

基于形态结构特征的对流单体自动分割方法研究

张军,贺婷婷,侯谨毅[†],王萍

(天津大学 电气自动化与信息工程学院,天津 300072)

摘要:风暴单体是形成各类强对流灾害天气的基本单元,它们的雷达回波形状复杂、内部分布不一、外层相互交织,从而造成单体分割困难.提出了一种基于形态结构特征的对流单体自动迭代分割方法.以雷达图片中区域树结构单体分割结果为初始输入,在每次迭代分割过程中,首先计算各个分割结果的 3 个形态结构特征,然后通过一个预先训练的支持向量机(Support Vector Machine, SVM)分类器判断分割结果是否为对流单体,对非单体的分割结果进行再次分割.通过 3 种不同类型的风暴案例进行测试,结果表明,本文方法能够有效地识别出聚集的单体和处于分裂/合并状态的单体,并且能够获得单体的完整结构.在定量评估测试中,本文算法获得了 0.84 的临界成功指数评分,高于传统的风暴单体识别与跟踪算法(Storm Cell Identification and Tracking, SCIT)方法(0.55)和单阈值方法(0.49).

关键词:天气雷达;风暴单体;形态学特征;迭代分割;计算机视觉

中图分类号:TP391.4

文献标志码:A

Research on Automatic Segmentation Method of Convective Cell Based on Morphological Structure Characteristics

ZHANG Jun, HE Tingting, HOU Jinyi[†], WANG Ping

(School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China)

Abstract: Storm cells are the basic units that form various types of severe convective weather. Their radar echoes have complex shapes, uneven internal distribution, and intertwined outer layers, which makes cell segmentation difficult. This paper proposes an automatic iterative segmentation method for convective cells based on their morphological structure characteristics. Taking the cell segmentation results based on the region-tree structure on the radar image as the initial input, in each iterative segmentation process, the three morphological structure features of each segmentation result are first calculated, and then a pre-trained SVM classifier is used to determine whether the segmentation result is a convective cell. The segmentation results that are not cells are segmented again. The method in this paper was tested through three storm cases with different types. The results show that the method can effectively identify aggregated cells and cells in a split/merged state, and can obtain the complete structure of cells. In the quantitative evaluation test, the algorithm presented in this paper obtained a Critical Success Index score of 0.84, which is higher than that of the traditional SCIT method (0.55) and the single threshold method (0.49).

Key words: weather radar; storm cell; morphological characteristics; iterative segmentation; computer vision

* 收稿日期:2020-09-02

基金项目:天津市面上基金(14JCYBJC21800), Tianjin Market Fund(14JCYBJC21800)

作者简介:张军(1964—),男,天津人,天津大学自动化学院副教授

[†] 通信联系人, E-mail:houjinyi@tju.edu.cn

由于气候变化,近年来对流风暴的频率和强度都显著增加^[1-3].极端的对流风暴事件会造成严重的社会经济损失.由于对流风暴高度动态的空间和时间过程,对流事件的研究仍然是一个具有挑战性的问题.天气雷达是监测强对流天气(冰雹、大风、龙卷和暴雨)的主要工具之一,利用天气雷达,可以更详细地分析对流风暴的形成和运动过程.

在天气雷达的反射率强度图像上,对流风暴经常表现为多单体共存的结构.其中,有些单体交织在一起.虽然强对流灾害可能由这些相互交织的单体群共同引发,但不同的对流单体所引发的灾害类型、强度各有不同.因此,强对流天气的自动识别及预警的前提是正确分割出每个对流单体.

自20世纪50年代以来,已经开发了许多基于天气雷达的对流单体自动识别算法.由于对流单体的强度较高,在雷达反射率图片上表现为局部极大值区域.最简单的识别局部极大值区域的方法是基于阈值的方法,将雷达图片上大于某个阈值的一片联通区域识别为对流单体.目前两种主流的对流单体识别方法,雷暴识别、跟踪、分析和临近预报方法^[4](Thunderstorm Identification, Tracking, Analysis and Nowcasting, TITAN)和 SCIT^[5]方法都是基于阈值的方法.其中 TITAN 是一种固定阈值方法,而 SCIT 是一种动态阈值方法.

一般来讲,动态阈值方法是一种更加有效的方法,能够有效的减少对流单体的分割错误.其他的一些基于动态阈值方法包括:侯正俊等^[6]根据不同下垫面选择不同阈值增强了 TITAN 算法的地区适应能力;杨吉等^[7]以 SCIT 为基础设计了一种基于三维雷达拼图的单体分割算法;Crane^[8]通过统计降水区域中的最大阈值获得一个动态阈值对单体进行分割.

基于阈值的方法并不是唯一的能够识别局部极大值区域的方法. Lakshmanan 等先后提出了基于 K-均值聚类^[9]和分水岭算法^[10]的单体识别算法,同样通过检测局部极大值区域来检测对流单体,但是,这种方法容易受到噪音的干扰,容易产生过分割,需要对分割结果进一步合并处理.另外, Wang 等^[11]提出了基于种子点生长和膨胀避让的单体识别算法,该方法在阈值分割的基础上,利用数学形态学方法获得对流单体的更加完整的结构. Hou 等^[12]使用区域树结构描述方法,将天气雷达图片描述为一个区域树形结构,然后通过剪枝法获得图片中的局部极大值区域,并以此作为对流单体的识别结果.

以上传统的方法通过检测天气雷达反射率图片

中的局部极大值区域来识别对流单体.在实际应用中,简单的通过检测局部极大值区域来识别对流单体存在一些问题:首先,局部极大值区域不一定是对流单体,还可能是流场中的噪音和扰动;其次,传统的对流单体识别算法只考虑对流单体的强度信息,而没有考虑对流单体的形态结构信息,会得到错误或不完整等单体识别结果;最后,传统的单体识别算法都是一个开环的结构,无反馈过程,换句话说,以往的单体识别算法不会检验分割结果是否正确,以及是否需要进一步对识别结果进行拆分等处理.

考虑到以上问题,本文设计了一种基于形态结构特征的对流单体动态识别算法.该算法在识别过程中加入了反馈环节,将单体识别过程设计为一个迭代识别过程,在每一步的迭代中,一个分类器通过分割结果的形态结构特征判断每个识别结果是否是对流单体.若分割结果不是对流单体,那么进行进一步拆分.设计算法提取每个分割结果的形态结构特征也是本文工作的一个关键点.

1 数据和算法

1.1 数据和预处理

本文研究案例的雷达数据是由位于中国天津市塘沽区的 S 波段多普勒雷达收集的.雷达站点位于中国北京东南 100 km 处的(117°43'E, 39°00'N).雷达体扫模式为 VCP 21,即每 5 min 进行一次体积扫描.为了训练和测试本文算法,我们一共收集了 10 个案例.这 10 个案例中包含不同模式的对流风暴,每个案例的持续时间约为 2 h.具体的案例相关信息见表 1.其中,6 个案例作为训练集,而剩余的 4 个案例作为测试集.

本方法设计用于处理 0.5°仰角的雷达反射率图像,原始的雷达数据被转换到笛卡尔坐标系下,其中,笛卡尔网格的水平空间分辨率为 0.01° 0.01°(1 km 1 km).为了滤除雷达图片上的噪音和地面杂波,对雷达图片数据进行滤波操作,因此,利用中值滤波消除图片中的噪音点.

