

凝聚层次聚类的 MODIS 海冰外缘线提取算法与应用

王宁^{1,2,3}, 张晰³, 纪永刚³, 鲁统臻³, 于波²

(1. 山东省海洋生态环境与防灾减灾重点实验室, 青岛 266033; 2. 国家海洋局北海预报中心, 青岛 266033;
3. 国家海洋局第一海洋研究所, 青岛 266061)

摘要: 本文论述了海冰外缘线在我国海冰监测和预报中的重要作用, 分析了常规的海冰外缘线提取方法存在的不足。同时, 结合 MODIS 遥感影像和渤海悬浮泥沙分布的特点, 讨论了凝聚层次聚类的 MODIS 影像分割算法^[1]。该算法以影像光谱特性和形状特性作为判定规则, 通过加入影像网格化、碎斑和噪声去除等分析, 在优化冰水识别参数及分割结果提取海冰外缘线的基础上, 将该算法提取的结果进行了分析。实验表明, 该算法能够较好地去除碎斑和噪声。同时对 MODIS 遥感影像的分割, 得到理想的结果。

关键词: 凝聚层次聚类算法; 海冰外缘线; MODIS; 冰水识别

DOI: 10.3724/SP.J.1047.2011.00266

1 引言

海冰外缘线是指海水区域和海冰区域之间的交界线。海冰外缘线位置及其变化的监测, 对海冰边缘区域的海上交通、海洋渔业、近岸海上操作、近海油气开发、气象研究和其他海上军事活动, 具有重要的作用。诸如, 商船破冰需要明确海冰外缘线的位置, 避开海冰; 油气开发需要知道海冰边缘位置, 以便规划钻井和其他活动等。在极地和高纬度地区, 海冰外缘线是一个重要的气象指数, 对长期的全球海冰监测是必需的^[2], 同时, 对海上交通和作业, 海冰资源的开发都具有重要意义。

海冰外缘线的位置是反映结冰海域冰情的重要参数。在我国的海冰监测、预报业务中, 根据海冰外缘线到渤海各湾湾底的距离确定冰情等级。

常规的海冰外缘线提取方法建立在冰水识别基础上的, 例如, 对于卫星图像上为冰的一个像元 A, 判别其邻域中的 8 个像元的冰水变化。如果在经向或者纬向或者对角线方向的两个像元有冰水变化, 则认为 A 像元为冰外缘线的点^[3]。

在这种判别方法下, 如果在冰外缘线内的冰区内部存在着水区, 这些水区的边缘也会被判别为冰外缘线。对于这种误判, 采用人工方法去除。

图 1 为 2009 年 1 月 26 日 11 点 04 分 MODIS 数据假彩色合成图(R=1 通道, G=2 通道, B=1 通道)与采用常规的海冰外缘线提取方法得到的结果。

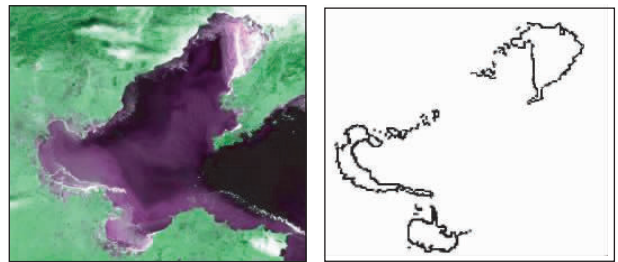


图 1 常规的海冰外缘线提取结果

Fig. 1 Result of conventional sea ice outflow boundary extraction

根据图 1 所示, 利用常规的海冰外缘线提取方法提取海冰外缘线时, 易对影像产生误判, 得到的结果存在大量的杂点、碎斑, 而且计算量大, 计算速度慢, 对于提取结果, 需要人工干预去除杂点和碎斑。

本文针对海冰预报业务需求, 利用凝聚层次聚

收稿日期: 2009-10-29; 修回日期: 2010-07-28.

