

# 支持向量机与区域增长相结合的CT图像并行分割

刘露<sup>1</sup> 楚春雨<sup>1</sup> 马建为<sup>1,2</sup> 刘宛予<sup>2</sup>

(哈尔滨理工大学自动化学院 哈尔滨 150080)<sup>1</sup>

(哈尔滨工业大学 HIT-INSA 中法生物医学图像联合研究中心 哈尔滨 150001)<sup>2</sup>

**摘要** 针对经典区域增长算法中生长规则确定的困难和单纯使用支持向量机分割速度慢的问题,提出了一种支持向量机与区域增长相结合的图像并行分割方法。首先,从已知分割结果的图像中选取一定数量的目标区域与非目标区域样本点作为支持向量机分类器的训练样本并训练支持向量机,然后利用训练好的支持向量机自动寻找种子点并进行区域增长,在区域增长过程中使用支持向量机分类器作为增长规则,最后,针对边缘和噪声像素点进行必要的后处理。测试实验获得了较好的分割效果和较快的分割速度且能实现自动分割,表明所提出的方法是可行有效的。

**关键词** 支持向量机(SVM),区域增长,CT图像,并行分割

中图分类号 TP391.41 文献标识码 A

## CT Image Segmentation Based on Support Vector Machine and Regional Growth

LIU Lu<sup>1</sup> CHU Chun-yu<sup>1</sup> MA Jian-wei<sup>1,2</sup> LIU Wan-yu<sup>2</sup>

(School of Automation, Harbin University of Science and Technology, Harbin 150080, China)<sup>1</sup>

(HIT-INSA Sino-French Research Center for Biomedical Imaging, Harbin Institute of Technology, Harbin 150001, China)<sup>2</sup>

**Abstract** In order to solve the difficulty of determining the growth rules in conventional regional growth algorithm and the slowly of support vector machine segmentation algorithm, an image segmentation method combined support vector machine and regional growth was proposed. Firstly, selected a certain numbers of sample point from target area and non-target area and trained the support vector machine classification, then used the trained classification search seed point and regional growing, the support vector machine classification was used as growth rules, the last, some necessary retrogressing were used for the edge and noise. The experimental results show that this algorithm is feasible and it performs better than conventional region growth segmentation algorithm and faster than conventional support vector machine segmentation algorithm.

**Keywords** Support vector machine, Regional growth, CT image, Parallel segment

## 1 引言

医学图像分割技术是医学图像处理和分析中的关键技术。图像分割技术的发展不仅影响到医学图像处理中其它相关技术的发展,如可视化、3D重建、不同模式医学图像的配准和融合等,而且在生物医学图像的分析中也占有极其重要的地位。医学图像从本质上是模糊的<sup>[1]</sup>,在医学图像中存在许多不确定的因素,如灰度、纹理和区域的边界等。另外,通常医学图像数据量较大,因此,在医学图像分割过程中不仅要考虑分割效果,同时还要考虑分割速度问题。正是由于医学图像的这些特殊性,使得医学图像分割问题缺乏普遍适用的理论和方法,没有得到完美的解决。

针对经典的区域增长算法的缺陷以及单纯使用支持向量

机的分割方法速度慢的问题<sup>[2]</sup>,提出了一种支持向量机与区域增长相结合的医学图像并行分割方法。该方法分为4个步骤:首先,从已知分割结果的图像中选取一定数量的目标区域与非目标区域样本点作为支持向量机分类器的训练样本并训练支持向量机;其次,以一定的粒度 $N$ ,即每隔 $N$ 行 $N$ 列取一个像素点,利用训练好的支持向量机判别该点是否属于目标区域,如果该点属于目标区域,则将该点记为种子点;再次,以这些种子点做起点进行区域增长,使用支持向量机判别结果作为区域增长的终止条件;最后,针对边缘和噪声像素点进行必要的后处理。

实验证明本文提出的分法是合理可行的。这种方法与经典的区域增长算法相比,避免了手动选取种子点的交互操作和增长终止条件的选择困难问题,与单纯使用支持向量机的

到稿日期:2009-12-18 返修日期:2010-02-25 本文受国家国际科技合作重大专项(No. 2007DFB30320),黑龙江省教育厅科技计划项目(No. 11531048)和哈尔滨市科技创新人才研究专项资金项目(No. 2008RFQXS062)资助。

