺分割结果.

1.3 基于超像素的疑似肺结节区域分割

将分割出的肺部 CT 图像分解成超像素,能够将疑似肺结节区域与正常器官划分到不同的超像素中,这可以使疑似肺结节检测更具有针对性.此外,基于 SLIC 的超像素分割算法时间复杂度较小,且能控制超像素的数量和边缘贴合程度,因此本节仍然使用该方法分割疑似肺结节.

图 4a 为 SLIC 超像素合并的结果,亮度较高区域为疑似肺结节部分.图 4b 为最终的提取结果,图中红色边框区域为疑似肺结节区域.通过图 4b 可以发现,很多正常肺部组织也包含在疑似肺结节区域中,这些区域需要通过进一步处理来排除.

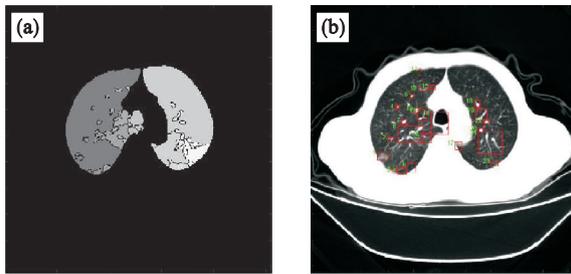


图4 利用超像素分割提取疑似肺结节

Fig. 4 Extraction of suspected nodule area with superpixel segmentation

(a)—超像素合并结果; (b)—最终提取结果.

2 基于形状变化的肺结节检测

在 CT 图像中,肺结节的 2D 特征主要包含边缘形态和边缘周长.在单一 CT 图像切片中,血管和气管等肺组织可能出现与肺结节相似的特点,对检测产生干扰.而在 3D 空间中,肺结节的形状与球体接近,血管和气管等组织呈现延伸形态,两者特点明显不同.所以利用 3D 特征区分肺结节与肺部组织的可靠性更高.

首先,不同大小的肺结节跨越的 CT 图像帧

数也不同.例如,如果肺结节直径为 3 mm,CT 扫描距离 2.5 mm,则该肺结节的连续帧数则不应超过 2.为了评价这种连续性,本文使用连续性系数 (continuity_pos) 表示肺结节的单向连续帧数. continuity_pos 的计算方法如式(7)所示:

$$\text{continuity_pos} = \begin{cases} 1, & \text{size} < 3; \\ 2, & 3 \leq \text{size} \leq 5; \\ 3, & \text{size} > 5. \end{cases} \quad (7)$$

式中:continuity_pos 表示包含同一个肺结节的单侧邻域范围;size 表示疑似肺结节区域包含的像素点数. size 的计算方法如式(8)和式(9)所示:

$$A = \sum_{i=1}^m \sum_{j=1}^n B^{(k)}(i, j), \quad (8)$$

$$\text{size} = \frac{2}{\sqrt{\pi}} \sqrt{A}. \quad (9)$$

式中: m 和 n 分别为第 k 个超像素区域的行数和列数; $B^{(k)}(i, j)$ 为第 k 个超像素的二值化区域中位于 (i, j) 处像素点的像素值.

在 3D 空间中,中心点连续性是判断物体形态的重要指标.球体(如肺结节)的中心点在连续多帧的图像中保持不变,而形状不规则的物体(如血管或支气管)中心点通常变化较大.此外,计算中心点连续性的时间复杂度较低.

图 5 展示了肺结节与支气管在中心点连续性上的差异.图 5 为连续 4 张 CT 图像,顺序为胸部内由下到上.图中的红色边框区域为支气管截面图像,这些区域在 2D 空间中判定为疑似肺结节区域.通过位置的变化可以看出,红色边框的分布从分散到聚合,表现为空间延伸的树形结构^[8].图 5 中的绿色边框区域为肺结节截面图像,在 4 个子图中始终保持中心点位置一致,与红色边框的特点明显不同^[9].本文利用质量中心点连续性描述这种差异.质量中心点 (C_x, C_y) 表示疑似肺结节 CT 截面的中心位置:

$$\left. \begin{aligned} C_x &= \frac{\sum \sum x \cdot f_b^{(k)}(x, y)}{\sum \sum f_b^{(k)}(x, y)}, \\ C_y &= \frac{\sum \sum y \cdot f_b^{(k)}(x, y)}{\sum \sum f_b^{(k)}(x, y)}. \end{aligned} \right\} \quad (10)$$

式中, $f_b^{(k)}(x, y)$ 为疑似肺结节检测二值化结果的第 k 个超像素中 (x, y) 点的像素值.为了降低算法的时间复杂度, (C_x, C_y) 的计算只针对疑似肺结节.通常肺结节的 (C_x, C_y) 在 continuity_pos 范围内比较稳定,而血管或支气管则会呈现明显偏移.