基金项目: 国家“863”计划资助项目“遥感海洋生态环境监测技术系统”(2007AA092102); 国家自然科学基金项目“多时相 SAR 渤海薄冰厚度探测”(40906093)。

作者简介: 王宁(1978—), 女, 山东龙口人, 硕士研究生。从事海洋遥感应用研究。E-mail: wn780324@126.com

类算法的优势,在刘荣杰等人研究的基础上^[1],将凝聚层次聚类的遥感影像分割算法应用到海冰外缘线提取上。

2 凝聚层次聚类的遥感影像分割算法

聚类就是将数据对象分组变成多个簇,使同一个簇中的对象具有较高的相似性,而不同簇中的对象具有较大的相异性。良好的聚类方法产生的聚类结果具有簇内对象高度相似,簇间对象很少相似的特性。

层次聚类按层次分解方向可以分为凝聚的和分裂的层次聚类。凝聚的层次聚类是一种自底向上的策略,首先,它将每个对象作为一个簇,然后,合并这些原子簇为越来越大的簇,直到所有的对象都在一个簇中,或者某个终结条件被满足。分裂的层次聚类是一种自顶向下的策略,与凝聚的层次聚类相反,首先是将所有对象置于一个簇中,然后逐渐细分为越来越小的簇,直到每个对象自成一簇,或者达到了某个终结条件^[4]。

根据 MODIS 影像渤海海冰的特点,本文采用凝聚层次聚类方法^[5]提取了海冰外缘线。

2.1 算法的原理与流程

该算法将遥感影像中的每个像元点在初始状态时都当作一类,并根据一定的判定规则进行逐级合并,直到不能再合并为止。整个分割过程如图 2 所示,在初始状态时每个格子代表一个像元点,具有不同的像元值。

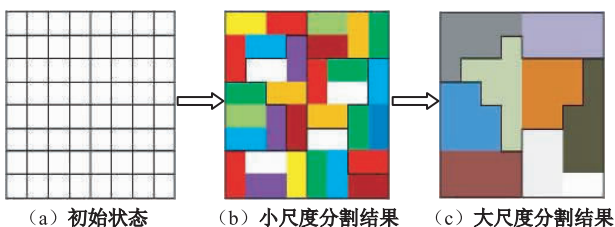


图 2 影像聚类过程示意图

Fig. 2 Diagram of image cluster process

算法结合遥感影像的特点,针对现有聚类算法聚类数目、初始聚类中心难以确定,以及局部最小的问题,作了如下分析:

(1)聚类相似性度量融合遥感影像的光谱特性和形状特性。

MODIS 遥感影像是中低分辨率多光谱数据,包含有丰富的光谱特性、空间特性,单纯利用遥感影像的光谱特性无法进行很好的分割,基于多特征(包括光谱特征、形状特征、干扰特征、区域地学特征等)的遥感影像才能得到较可靠的分割结果。本算法选用遥感影像的光谱特征和形状特征,用以影像分割,提高精度。

(2)以每个像元点为初始聚类中心,初始条件容易确定,能够发现任意形状的区域。

针对目前分割算法中存在的问题,本算法以每个像元点为初始聚类中心进行聚类,不断与周围区域进行合并,直到所有的区域无法合并,这种合并策略还可以发现任意形状的区域,避免了现有算法各向同性的特点,符合遥感影像的地物分布。

(3)减少影像分割后的杂点、碎斑。

针对影像分割后杂点、碎斑过多的问题,算法设定一个判断阈值,如果某一个区域内像元的个数小于给定的阈值,则把这个区域与周围最相似的区域合并,这样能够有效地减少影像分割后的杂点、碎斑,消除噪声带来的干扰。

算法的相似度度量即为影像相邻区域合并判断规则。因此,本文针对凝聚层次聚类算法的原理,给出了融合遥感影像的光谱特征和形状特征的凝聚层次聚类分割算法。

下面是融合遥感影像的光谱特性和形状特性的影像相邻区域合并判断规则:

任何一个区域的异质性质 H 可由四个变量计算得到,如公式(1):

$$H = W_{color} \times H_{color} + W_{shape} \times H_{shape} \quad (1)$$

其中, H_{color} 是光谱异质性值, H_{shape} 为形状异质性值, W_{color} 是光谱异质性权重因子, W_{shape} 为形状异质性权重因子。且 $0 \leq W_{shape} \leq 1$, $0 \leq W_{color} \leq 1$, $W_{color} + W_{shape} = 1$ 。如果用户想着重强调遥感影像光谱特性,可将 W_{color} 设置大一些。如果认为遥感影像的形状特征更为重要,可将 W_{shape} 设置的大一些。