1.2 对流单体识别方法

1.2.1 算法概述

图 1 为本文设计方法的流程图.整体流程包括 3 个主要部分:初分割、迭代分割和判别器训练.在初分割部分,首先利用区域树方法检测风暴图像中的极大值区域,然后以极大值区域为核心,采用分水岭方法分割风暴图像;在迭代分割部分,首先提取分割

结果的形态结构特征, 然后通过分类器判断分割结果是否为对流单体, 若不是对流单体, 则进行进一步拆分; 迭代分割过程中用到的分类器通过人工标记的单体样本图片训练得到.

表 1 对流风暴案例列表

Tab.1 List of convective storm cases

编号	起始时间	终止时间	风暴类型	体扫数目
1	2018-05-01	2018/05/01	线状单体	21
	19:21:30	21:12:43		
2	2018/05/01	2018/05/01	线状单体	22
	21:48:53	23:31:30		
3	2018/05/14	2018/05/02	团状单体	78
	20:00:06	03:01:06		
4	2018/05/14	2018/05/14	团状单体	22
	19:33:18	22:54:54		
5	2018/05/14	2018/05/14	孤立单体	26
	20:56:42	22:54:54		
6	2018/05/14	2018/05/14	团状单体	23
	03:56:27	05:50:34		
7	2018/05/14	2018/05/14	线状单体	18
	01:56:17	03:51:13		
8	2018/05/15	2018/05/15	线状单体	19
	01:49:33	03:51:33		
9	2018/05/29	2018/05/30	孤立单体	25
	22:48:40	00:43:33		
10	2018/05/30	2018/05/30	孤立单体	17
	23:17:16	00:45:34		

1.2.2 对流单体的初分割

对流单体初分割方法是建立在采用区域树结构描述对流风暴图片的基础上. 区域树结构描述对流风暴雷达图片的方法见文献[12].

图 2 给出了构造雷达反射率图片区域树结构的示意图. 图 2(a)为一个对流风暴的示意图. 图 2(b)中的 P_1 到 P_5 为不同阈值下的分割图片, 其中的联通区域用标号标出. 图 2(c)为描述对流风暴的抽象数据结构, 其中每个节点对应一个区域, 并且存储了对应区域的相关信息. 节点间的连线代表了区域的包含关系.

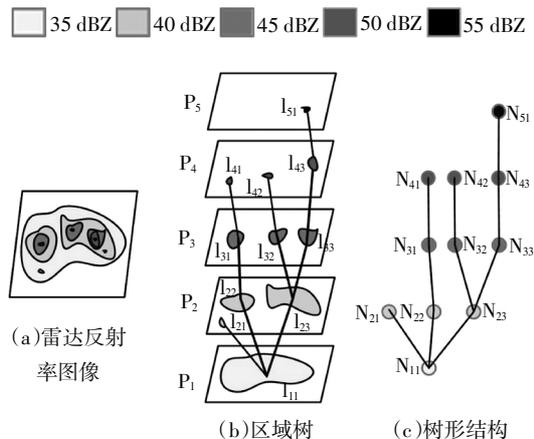


图 2 构建区域树结构的示意图

Fig.2 Schematic diagram of building the region tree structure

当利用树结构描述对流风暴的时候, 需要引入一些树结构相关的定义. 树中节点间的连线称为边, 边上强度值较低的节点称为父节点, 而强度值较高的节点称为子节点. 没有子节点的节点称为叶节点, 存在多个子节点的节点为分支节点, 没有父节点的节点为树结构中的根节点.

当对流风暴雷达反射率图像被描述为区域树结构之后, 可以借助区域树结构实现对风暴图片滤波和检测其中的局部极大值区域. 在区域树结构中, 将面积小于 5 的所有区域剔除, 然后将剩下的区域叠加在一起, 就实现了对风暴结构的滤波. 如图 3(a)所示, 中间为去除小块区域之后的树结构, 右侧为重新叠加区域得到的滤波结果.

在利用区域树结构寻找局部极大值区域的时候, 首先遍历树中的每一个叶节点, 沿着其父节点向根部探索, 当遇到一个分支节点时, 探索过程结束, 探索路径上的节点集合对应着一个局部极大值区域. 将探索路径上的节点对应的区域叠加在一起, 得到局部极大值图像. 图 3(b)给出了一个采用这种方式获得局部极大值区域集合. 其中间图上的矩形框

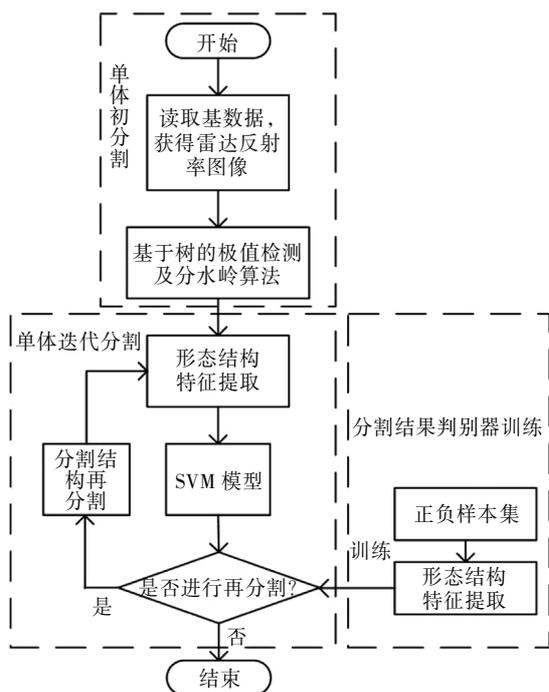


图 1 算法流程图

Fig.1 Flow chart of the algorithm

标记出所有探索得到的路径,将这些路径上的区域重叠在一起,得到了此案例中的局部极大值区域,见图3(b)右侧。

算法获得的局部极大值区域与 SCIT 方法获得的对流单体检测结果是相似的,其缺陷很明显,当对流风暴内部有多个细小的反射率核的时候,对流单体的检测结构不完整.为了获得对流单体的完整结构,我们使用了基于距离变换的分水岭分割算法^[16-18],利用距离变换方法计算出风暴区域内部所有点到局部极大值区域的最短欧式距离.如果一个点在局部极大值内部,那么距离变换结果为0.风暴区域的距离变换示例如图3(c)中部所示.对距离变换的结果采用分水岭方法分割,得到对流单体的初分割结果,见图3(c)右侧.可见,初分割方法能够获得对流单体的完整结构。

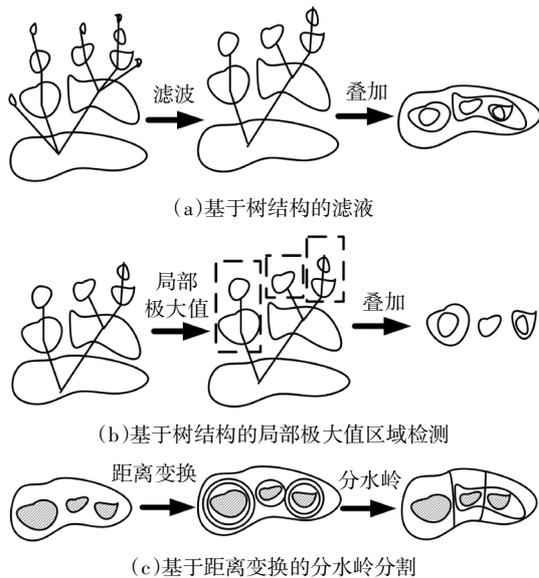


图3 基于区域树结构的初分割方法示意图

Fig.3 Schematic diagram of the initial segmentation based on region tree structure