刘露(1972-),男,副教授,硕士生导师,主要从事医学图像处理与模式识别方面的研究, E-mail: liulu@hrbust.edu.cn; 楚春雨(1986-)男,硕士生,主要从事医学图像处理与模式识别方面的研究; 马建为(1977-),男,博士生,讲师,主要从事医学图像并行算法与医学网格应用方面的研究; 刘宛予(1964-),男,教授,博士生导师,主要从事医学图像处理、生物医学仪器、现代公路检测技术、在线检测技术、机器视觉与图像处理等方面的研究。

分割方法相比减少了支持向量机判别的次数,提高了其分割速度。另外,此方法还具有容易并行化的特点。我们在 EGEE(Enable Grids for Escience)网络环境下对此方法进行了并行化实现,进一步提高了其分割的速度。

## 2 方法描述

### 2.1 支持向量机及其图像分割原理

支持向量机(Support Vector Machine, SVM)算法是一类基于统计学习理论的构造性学习技术<sup>[3]</sup>,能够克服传统的学习技术在处理高维问题时所遇到的维数灾难,并且具有泛化能力强的优点。

SVM以结构风险最小化为原则,其优化目标是 minimized 泛化误差限,故具有很好的泛化能力。

在模式分类过程中,SVM先将输入样本  $x$  通过非线性映射  $\varphi(x)$  映射到高维空间  $H$ ,然后在映射特征空间内进行线性分类。SVM的分类函数可以表示为如下形式:

$$f_{SVM}(x) = w^T \varphi(x) + b \quad (1)$$

式中,参数  $w$  和  $b$  在训练样本集合上最小化。

结构风险泛化函数:

$$J(w, \xi) = \frac{1}{2} w^2 + C \sum_{i=1}^N \xi_i \quad (2)$$

$$s. t. y_i f_{SVM}(x_i) \geq 1 - \xi_i, \xi_i \geq 0 (i=1, 2, \dots, N)$$

式中,  $C$  是根据具体问题选择的正参数,  $\xi_i$  是松弛变量。式(2)中的风险泛函是在经验风险(第2项训练误差)和模型复杂性(第1项)之间的折中考虑,用参数  $C$  控制折中程度,可以避免过度拟合,即避免分类能力变差。

满足  $y_i f_{SVM}(x_i) \leq 1$  的训练样本  $(x_i, y_i)$  称为支持向量。引入核函数  $K(x, z) \equiv \phi(x) T \phi(z)$  将决策函数式(1)改写为如下形式:

$$f_{SVM}(x) = \sum_{i=1}^N a_i K(x, s_i) \quad (3)$$

式中,  $s_i (i=1, 2, \dots, N)$  表示支持向量。

利用核函数  $K(\cdot, \cdot)$  可以直接从式(3)中得到决策函数,而不需要考虑潜在的映射  $\phi(\cdot)$ 。

本文使用了 Gaussian 径向基函数(RBF)作为核函数:

$$K(x, z) = \exp(-\|x - z\|^2 / 2\sigma^2) \quad (4)$$

式中,  $\sigma \geq 0$  是控制核宽度的参数。

采用 SVM 对图像进行分割,首先是选择样本点生成对应的特征向量进行学习,其次计算待分割图像的每个像素点的特征向量并进行分类。其步骤如下:

(1)选取目标区域和非目标区域的样本点,并标记为两个类别;

(2)选取每个样本点所在的小区域(如  $5 \times 5$  邻域)提取特征形成特征向量  $x_i$ ,将它与类别标记数据  $y_i$  一起构成支持向量机的训练样本;

(3)使用得到的训练样本训练支持向量机;

(4)使用训练好的支持向量机对待分割图像的每个像素点进行判别。

### 2.2 区域增长分割方法的原理

区域增长图像分割方法的基本思想是将具有相似性质的像素集合起来构成区域。基本的方法是从一组“种子”点开始将与种子性质类似的相邻像素附加到生长区域的每个种子上。该方法有4个关键的问题:1)选取合适的种子点;2)定

义恰当的相似性准则;3)定义连通或相邻信息;4)定义一个区域生长终止规则。种子点的选择通常根据所要解决问题的性质,选择一个或多个种子点。相似性准则不仅取决于面对的问题,而且与图像类型有关。对单色图像而言,可以采用基于灰度和空间性质的描述子对区域进行分析。定义连通或相邻信息,是指从种子点逐步生成一个连通区域。连通性的不同定义往往决定符合相似性质准则的像素点是否能归入正在生长的区域。若不再有像素点归入区域,区域生长自动结束。也可在自动生长结束前,定义某种限制性准则,令区域生长结束。一般来说针对同一目标区域,其中不同空间位置的子区域所包含的灰度或纹理信息基本相同,这些子区域之间的灰度和纹理信息的相对变化也比较一致。因此可以选取合适的特征来描述这些区域的差别。