光谱异质性值 H_{color} 取决于区域内像元各个波段的方差以及像元的数目(见式(2)):

$$H_{color} = \frac{1}{m} \sum_{k=1}^m [N_{merge} \times \sigma_k - (N_{area1} \times \sigma_k^{area1} + N_{area2} \times \sigma_k^{area2})] \quad (2)$$

其中, m 是遥感影像波段数,即通道数目。 N_{merge} 表示合并后新区域的像元个数, N_{area1} 表示合并区域 1 的像元个数, N_{area2} 表示合并区域 2 内的像元个数, σ_k 表

示区域内所有像元的第 k 个波段的方差(式(3))。

$$\sigma_k = \sqrt{\frac{1}{n} \sum_{i=1}^n (V_{i,k} - \frac{1}{n} \sum_{i=1}^n V_{i,k})^2} \tag{3}$$

其中, n 表示区域内像元的个数, $V_{i,k}$ 表示区域内第 i 个像元的第 k 个波段的值。

形状异质性值 H_{shape} 由区域周长 L 和区域内像元的个数 N 来表示(见式(4)):

$$H_{shape} = N_{merge} \times \frac{L_{merge}}{\sqrt{N_{merge}}} - (N_{area1} \times \frac{L_{area1}}{\sqrt{N_{area1}}} + N_{area2} \times \frac{L_{area2}}{\sqrt{N_{area2}}}) \tag{4}$$

其中, L_{merge} 表示合并后区域的周长, N_{merge} 表示合并后区域内的像元个数, L_{area1} 、 N_{area1} 分别表示合并前区域 1 的周长和像元个数, 而 L_{area2} 、 N_{area2} 则表示合并前区域 2 的周长和像元个数。

本文根据计算两个相邻区域的异质性值 H , 与输入的分割阈值 T 进行比较, 如果 $H < T$, 则对两个相邻的区域合并, 否则不合并。该合并判断方法的好处是综合了遥感影像的光谱特征和形状特征, 使得聚类效果更精确。另外, 可以通过设置不同的分割阈值, 实现遥感影像的不同尺度分割。而且可以根据设置不同的因子权重来突出影像的光谱特征或者形状特征, 以获得更好的分割效果。

算法流程如下:

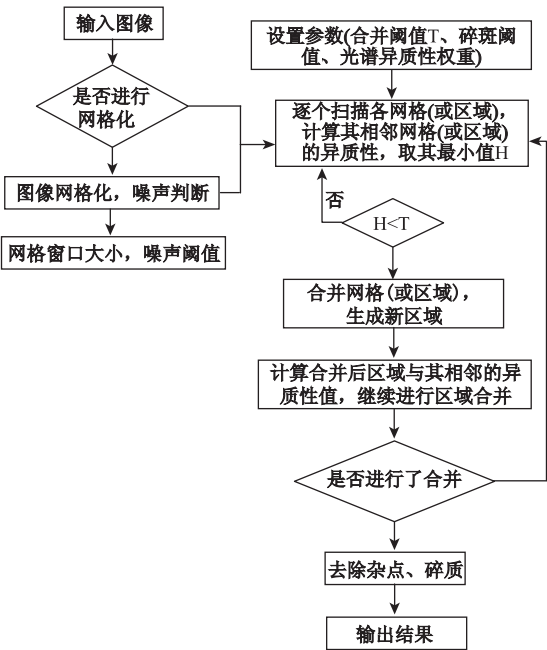


图 3 凝聚层次聚类算法的流程图
Fig. 3 Flow chart of agglomerative hierarchical clustering algorithm

2.2 凝聚层次聚类的 MODIS 彩色图像分割

凝聚层次聚类的遥感影像分割算法, 以影像光谱特征和形状特征作为判定规则, 通过加入影像网格化、碎斑和噪声去除等功能, 可有效提高影像分割结果的可用性和准确性^[1]。

本文选用辽东湾、莱州湾和渤海湾均有结冰现象的 2009 年 1 月 13 日 MODIS 数据(见图 4), 按流程图 3 采用不同的阈值进行分割, 得到如图 5-6 结果。