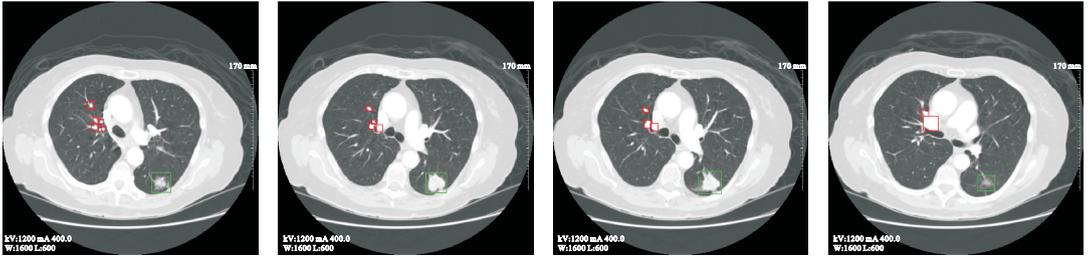


图 5 CT 图像中肺结节与支气管的连续性对比

Fig. 5 Comparison of position continuity between pulmonary nodule and bronchus in CT image

3 实验分析

本文实验的硬件环境:CPU 为 Intel Core i7-4790 k,内存 8 G. 程序的运行环境是 Windows 7 操作系统下的 Matlab R2013a. 对于实验的数据集,本文采用上海市胸科医院提供的 200 组病例和国际公认的 LIDC 数据集 1 000 组.

3.1 中心连续性实验

图 6 为肺结节的连续多帧标志图像,并用白色区域表示. 这 3 组图像是对 LIDC 数据库中的 xml 文件处理后得到的.

通过图 7 中 3D 模型的描述,可以很容易地区分肺结节与血管、支气管及胸膜等肺组织,而其中最明显的特征就是肺结节具有质量中心连续性.

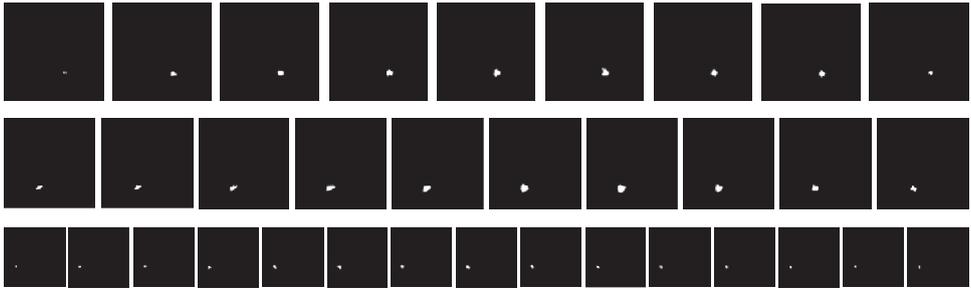


图 6 肺结节质量中心点的连续标志图像

Fig. 6 Sequent images of pulmonary nodule mass center

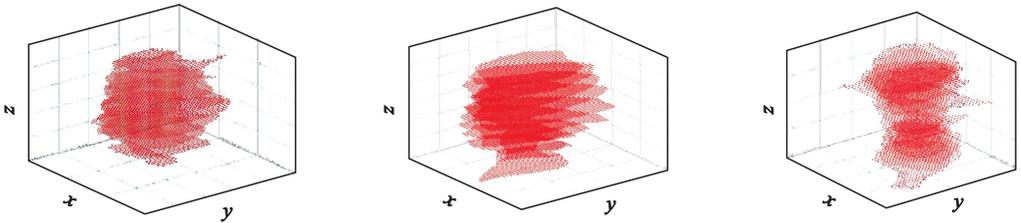


图 7 肺结节的 3D 形状

Fig. 7 3D shape of pulmonary nodule

利用中心点偏移程度判断肺结节时,选择不同的阈值会对结果造成一定的影响.

图 8 所示为选定不同阈值的结果对比,实验所用 CT 图像的分辨率为 512×512 .

敏感度描述的是算法检测出的肺结节占有所有肺结节的比例,而假阳性率 (F_p) 描述的是每个病人的 CT 图像集中误检的肺结节数量. 在图 8 中,相同 F_p 下阈值为 5 时敏感度最高. 所以,对于分辨率为 512×512 的 CT 图像,阈值的最优值为 5 个像素点.