1.2.3 对流单体的迭代分割

对流单体的初分割过程只考虑了对流单体的强度信息,但是当对流单体发生分裂与合并的时候,容易造成错误的分割结果.因此需要对初分割的结果进行进一步处理,检查每个分割结果的形态结构特征以判断其是否为一个对流单体.当对流单体的形态结构表现为多个单体相互联合的情况时,需要对其进一步拆分.因此,本文设计了一个迭代的再分割过程。

对流风暴的初分割结果送入迭代分割过程中,在迭代过程中的每一步,首先提取分割区域的形态

结构特征,然后利用一个判别器决定该区域是否需要进一步进行分割,若该分割区域满足对流单体的判据,则输出该分割结果;否则,对分割结果进一步拆分,不断重复以上迭代过程.在迭代过程中,重点是如何提取分割结果的几何形态结构,以及如何对分割区域进行再分割。

在迭代分割过程中,计算每一个分割区域的以下3个特征:面积衰减率、质心偏移率以及凹陷度.分割区域的特征计算需要借助区域树结构描述方法.如图4所示,初分割结果相当于将原始的区域树结构拆分为一组由单一路径构成树集合,每一个子树对应一个初分割区域.对于每一个子树,由于其只有一个单一的区域链,因此可以记录为一个区域序列: $T = \{R_1, R_2, \dots, R_n\}$. 每个区域的面积和质心点同样构成序列 $\{S_1, S_2, \dots, S_n\}, \{x_1, x_2, \dots, x_n\}$. 下面给出三种特征的计算方法:

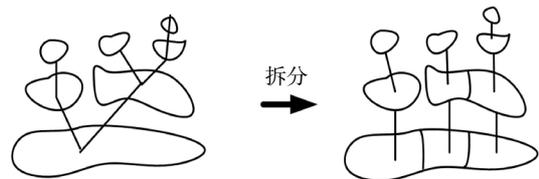


图4 区域树结构的初分割效果

Fig.4 Initial segmentation effect of region tree structure

1) 面积衰减率. 首先计算相邻两个区域的面积比值:

$$t_i = S_i / S_{i-1} \quad (1)$$

这里 $i = 2 : n$. 然后计算集合 $\{t_i\}, i = 2 : n$ 的方差 d_t , 将 d_t 作为面积衰减率的一个度量. 面积衰减率度量了分割区域内部不同强度的面积的波动情况. 对于一个对流单体, 其内部各强度区域的面积变化较为平缓, 而当分割区域内部包含多个单体时, 其内部各强度区域的面积会存在一个跳变, 导致 d_t 较大, 更趋向于被判定为需要进行再分割。

2) 质心偏移率. 与面积衰减率计算公式相似, 首先计算相邻两个区域的质心偏移程度:

$$c_i = d(x_{i-1}, x_i) / \sqrt{S_{i-1}} \quad (2)$$

式中: $d(x_{i-1}, x_i)$ 为区域 R_{i-1} 和 R_i 的质心欧式距离; $\sqrt{S_{i-1}}$ 用于对距离进行归一化. 然后计算 $\{c_i\}, i = 2 : n$ 的方差 d_c , 将其作为质心偏移率的一个度量. 通过分析, 质心偏移率越大, 相邻两反射率阈值下的连通域质心点距离越远, 分割区域内部更趋向于需要进行再分割。

3) 凹陷度. 具有分裂趋势的单体通常会有凹陷

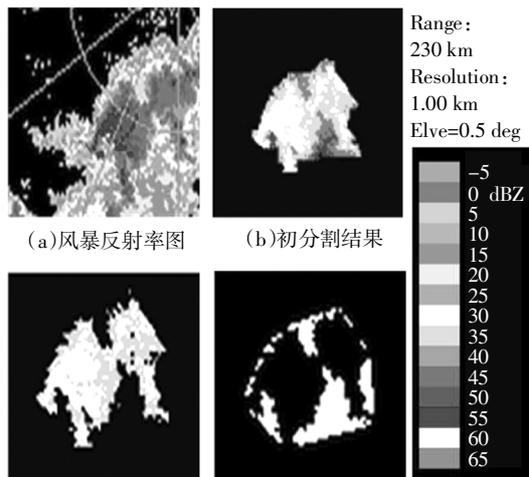
的较高反射率的区域, 而一个完整的对流单体则不会有凹陷较高反射率的区域, 如图 5(a)所示. 因此可以求取分割区域的凹陷度作为识别对流单体的一个结构特征.

图 5 给出了计算分割区域凹陷度的方法, 首先计算每个区域 $\{R_1, R_2, \dots, R_n\}$ 的凸包多边形的面积 $S_{a1}, S_{a2}, \dots, S_{an}$, 再计算每个区域面积与区域凸包多边形的面积比值 e_i , 然后计算不同阈值下最大的凹陷面积比 e_m , 作为凹陷度的度量.

$$e_i = S_i/S_{ai} \tag{3}$$

$$e_m = \max\{e_1, e_2, \dots, e_n\} \tag{4}$$

同时记录最大凹陷面积比所属的强度 I_m . 凹陷度值越大, 说明分割区域的凹陷程度越厉害, 则越倾向于被判定为需要进一步分割. 在图 5 中, 图 5(b)是分割结果, 其中具有 45 dBZ 的反射率区域 (见图 5(c))具有最大的区域凹陷面积比, 该区域的凸包多边形如图 5(d)所示.



(c)最大凹陷面积比区域 (d)图(c)的凸包

图 5 凹陷度计算示意图

Fig.5 Schematic diagram of concavity calculation

对于每一个分割结果, 计算其 3 个形态结构特征, 得到一个三维特征向量, 在迭代过程中, 将该向量输入到二分类 SVM 中, 对于被判定为需要进一步分割的多单体风暴, 采用数学形态学开运算对其再分割. 开运算是指对二值图片先腐蚀后膨胀, 开运算操作可以平滑对象的轮廓, 消除细小的尖刺, 断开窄小的连接. 详细的再分割流程图见图 6, 该流程的说明如下:

①利用最大凹陷面积比强度 I_m 对原始单体区域进行二值化, 得到区域核.

②对强度 I_m 下的区域核进行一次开运算处理,

对开运算处理过的区域核进行中值滤波, 滤除断开的细小的毛刺区域.

③识别所有开运算及滤波之后的联通区域, 并且根据面积进行排序. 找出面积最大的联通区域的面积, 计算其他区域与最大面积的比值.

④若所有的面积比值都小于 0.5, 说明没有完成区域拆分, 转步骤②. 否则, 说明开运算将该区域拆分为两个及两个以上的子区域.

⑤将面积最大的两个连通域作为符合要求的核, 基于原本单体的掩膜进行分水岭分割操作, 得到两个分割区域.

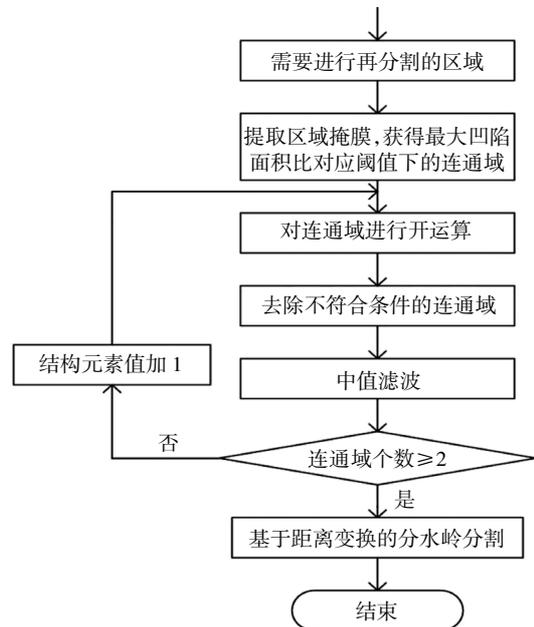


图 6 迭代分割方法流程图

Fig.6 Flow chart of the iterative segmentation method

图 7 展示了一个案例来说明如何使用数学形态学方法对单体区域进行再分割. 首先, 在图 7(a)中的所有反射率区域中, 利用 I_m 提取图 7(a)中的连通区域, 得到如图 7(b)所示的图像. 对图 7(b)图像进行多次的开运算和中值滤波处理, 直到该区域被拆分为两个核区, 如图 7(c)所示, 然后利用图 7(c)中的多个核区对原始的分割区域进行距离变换和分水岭分割操作. 得到两个再分割的区域, 如图 7(d)所示. 此时, 再分割过程结束.