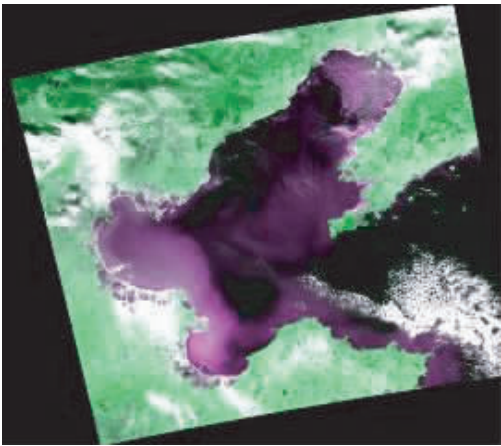


图 4 2009 年 1 月 13 日 MODIS 假彩色图像
Fig. 4 MODIS false color image, January 13, 2009

光谱异质性阈值为 40, 颜色差异性阈值为 10, 碎块大小阈值为 400, 光谱异质性权重为 0.8, 形状异质性权重为 0.2, 分割后的 MODIS 图像如图 5。

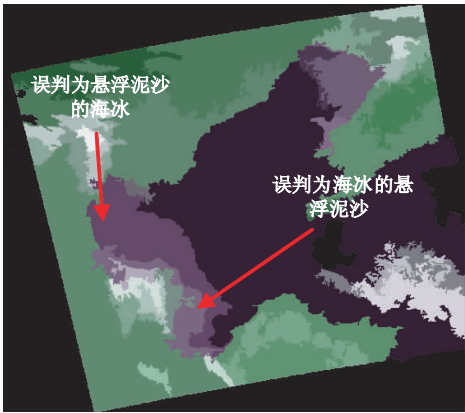


图 5 MODIS 图像的分割结果之一
Fig. 5 Segmentation result A of MODIS image

光谱异质性阈值为 40, 颜色差异性阈值为 10, 碎块大小阈值为 400, 光谱异质性权重为 0.9, 形状异质性权重为 0.1, 分割后的 MODIS 图像如图 6。

从图 5 和图 6 看出, 设置不同的图像分割阈

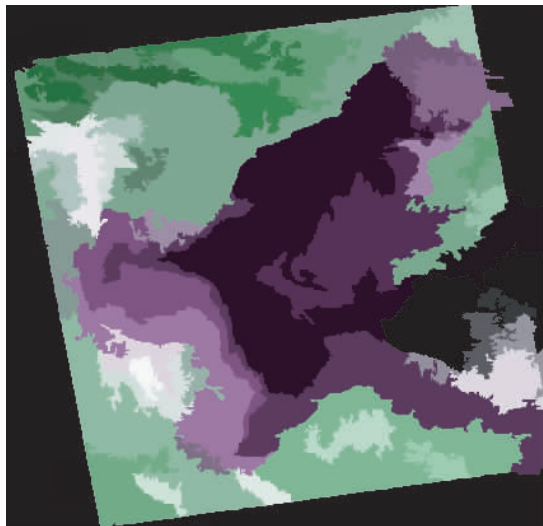


图 6 MODIS 图像的分割结果之二

Fig. 6 Segmentation result B of MODIS image

值,得到的分割结果也不同。直接用 MODIS 彩色图像进行分割,如果各阈值设置不准确,容易误判海冰与悬浮泥沙。对于非遥感专业用户来说,设置准确的光谱异质性阈值、颜色差异性阈值、光谱异质性权重和形状异质性权重比较困难。基于冰水识别的二值数据图像能将这几个方面对图像分割的影响降低到最小。

为了满足海冰预报业务的需求,本文在冰水识别的基础上,对得到的冰水二值图进行图像分割,提取海冰外缘线。

3 凝聚层次聚类的遥感影像分割算法的改进

该算法的改进是在对冰水识别进行参数优化的基础上进行的^[6]。

3.1 冰水识别参数优化

用于 MODIS 海冰检测的 NASA 标准算法^[7]对于厚冰识别效果很好,但对于薄冰和悬浮泥沙等悬浮物浓度高的地区不易区分^[8]。由于黄河口泥沙对海水水色改变较大,经常在渤海湾和莱州湾出现海水反射率与海冰反射率接近乃至等同的情况,所以,往往会发生对这个地区海冰的误判。在红外通道,悬浮泥沙与海水的亮温基本相同,而与海冰的亮温有明显的差别。利用吴龙涛等人^[9]的研究结果,采用 MODIS 第 1 通道和第 31 通道联合进行