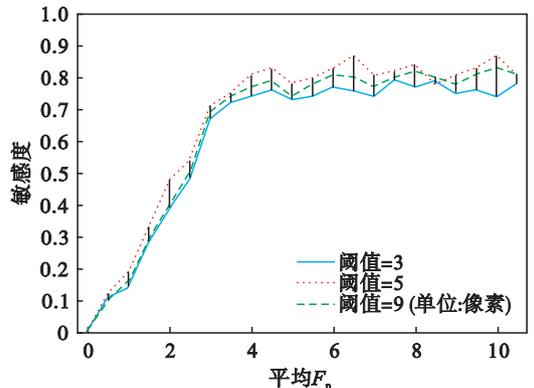


图 8 不同阈值结果对比

Fig. 8 Comparison of results with different thresholds

图 9 所示为利用肺结节的质量中心点连续性得到的肺结节检测结果,算法判定的阳性肺结节

用黄色边框的矩形区域表示。

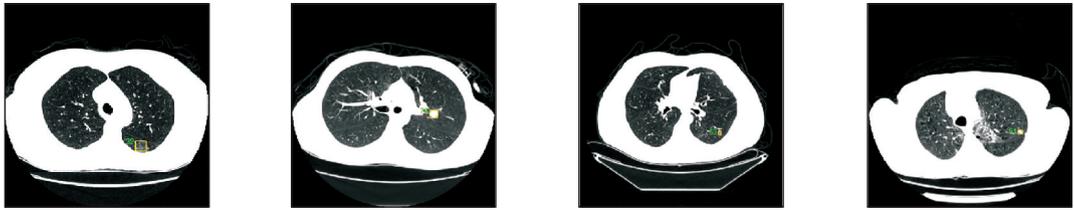


图 9 基于中心点连续性的肺结节检测结果

Fig. 9 Result of detected malignance pulmonary nodules based on center continuity

3.2 肺结节检测实验

本节选择文献[4],文献[5]和文献[10]算法与本文算法进行对比.本文的数据来源于上海市胸科医院和 LIDC 数据库,从中选取 190 个病人样本,共包含 286 个肺结节.

根据表 1 的数据对比,本文算法的敏感度仅低于文献[5]的算法,且明显高于其他两种算法.文献[5]的高敏感度主要来自两方面,一是采用维度更高的特征来描述肺结节,这会增加算法的计算量;二是降低阈值,这样可能会提高假阳性率,而本文算法的 F_p 为 1.76,在 4 种算法中最低.所以,本文的算法在保证假阳性率最低的情况下,使敏感度达到最大.实验结果证明,利用肺结节在 3D 空间上的中心点连续性检测肺结节是一种有效的方法.

表 1 肺结节检测算法结果对比

Table 1 Comparison of algorithms for pulmonary nodule detection

算法	病人样本数量	肺结节数量	肺结节检测数量	敏感度 %	F_p
文献[10]	125	259	172	66.40	3.00
文献[5]	100	123	120	98.03	2.45
文献[4]	60	211	169	80.00	3.90
本文算法	190	286	247	86.36	1.76

4 结 语

本文针对 2D 空间中肺结节检测准确率低的问题,提出了基于 3D 空间中心点连续性的肺结节检测算法.通过与其他算法的比较,本文算法在敏感度和 F_p 方面均表现良好,但是仍然存在一些问题需要解决,其中主要问题为超像素分割的结果中包含的假阳性肺结节较多,增加了算法的时间复杂度.

参考文献:

- [1] Brandman S, Ko J P. Pulmonary nodule detection, characterization, and management with multidetector computed tomography[J]. *Journal of Thoracic Imaging*, 2011, 26(2): 90-105.
- [2] Choi W J, Choi T S. Automated pulmonary nodule detection system in computed tomography images: a hierarchical block classification approach[J]. *Entropy*, 2013, 15(2): 507-523.
- [3] El-Baz A, Beache G M, Gimel'farb G, et al. Computer-aided diagnosis systems for lung cancer: challenges and methodologies[J/OL]. [2016-06-20]. https://www.researchgate.net/publication/235691825_Computer-Aided_Diagnosis_Systems_for_Lung_Cancer_Challenges_and_Methodologies.
- [4] Alilou M, Kovalev V, Snezhko E, et al. A comprehensive framework for automatic detection of pulmonary nodules in lung CT images[J]. *Image Analysis & Stereology*, 2014, 33(1): 13-27.
- [5] Demir Ö, Çamurcu A Y. Computer-aided detection of lung nodules using outer surface features[J]. *Biomedical Materials and Engineering*, 2015, 26(sup1): 1213-1222.
- [6] Foncubierta-Rodríguez A, Müller H, Depeursinge A. Retrieval of high-dimensional visual data: current state, trends and challenges ahead[J]. *Multimedia Tools and Applications*, 2014, 69(2): 539-567.
- [7] Achanta R, Shaji A, Smith K, et al. SLIC superpixels compared to state-of-the-art superpixel method[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(11): 2274-2282.
- [8] Krishnamurthy S, Narasimhan G, Rengasamy U. Three-dimensional lung nodule segmentation and shape variance analysis to detect lung cancer with reduced false positives[J]. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 2016, 230(1): 58-70.
- [9] Li X, Wang X, Dai Y, et al. Supervised recursive segmentation of volumetric CT images for 3D reconstruction of lung and vessel tree[J]. *Computer Methods and Programs in Biomedicine*, 2015, 122(3): 316-329.
- [10] Tan M, Deklerck R, Jansen B, et al. A novel computer-aided lung nodule detection system for CT images[J]. *Medical Physics*, 2011, 38(10): 5630-5645.