1.2.4 分割结果判别器训练

在迭代分割过程中, 分割结果的判别和进一步分割不断交替进行, 判别过程利用了一个二分类的 SVM 分类器判断是否需要进一步分割. SVM 分类器可以解决小样本下机器学习的问题, 提高泛化性能^[9], 考虑到训练集样本的数据量及三

特征的空间分布,本文选择 SVM 分类器判断分割区域是否为一个对流单体.

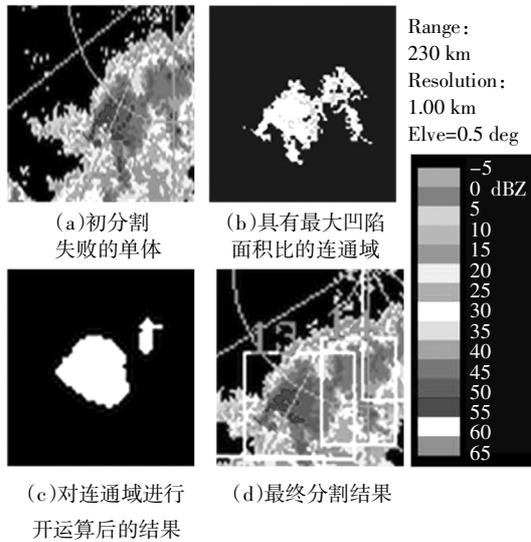


图7 区域再分割示例
Fig.7 An example of region re-segmentation

为了获得 SVM 模型,需要通过人工标记样本,样本包括需要再进一步拆分的分割结果和不需要进行拆分的对流单体,人工获取样本的过程相当于在图1所示的流程图中用人工取代 SVM 模型.在每一次分割的迭代过程中,通过人工判断分割的结果是否为对流单体,如果是对流单体,那么标记为正样本,否则,将其标记为负样本.该标记过程不断重复,直到所有的单体都被识别出来.

对所有人工标记的正负样本,计算其3个几何特征(如前所述).SVM 的训练过程其实是为了找到正负样本在高维特征空间的一个超平面,正样本和负样本分别分布在超平面的两侧.

2 实验与结果

本文算法与两种传统的对流单体识别算法进行对比来验证算法的有效性,这两种对流单体识别算法分别是 SCIT 方法和单阈值算法. SCIT 方法是一个基于动态阈值的极大值区域检测算法,而单阈值方法是一个基于固定阈值的方法.评估系统平台为个人 PC 机,处理器为英特尔酷睿 5,系统内存为 8 Gb,评估实验软件及评估算法通过 C++语言编写.

评估过程包括定性评估和定量评估,在定性评估中,首先分析了本文设计的对流单体几何特征对于分割对流单体的有效性,然后通过几个典型的案例展示,在对流风暴演化的过程中(如生长、消失、分

裂、合并等过程),本文算法如何分割不同风暴类型(如线性风暴、团状风暴和孤立单体风暴)中的对流单体,以及本文方法与传统方法的区别.在定量评估过程阶段,通过人工评估对流单体的识别结果,给出不同单体分割方法的主观定量评估结果.

2.1 定性评估

1)特征有效性.图8(a)(b)(c)分别是3个特征的可分性对比图,左侧分布图表示无需再分割的单体特征,右侧分布图表示需要再分割的分割结果的特征.

可以看出:无需再分割的单体,其面积衰减率、质心偏心率和凹陷度集中分布在低值区域,而内含多单体区域的这三个特征较分散的分布于高值区域.因此可以说三特征描述的单核单体和多核“单体”分别来自两个均值、方差均具显著性差异的总体.简单地说:三特征各自的分类能力是显著的;假设令每个特征独立承担辩证上有待再做分割处理的任务,并取图8中低值类上限值做阈值,面积衰减率的漏分辨率仅为9.09%,质心偏心率和凹陷度的漏分辨率也分别低于9.1%和21.05%.联合建立三特征空间样本分布如图8(d)所示,显然,这两类样本在三维空间的分布状态为支持向量机将其变换到高维空间后获得优质分类能力奠定了重要基础,详见后面的分析.

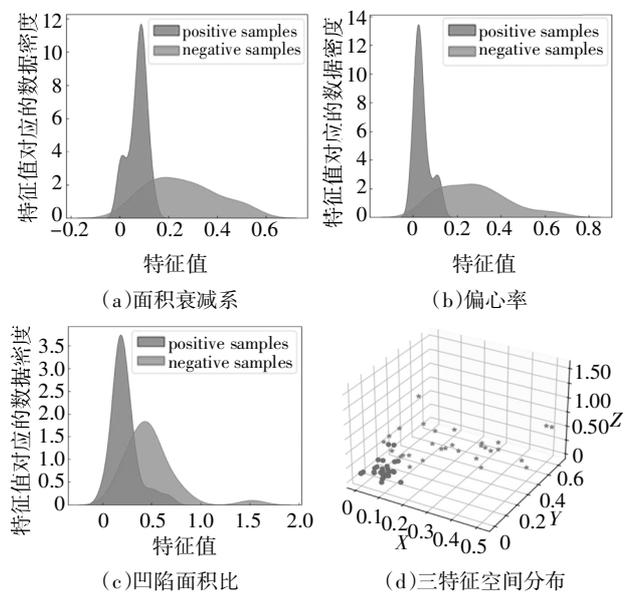


图8 正负样本的形态特征分布
Fig.8 Distributions of morphological characteristics of positive and negative samples

2)案例分析.在本节中,采用一个线性风暴展示本文算法分割对流单体的各步中间结果,然后使用不同类型的风暴说明本文方法与传统方法的区别.

线性对流风暴. 图 9 展示了一个线性风暴在天气雷达 0.5°仰角上的观测结果. 雷达反射率图片的大小为 512 × 512, 图片中心为雷达所在的位置, 3 个距离圈分别标记出离雷达中心 50 km、150 km、230 km 的范围, 各条斜线的角度分别为 30°、60°、…、360°. 图 9 中虚线框中的部分存在多个聚集的单体, 而且其上部的单体正在发生分裂, 我们使用线性风暴的这一部分展示本文算法是如何工作的.