### 2.3 支持向量机与区域增长相结合

本文所提出的支持向量机与区域增长相结合方法框图如图1所示。

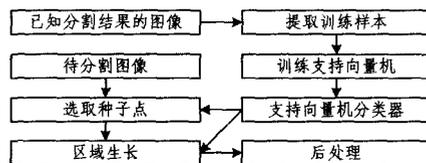


图1 支持向量机与区域增长相结合方法框图

该方法分为4个步骤。

(1)选取样本训练分类器:从已知分割结果的图像中选取一定数量的目标区域与非目标区域样本点作为支持向量机分类器的训练样本并训练支持向量机;

(2)选取种子点:以一定的粒度  $N$ ,即每隔  $N$  行  $N$  列取一个像素点,利用训练好的支持向量机判别该点是否属于目标区域,如果该点属于目标区域,则将该点记为种子点;

(3)区域生长:以这些种子点作起点进行区域增长,使用支持向量机对像素点进行判别,决定是否将该点标记为目标区域;

(4)后处理:针对边缘和噪声像素点进行必要的后处理(平滑、滤波、数学形态学等操作)。

该方法有如下优点。

(1)单纯使用支持向量方法对图像进行分割,需要对图像中的每个点使用支持向量机进行分类判别,而本文提出的方法通常不需要对所有点进行判别,因此在速度上要优于单纯使用支持向量机的分割方法。

(2)与经典的区域增长算法相比,避免了手动选取种子点的交互操作以及增长规则和终止条件的选择困难问题。在该方法中种子点的选取使用支持向量机自动搜索,只需根据具体情况确定合适的粒度  $N$  即可。而在区域增长过程中直接使用支持向量机自动判别。

(3)该方法只需提供一已知分割结果的图像用于训练支持向量机,训练好后就可以实现同类图像的自动分割。

### 2.4 基于 EGEE 的图像并行分割

我们在 EGEE 网络环境下采用主从模型(Master-Slave)实现并行分割算法,其主要执行顺序为:主进程 Master 将图像空间划分为若干任务块并编号,按顺序将带有编号的任务信息分发给每个子进程 Slave;每个子进程收到任务后采用本文提出的分割算法分割子图像数据,处理完后将数据和该任

务的编号一同返回给主进程,然后等待下一个任务的到来;主进程接收到子进程返回的子图像数据后,由任务编号将结果存放于整个图像缓冲区的对应位置,然后查看任务是否全部返回,若是,则保存分割结果的任务完成,否则检查任务池中是否还有任务剩余,若有,则按序取出其中之一,分配给刚返回数据的子进程,若没有,则等待回收尚未完成的任务。图2给出了这一过程的流程图。

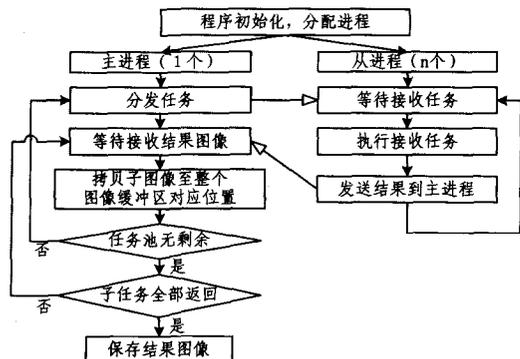


图2 并行分割流程图

在 Scientific Linux CERN 4 操作系统下实现了该并行算法,并在 EGEE 网格环境下分别指定不同计算结点数对其进行了测试,程序运行时间如表1所列。

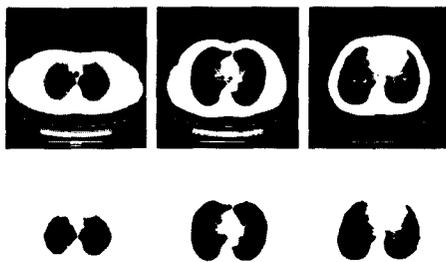


图3 分割效果图(分别为肺的上段、中段和下段)