冰水识别,能减少悬浮泥沙对冰水识别的影响。

本文根据海冰预报业务的需求,对全渤海或单独对辽东湾、渤海湾或莱州湾进行冰水识别,步骤如下:

(1)如果用户选择的是对全渤海进行冰水识别,根据渤海海域海冰、悬浮泥沙分布的特点^[10],将渤海海域分成 3 个子区块:辽东湾、渤海湾和莱州湾,对 3 个子区块分别采用不同的阈值进行冰水识别。

(2)分别对 3 个子区块进行反射率和辐射率直方图统计分析,根据分析结果提取出 3 组阈值,利用 $CH1 > T1$ 和 $CH31 > T2$ 提取 3 个子区块的海冰,其中,CH1 和 CH31 分别为 MODIS 第 1 通道和第 31 通道的数据,T1 和 T2 分别为第 1 通道和第 31 通道的阈值。合并 3 个子区块,输出全渤海海冰检测结果。

(3)如果用户选择的是某个单独的子区块,利用 $CH1 > T1$ 和 $CH31 > T2$ 提取出海冰。参数优化结束后,得到冰水识别二值图,其中,海冰为 1,海水和陆地为 0。

3.2 冰水识别二值图的凝聚层次聚类分割结果

本文利用参数优化的方法对 2009 年 1 月 26 日 11 点 04 分 MODIS 数据(见图 1)进行冰水识别,得到冰水二值图;对得到的二值图采用凝聚层次聚类的遥感影像分割算法进行分割,结果如图 7-8。



图 7 渤海海冰冰水识别结果二值图

Fig. 7 Two-value image of Bohai Sea ice-water identification result

光谱异质性阈值为 40,颜色差异性阈值为 10,

碎块大小阈值为 400,光谱异质性权重为 0.8,形状异质性权重为 0.2 的渤海海冰冰水识别二值图分割结果如图 8:

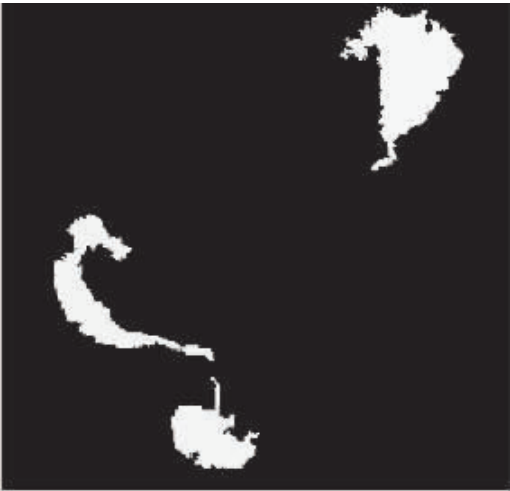


图 8 渤海海冰冰水识别二值图分割结果
Fig. 8 Segmentation result of the two-value image of Bohai Sea ice-water identification

从图 8 可看出,由于二值图的数据只有两个值 0 和 1,光谱异质性和颜色差异性很好区分,所以,在分割过程中不需过多考虑这两个阈值,只需考虑碎块大小阈值即可。碎块大小阈值是针对影像分割后杂点、碎斑过多的问题设定的判读阈值,如果某一个区域内像元点的个数小于给定的阈值,则把这个区域与周围最相似的区域合并,这样能够有效地减少影像分割后的杂点、碎斑,消除噪声带来的干扰。

4 影像分割的海冰外缘线提取及验证

4.1 影像分割的海冰外缘线提取

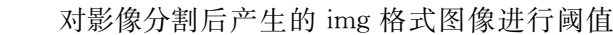
对影像分割后产生的格式图像进行阈值分割,利用边缘检测方法实现海冰外缘线提取。本文选用最经典的梯度算子——Sobel 算子^[11]提取海冰外缘线。

图 9 是对 2009 年 1 月 13 日 MODIS 数据,利用上述方法和算子在冰水识别的基础上,提取的海冰外缘线结果。

4.2 提取结果验证

本节针对基于冰水识别二值图的凝聚层次聚类影像分割算法进行结果验证。



图 9 冰水识别的边缘检测结果
Fig. 9 Examination result of sea ice outflow boundary based on ice-water identification