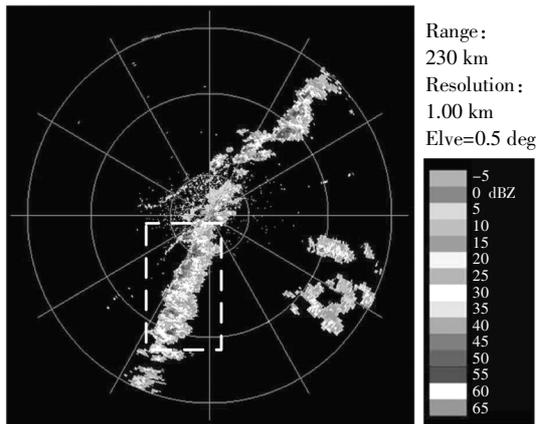


图 9 线状风暴案例
Fig.9 Linear storm case

图 10 展示了初分割方法各个步骤的输出结果, 以及最终的初分割结果. 天气雷达的反射率图片首先需要从极坐标系转换到笛卡尔坐标系, 如图 10(a) 所示, 然后对雷达图片进行中值滤波, 去除噪点, 得到图 10(b). 利用图 10(b) 的区域树结构, 可以得到图中的局部极大值区域, 如图 10(c) 所示, 图中阴影部分为局部极大值区域. 同时, 图 10(c) 中还展示了 30 dBZ 的风暴区域 (图中白色部分). 在风暴区域内部以局部极大值区域为核进行分水岭分割, 可以得到对流风暴中单体的初分割结果, 如图 10(d) 和图 10(e) 所示.

从图 10(e) 展示的对流单体初步分割的结果可以看出, 正在发生分裂或者合并的相邻单体, 由于这些单体共用一个内核, 难以通过局部极大值区域检测分开. 为了能够区分并分割这些单体, 需要结合对流单体的形态结构信息, 对初分割结果进一步拆分.

在迭代分割过程中, 第一步的输入为图 10(e) 中的所有初分割区域结果. 首先计算图 10(e) 中每一个分割结果的向量形态结构特征, 将特征向量输入到 SVM 分类器中, 根据 SVM 分类器的判断, 图 10(e) 中标号为 14 号的分割结果不符合对流单体的判据, 而其余的 3 个分割结果都满足对流单体的判定条件, 因此需要对 14 号分割结果进一步拆分.

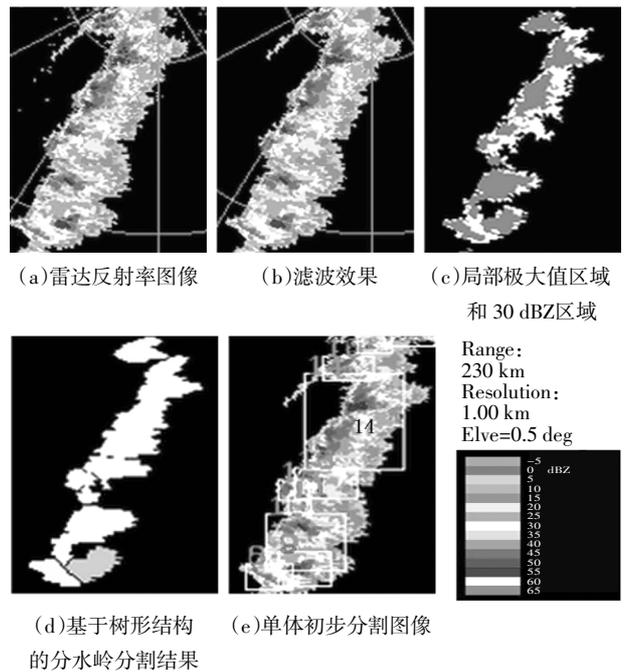
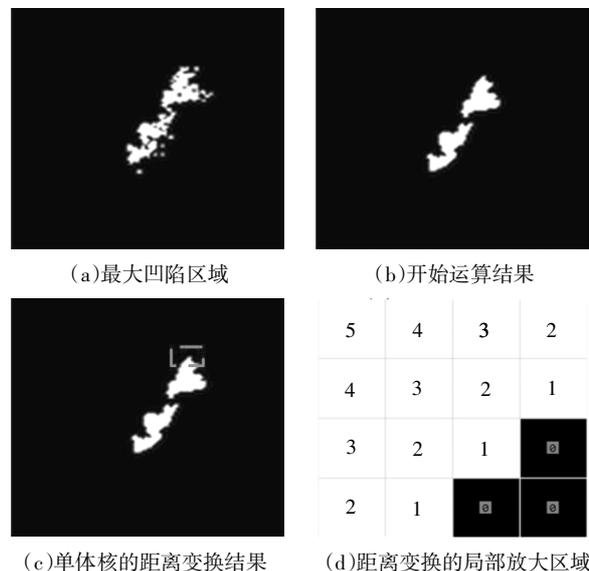


图 10 初分割方法的输出结果
Fig.10 Results of the initial segmentation procedures

图 11 展示了对图 10(e) 中 14 号分割结果进一步拆分的各个中间步骤. 在 14 号分割结果的特征提取过程中, 记录了具有最大面积凹陷比率的区域强度 I_m 以反射率强度为 I_m 的区域作为再分割目标区域, 如图 11(a) 所示. 对图 11(a) 中的区域不断进行开运算和滤波处理, 直到将该区域拆分为多个区域并且满足拆分的判据, 拆分结果见图 11(b). 在图 11(a) 所示的区域中, 以拆分结果为核, 进行距离变换, 得到图 11(c), 对距离变换的结果图进行分水岭分割 (图 11(d) 是距离变换的局部放大示意图) 可以得到图 11(e) 的再分割结果.



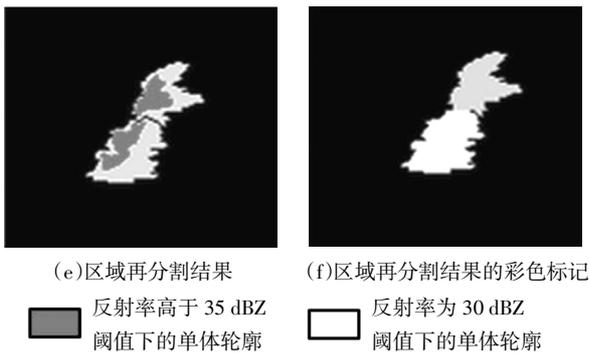


图 11 再分割各步骤结果
Fig.11 Results of the re-segmentation procedures

接下来，对再分割得到的单体进行特征提取、SVM 分类，判断是否需要进一步迭代分割。在此案例中，图 11(e)中的两个分割结果的集合结构特征全部满足对流单体的判据，因此无需进行再次分割，对流单体识别结束。最终利用本文方法得到的分割结果如图 12(a)所示。

图 12 中同时列出了利用单阈值分割算法及 SCIT 算法的单体分割结果。为了详细对比 3 种算法的分割结果，不同算法的单体分割结果采用两种方式标记出来，图 12(a)(c)(e)采用彩色掩膜的方式标记出本文算法、单阈值方法及 SCIT 方法的单体分割结果；而图 12(b)(d)(f)再采用矩形框的形式标记出 3 种算法的单体分割结果。