表1 不同计算结点数时程序运行时间

计算结点数	1	2	4	8	16
运行时间(s)	1786	962	468	248	132

**结束语** 本文提出了一种基于 EGEE 的支持向量机与区域增长相结合的医学图像并行分割方法。该方法从已知分割结果的图像中选取样本训练支持向量机作为区域增长的种子点搜索器和增长规则的判别器,避免了经典区域增长算法中手动选取种子点和门限设置的困难,与单纯使用支持向量机的分割方法相比减少了需要判别的次数,增加了分割的速度,并在 EGEE 环境下实现了该算法的并行化,进一步提高了分割的速度。本文提出的方法在医学图像领域中具有一定的应用前景。

### 参考文献

- [1] Jayaram K U, Supun S. Fuzzy Connectedness and Object Definition: Theory, Algorithms, and Applications in Image Segmentation[J]. Graphical Models and Image Processing, 1996, 58(3): 246-241
- [2] Wan Shu-Yen, Higgs, William. Symmetric region growing[J]. IEEE Transactions on Image Processing, 2003, 12(9): 1007-1015
- [3] Vapnik V. Statistical learning theory [M]. New York: Wiley, 1998
- [2] Helbing D. Traffic and related self-driven many-particle systems [J]. Review Modern Physics, 2001, 73: 1067
- [3] Kerner B S. The Physics of Traffic[M]. Berlin, New York: Springer, 2004
- [4] Lee H K, Barlovic R, Schreckenberg M, et al. Mechanical restriction versus human overreaction triggering congested traffic states[J]. Physical Review Letters, 2004, 92: 238-702
- [5] 董力耘, 薛郁, 戴世强. 基于跟车思想的一维元胞自动机交通流模型[J]. 应用数学和力学, 2002, 23: 331
- [6] 雷丽, 薛郁, 戴世强. 交通流的一维元胞自动机敏感驾驶模型[J]. 物理学报, 2003, 52(9): 2121-2126
- [7] Bazzan A, Klügl F, Ossowski S. Agents in traffic and transportation: Exploring autonomy in logistics, management, simulation, and cooperative driving[J]. Transportation Research C, 2005, 13(4): 251-254
- [8] 姚云, 骆勇. 城市干线交通协调控制仿真优化[J]. 重庆工学院学报: 自然科学版, 2009, 23(3): 23-27

### 3 实验与结果

本实验使用了  $512 \times 512 \times 20$  的已分割出肺实质的胸部 CT 图像(记为图像 A)和  $512 \times 512 \times 64$  的待分割胸部 CT 图像(记为图像 B)。实验中,分别从图像 A 的目标区域(肺实质)和非目标区域(肺实质以外的部分)随机地选取 500 个像素点共 1000 个样本点,从每个样本点的  $5 \times 5$  领域中提取了包括当前像素点灰度值、当前像素邻域均值、当前像素邻域中值以及由当前像素邻域的灰度共生矩阵( $\sigma = 1, \theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$ )导出的 5 个纹理特征(角二阶矩、对比度、相关性、方差、反差矩,取 4 个方向上的平均值)共 8 个特征,经归一化后形成 SVM 的输入向量  $x$ ; 训练支持向量机。在搜索种子点过程中粒度  $N$  取 32,即每隔 32 行 32 列取一点判别是否为种子点。应用本文的方法对图像 B 进行分割的结果如图 3 所示。

采用 C++ 语言结合 ITK (Insight Segmentation and Registration Toolkit) 开发包和 MPICH2.0 并行程序开发包

(上接第 233 页)

双面状态加减速能力和在单车道上的多代理相互作用。这显示该模型能模拟大部分实际观测结果,包括基本相位、速度-密度相位和 3 种熟知的交通相位。对速度-位置相位、1-min 压缩流动-时间相位和速度-时间相位和同步移动状态都做了总结。自由状态和同步状态的交叉关联功能分析与其它研究表现出良好的一致性,意味着同步移动的变化可以归因于自主代理在不同群体中的不同速度。

在未来的工作中,将集中研究相互作用的系数因  $\eta$ , 这会允许考虑超过两个前面代理的相互作用。而且,将进一步调查一个速度群的速度对同步移动状态的移动的影响。

### 参考文献

- [1] Chowdhury D, Santen L, Schadschneider A. Statistical physics of vehicular traffic and some related systems[J]. Physics Reports, 2000, 329: 199