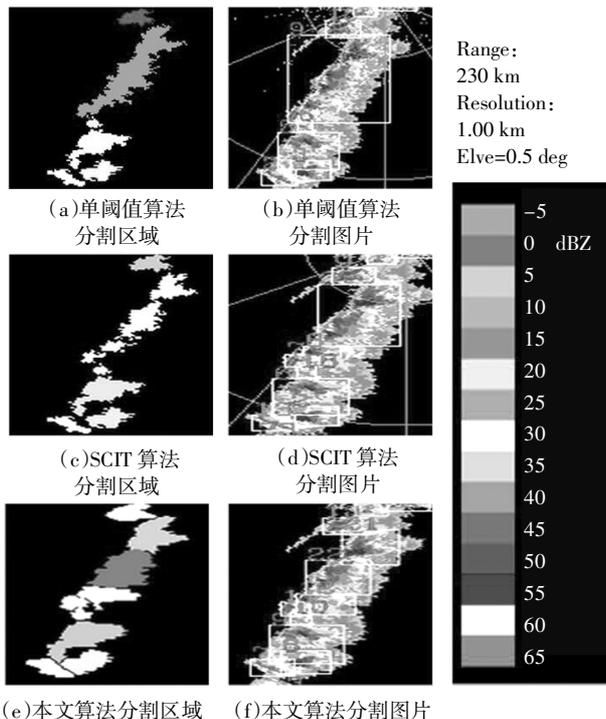


图 12 线性风暴分割结果
Fig.12 Segmentation results of line-typed storms

由图 12 可见，单阈值方法将风暴上部的 5 个相互黏连的、具有分裂趋势的对流单体识别为了一个结构，单体分割错误非常明显。相比之下，SCIT 方法能够将线性风暴中相互临近的对流单体分割出来，但是 SCIT 方法将一对正在分裂的对流单体识别为了一个单体。另外，SCIT 方法识别出的单体结构明显不够完整，这些识别结果中不包含 30 dBZ 及 35 dBZ 的反射率区域。在此案例中，本文算法获得了最优的单体分割结果，能够将所有对流单体正确分割，而且相比于 SCIT 方法，本文算法能够获得更加完整的对流单体结构。

团状风暴与孤立单体风暴的分割过程与线状单体相同，采用的雷达基数据格式相同，且雷达反射率图像与线状单体所示的分辨率相同，因此，在本小节中，案例分析过程参照线状单体，接下来，展示本文方法与单阈值方法及 SCIT 算法在团状风暴及孤立单体风暴情况下的单体分割结果。

图 13 和图 14 分别为 3 种单体分割算法对于团状风暴和孤立单体风暴中的单体的识别效果。图 13 和图 14 采用了与图 12 相同的展示方法，使用色标和矩形框标记对流单体的分割结果。

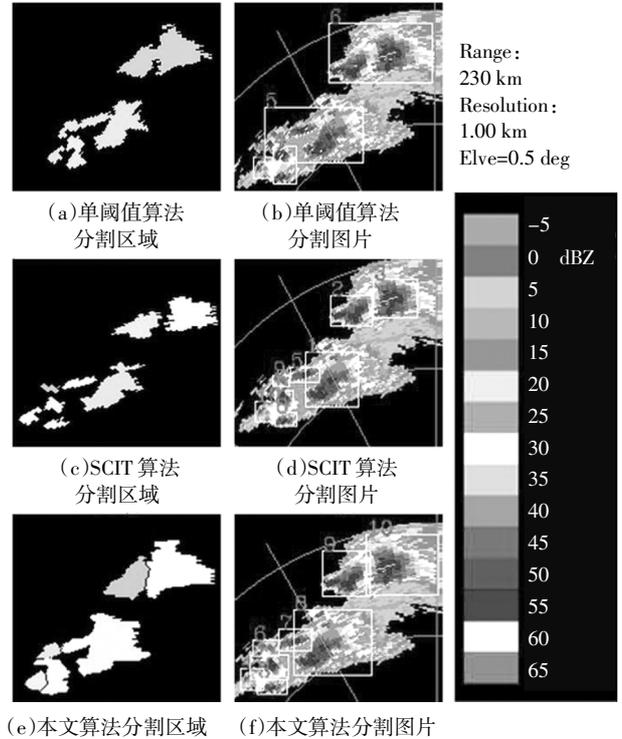


图 13 团状风暴分割结果
Fig.13 Segmentation results of clustered storms

从图 13 中可以看出，在团状风暴的案例中，多个单体相互临近，但是没有单体分裂/合并的现象出

现.在这种情况下,单阈值分割方法同样将一些相邻的对流单体识别为一个单体. SCIT 方法与本文算法获得了相似的识别结果,相对比本文方法,SCIT 方法的缺陷主要在于单体结构不够完整,忽视了单体结构的完整性,丢失了部分信息.

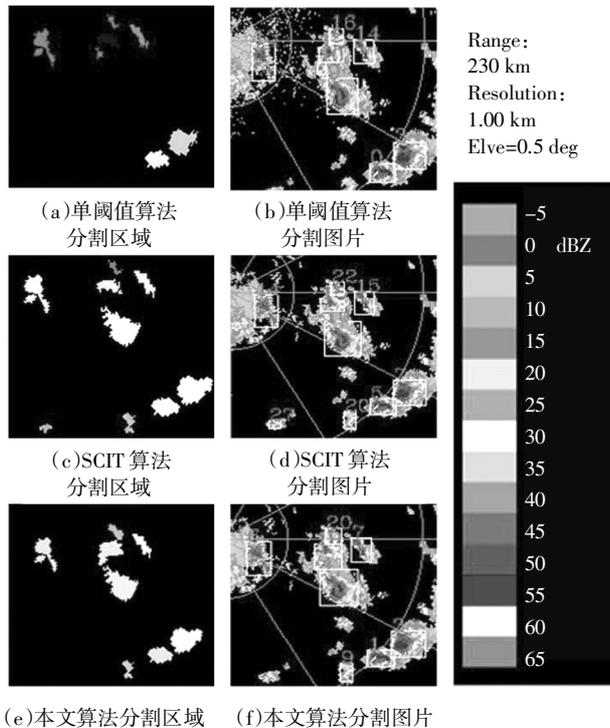


图 14 孤立单体风暴分割结果

Fig.14 Segmentation results of single-cell storms

从图 14 中可以看出,对于孤立单体风暴的案例,由于各个孤立单体风暴之间的距离较远,3 种单体分割算法获得了几乎相同的结果,而且各种方法都能够获得对流单体的完整结构.

2.2 定量评估

1) 对流单体分割临界成功指数. 为了定量评估各类对流单体分割算法的性能,需要人工检验各类分割算法的识别结果. 首先人工标记出每个算法正确分割的单体,错误分割的单体,以及漏识的单体;然后统计每种单体出现的数目,根据统计结果计算出单体分割算法的量化评分值.

在人工检验过程中,算法正确分割的对流单体为具有完整结构的,存活时间超过两个连续体扫的局部极值区域. 一个对流单体的空间结构在演化过程中应该保持相对稳定. 算法正确分割的单体个数记为 X . 错误的对流单体分割结果是指算法将由于扰动或者噪音造成的局部极大值区域识别为对流单

体. 所有错误的对流单体分割结果个数记为 Y . 漏识的单体是指由于面积较小,或强度过低等原因造成的算法不能够分割的对流单体. 所有漏识的对流单体个数记为 Z . 另外,当对流单体发生分裂或合并时,如果 n 相邻的对流单体被识别为一个对流单体,那么该对流单体识别结果被认为是错误的分割结果,同时认为算法漏识了 n 个对流单体. 由于分裂或合并造成对流单体分割错误则对分割的对流单体的数目进行单独的统计,其数目分别记为 Y' 和 Z' .

当通过人工标记的方式对算法的分割结果进行检验和统计之后,计算以下三项评分值:对流单体的识别率:

$$POD = X / (X + Z + Z') \quad (5)$$

误报率:

$$FAR = (Y + Y') / (X + Y + Y') \quad (6)$$

和临界成功指数:

$$CSI = X / (X + Y + Y' + Z + Z') \quad (7)$$

在定量评估过程中,以 CSI 作为主要的算法量化评价指标,该值的范围从 0~1,CSI 评分值越高,算法性能越好.

2) 形态学特征有效性验证. 本文提取了面积衰减率、偏心率及凹陷面积比 3 个形态学特征以判断单体是否需要迭代分割. 为了验证形态学特征的有效性,本文设计了关于特征有效性的测试,利用 CSI 作为算法的评价指标,在所有流程不变的基础上,使用最为基本的、在气象领域最为常见的特征:如面积、最大强度、最小强度等. 然后依次增加不同的形态特征,判断选择不同的特征对于算法有何种影响.

在 2.1 节特征有效性分析中,分析了 3 种形态特征对于正负样本具有较好的可分性,且计算出面积衰减率、偏心率及凹陷面积比的漏分割率分别为 9.09%、9.1% 和 21.05%. 图 15 为正负样本中气象常用特征(面积、最大强度、最小强度)的分布情况. 只有面积特征可以用于区分单体正负样本,但是面积特征的漏分辨率达到 43.40%. 通过对比可知,本文提出的 3 种形态特征更加适合于单体分割判别.

此外,为了进一步验证本文形态特征的有效性,我们首先使用气象基本特征:面积、最大强度、最小强度对单体进行分割,得到 CSI 评分值,然后,在使用基本特征训练分类器的基础上依次加入面积衰减

率、偏心率及凹陷面积比三个形态特征,并且计算对应的 CSI 评分值,结果如表 2 所示.由表 2 可知,当在训练分类器的过程中加入形态结构特征的时候,分类器的 CSI 评分值会得到改善,当加入 3 种形态特征时,算法的 CSI 评分值达到了最大(0.82).该实验结果显示,本文选取的形态特征有助于得到更好的单体分割效果.

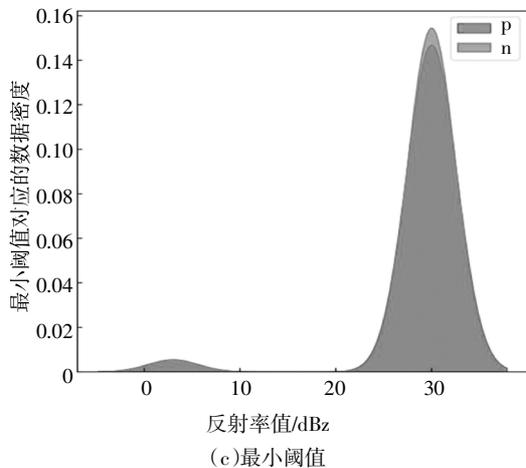
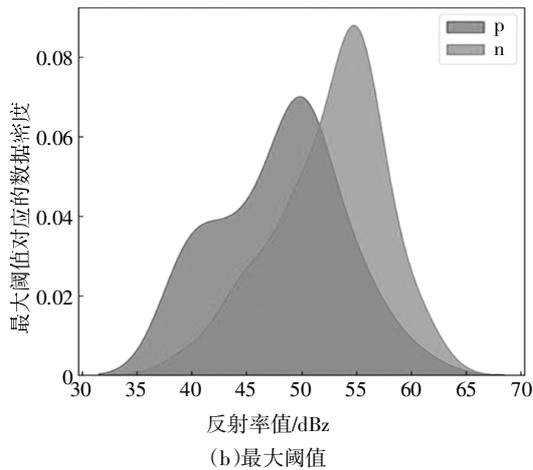
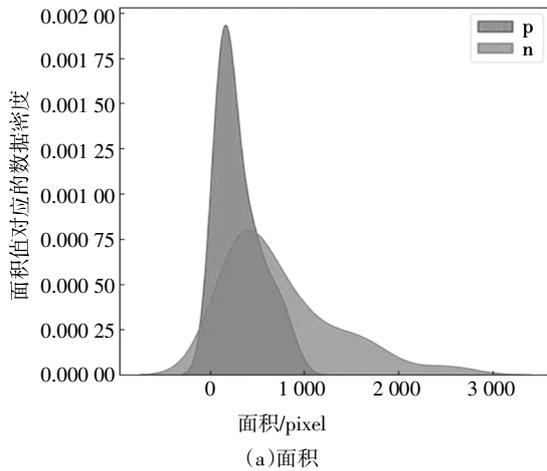


图 15 正负样本的常用特征分布

Fig.15 The distribution of common characteristics of positive and negative samples

表 2 特征有效性分析验证结果

Tab.2 Validation results of feature validity analysis

特征	评价指标					CSI
	X	Y	Y'	Z	Z'	
常用特征	110	21	26	6	2	0.67
加入面积衰减率	115	20	24	4	2	0.71
加入面积衰减率和质心偏心率	117	19	14	4	2	0.75
加入 3 个特征	132	14	9	4	2	0.82

3)迭代分割的有效性验证.在进行迭代分割的有效性评估之前,首先对算法的迭代次数进行了评估.由于迭代分割是在初分割的基础上进行的,往往迭代分割的次数不会太多,大部分单体的初分割结果在经历 1~2 次迭代分割后即可满足终止条件.因此,在单体分割的过程中设计了 4 种情况的对比:仅仅包含初分割(不含迭代分割),加入 1 次、2 次迭代分割和自适应迭代分割.其中,加入 1 次、2 次迭代分割和自适应迭代分割 3 种情况是在加入 SVM 分类器的条件下进行的,即通过 SVM 判断分割结果是否需要进一步迭代分割.这 4 种方法得到的单体分割评分值如表 3 所示.当加入迭代分割之后,分割评分值会有明显提升,这证明了迭代分割的有效性.

表 3 迭代分割的有效性验证结果

Tab.3 Validation results of iterative segmentation

迭代次数	评价指标					CSI
	X	Y	Y'	Z	Z'	
初分割	121	18	27	9	3	0.70
迭代 1 次(加入 SVM)	121	17	9	4	2	0.79
迭代 2 次(加入 SVM)	123	15	8	4	2	0.81
自适应迭代次数	126	12	6	4	2	0.84

4)SVM 分类器的选择及有效性能测试.本文中 SVM 分类器的训练用到了 60 个样本,其中正负样本各 30 例.由于我们关注的重点不是分类器的性能,而是利用一个分类器能够终止分割过程.因此,本文中分类器本身是可以替换的,选择哪一种分类器,相当于对于整体算法的一个调参过程.在测试中,我们尝试了不同类型的分类器,检测了各种分类器对本文算法性能的评分效果.

表 4 为 3 种分类器进行单体分割实验结果对

比,利用 CSI 作为评价指标,得到 3 种分类器对应的 CSI 评分值分别为 0.80、0.83 及 0.79. 可见,采用哪一种分类器对算法的分割效果影响不大,但使用 SVM 的分割结果略高,因此最终选取 SVM 作为本文算法使用的分类器.

表 4 3 种分类器进行单体分割实验结果对比

Tab.4 Comparison of experimental results of cells segmentation of three classifiers

分类器	评价指标					
	X	Y	Y'	Z	Z'	CSI
贝叶斯	120	12	9	6	2	0.80
SVM	122	15	4	4	2	0.83
随机森林	122	14	12	4	1	0.79

为了测试 SVM 的作用,在迭代分割中去掉 SVM 模型,也就是说,在每一个迭代分割过程中,所有分割结果不经过 SVM 模型,直接进入下一步迭代分割,用这种分割方法得到的单体分割评分值见表 5.

不加 SVM 分类得到的一次迭代与二次迭代分割结果分别为 0.76 和 0.71, 低于加入 SVM 的评分值,这是因为原本通过 SVM 分类器可能被判定为单体的初分割结果也被再次分割了,破坏了原本单体的结构,影响了分割结果. 可见,在迭代分割过程中加入 SVM 能够提高最终分割结果的评分值,这验证了 SVM 的有效性.

表 5 不加 SVM 控制迭代次数的实验结果

Tab.5 Experimental results without SVM to control the number of iterations

迭代次数	评价指标					
	X	Y	Y'	Z	Z'	CSI
迭代 1 次(不加 SVM)	112	18	11	4	2	0.76
迭代 2 次(不加 SVM)	109	21	17	4	2	0.71
自适应迭代次数	126	12	6	4	2	0.84

5)单体定量分割结果. 由于人工标记过程工作量较大,在实验中只使用了 6 个案例进行评估,这 6 个案例包含了 3 种风暴类型, 每个案例的时长约为 2 h. 表 6 给出了人工检验 3 种对流单体识别算法的统计评分结果. 从表 6 中可见,本文算法获得了最高的 CSI 评分值,达到了 0.84, 远高于单阈值方法和 SCIT 算法. 具体分析各种算法的统计结果,可以看出

本文算法性能获得提升的主要原因是算法在对流风暴发生分裂或合并的时候,有效的降低了漏报和误报的对流单体数目.

表 6 单体定量分割结果

Tab.6 Quantitative segmentation results of cells

分割算法	评价指标							
	X	Y	Y'	Z	Z'	POD	FAR	CSI
本文算法	198	7	12	17	3	0.90	0.08	0.83
单阈值算法	105	43	14	42	12	0.66	0.35	0.48
SCIT 算法	121	26	52	13	7	0.85	0.39	0.55

6)算法运行效率. 统计了 3 种算法在 10 个案例中的平均运行时间, 本文算法、单阈值方法及 SCIT 方法处理一张雷达反射率图片的平均时间分别为 1.858 s、0.051 s 和 0.143 s. 对于 3 种算法而言, 本文算法的运行时间最长, 这是由于本文方法是一种迭代的分割方法, 而且在迭代的过程中需要大量的数学形态学运算. 虽然本文方法运行时间较长,但相对于雷达的采样周期(5~6 min), 本文算法的运行时间可以忽略不计, 完全能够满足雷达数据在线处理的要求.

3 结论

针对对流风暴内部单体分割困难的问题, 本文提出了一种基于形态结构特征的对流单体自动分割方法. 与传统方法相比, 本文方法的特点是: 1) 在对流单体的识别过程中考虑到了对流单体的形态结构信息, 而不仅仅是反射率强度信息; 2) 对流单体的分割过程设计为一个迭代过程, 用于消除错误的单体分割结果; 3) 该方法是一种基于数据驱动的方法, 而传统方法是基于规则的目标识别方法. 实验结果显示, 本文算法在对流单体聚集或发生分裂与合并的时候, 能够更加有效的分割出对流单体的完整结构, 而且本文算法的量化评分值显著高于传统的 SCIT 方法和单阈值方法.

参考文献

[1] MAHONEY K, ALEXANDER M A, THOMPSON, G, *et al.* Changes in hail and flood risk in high-resolution simulations over Colorado's mountains[J]. *Nature. Climate, Change*, 2012, 2: 125-131.

[2] TAYLOR C M, BELUËÏÛ D, GUICHARD F, *et al.* Frequency of ex-

- trema Sahelian storms tripled since 1982 in satellite observations [J]. *Nature*, 2017, 544: 475—478.
- [3] WESTRA S, FOWLER H J, EVANS J P, *et al.* Future changes to the intensity and frequency of short-duration extreme rainfall [J]. *Reviews of Geophysics*, 2014, 52(3): 522—555.
- [4] DIXON M, WIENER G. TITAN: Thunderstorm identification, tracking, analysis, and nowcasting—A radar-based methodology [J]. *Journal of Atmospheric and Oceanic Technology*, 1993, 10 (6): 785—797.
- [5] JOHNSON J, MACKEEN P L, WITT A, *et al.* The storm cell identification and tracking algorithm: an enhanced WSR-88D algorithm [J]. *Weather and Forecasting*, 1998, 13 (2): 263—276.
- [6] 侯正俊, 潘多, 王磊. 改进气象雷达 TITAN 算法在灾害性天气预警中的应用研究 [J]. *大气科学学报*, 2018, 41(4): 561—568.
HOU Z J, PAN D, WANG L. Application research on disaster weather warning based on improved TITAN algorithm of Doppler weather radar [J]. *Transactions of Atmospheric Sciences*, 2018, 41 (4): 561—568. (In Chinese)
- [7] 杨吉, 刘黎平, 李国平, 等. 基于雷达回波拼图资料的风暴单体和中尺度对流系统识别、跟踪及预报技术 [J]. *气象学报*, 2012, 70(6): 1347—1355.
YANG J, LIU L P, LI G P, *et al.* A new techniques for storm cell and mesoscale convective systems identification, tracking and nowcasting based on the radar mosaic data [J]. *Acta Meteorologica Sinica*, 2012, 70(6): 1347—1355. (In Chinese)
- [8] CRANE R K. Automatic cell detection and tracking [J]. *IEEE Transactions on Geoscience Electronics*, 1979, 17 (4): 250—262.
- [9] LAKSHMANAN V, RABIN R, DEBRUNNER V. Multiscale storm identification and forecast [J]. *Atmospheric Research*, 2003, 67: 367—380.
- [10] LAKSHMANAN V, HONDL K, RABIN R. An efficient, general-purpose technique for identifying storm cells in geospatial images [J]. *Journal of Atmospheric and Oceanic Technology*, 2009, 26 (3): 523—537.
- [11] WANG P, LI C, ZHANG Y. An adaptive segmentation arithmetic adapted to intertwined irregular convective storm images [C]//2013 International Conference on Machine Learning and Cybernetics. Tianjin, China: IEEE, 2013: 896—900.
- [12] HOU J Y, WANG P. Storm tracking via tree structure representation of radar data [J]. *Journal of Atmospheric and Oceanic Technology*, 2017, 34 (4): 729—747.
- [13] ROBERT H M, SHAPIRO L G. *Computer and Robot Vision*, Vol. 1 [M]. Boston: Addison-Wesley Press, 1992: 28—48.
- [14] HAPP P N, DA COSTA G A, BENTES C *et al.* A cloud computing strategy for region-growing segmentation [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2016, 9(12): 5294—5303.
- [15] TREMEAU A, BOREL N. A region growing and merging algorithm to color segmentation [J]. *Pattern recognition*, 1997, 30(7): 1191—1203.
- [16] 王国权, 周小红, 蔚立磊. 基于分水岭算法的图像分割方法研究 [J]. *计算机仿真*, 2009(5): 255—258.
WANG G Q, ZHOU X H, YU L L. Image segmentation based on watershed algorithm [J]. *Computer Simulation*, 2009 (5): 255—258. (In Chinese)
- [17] 赵晓晴, 刘景鑫, 张海涛等. 色彩空间变换和基于距离变换的分水岭算法在白细胞图像分割中的应用 [J]. *中国医疗设备*, 2019, 34(7): 5—9.
ZHAO X Q, LIU J X, ZHANG H T, *et al.* Application of color space transformation and watershed algorithm based on distance transform in white blood cell image segmentation [J]. *Chinese Medical Devices*, 2019, 34(7): 5—9.
- [18] RAFAEL C, GONZALEZ R E, WOODS. *Digital image processing*, third edition [M]. Beijing: Publishing House of Electronics Industry, 2011: 497—501.
- [19] 张学工. 模式识别 [M]. 北京: 清华大学出版社, 2002: 60—81.
ZHANG X G. *Pattern recognition* [M]. Beijing: Tsinghua University Press, 2002: 60—81. (In Chinese)