海冰外缘线的现场调查是利用飞机或船舶进行监测的。由于条件限制,飞机或船舶不能获得与卫星同步的海冰外缘线资料,而且飞机或船舶获得的外缘线资料只是针对部分海域,不能在 MODIS 影像上很好地反映出来。本文利用晴空条件下的 MODIS 数据,通过人眼经验判读得到海冰外缘线位置,检验基于冰水识别二值图的凝聚层次聚类影像分割算法的准确性。

本文在冰水识别基础上,利用 2009 年 1 月 26 日 11 点 04 分的 MODIS 数据提取渤海海域海冰外缘线。图 10 中,海冰为 1,陆地和海水为 0。

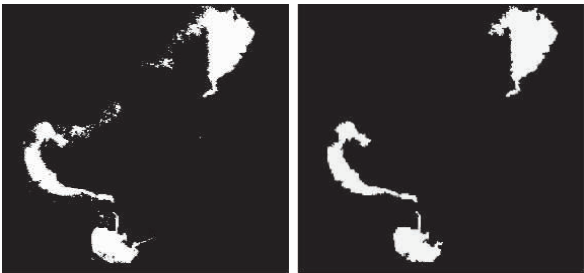


图 10 渤海海域海冰检测结果二值图与分割结果
Fig. 10 Two-value image of Bohai Sea ice-water examination and its segmentation result

海冰外缘线目测假彩色合成波段组合结果: R = 波段 1, G = 波段 2, B = 波段 1 的结果与利用程序半自动生成的海冰外缘线结果比较如图 11 所示:

通过图 11 的比较可知,利用冰水识别二值图

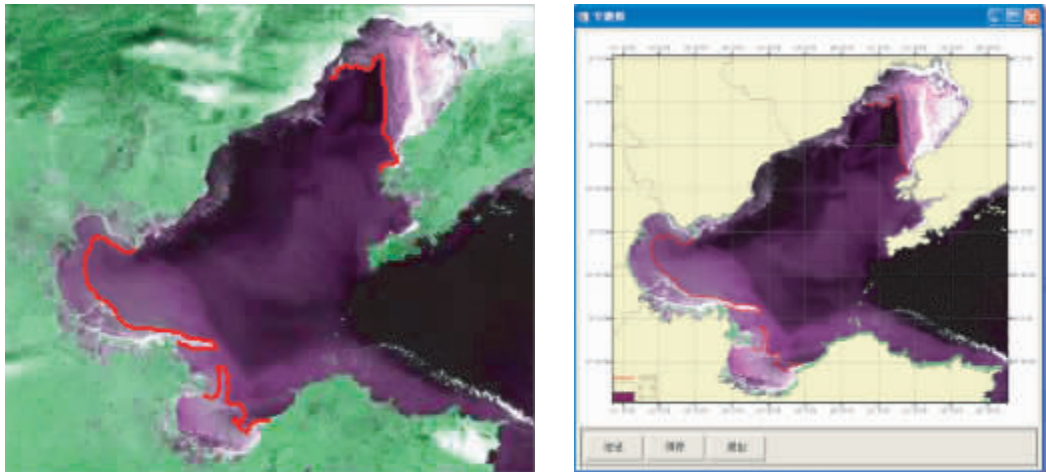


图 11 海冰外缘线目测与产品结果比较

Fig. 11 Comparison of the sea ice outflow boundary from MODIS false color image and from the product result

的凝聚层次聚类算法分割图像,提取海冰外缘线,得到的海冰外缘线位置与利用海冰反射率特征目测得到的结果一致,说明利用本算法提取的海冰外缘线与实际外缘线的位置是相符的。

5 结论与讨论

本文针对常规海冰外缘线提取方法的缺点,讨论了凝聚层次聚类的遥感影像分割算法在 MODIS 彩色图像分割中的应用,得到利用多个不同阈值分割的结果。采用分区域读取阈值的方法进行冰水识别,进一步减少了悬浮泥沙对海冰的影响;在此基础上,对凝聚层次聚类的遥感影像分割算法进行了改进,给出了基于冰水识别二值图的 MODIS 数据提取海冰外缘线的方法。

在业务化海冰信息提取过程中,检验算法的准确性,不断完善海冰外缘线提取算法,还有待今后深化研究。

参考文献:

[1] 刘荣杰. 基于凝聚层次聚类的高分辨率遥感影像分割算法研究[D]. 青岛大学,2008.
[2] 纪永刚. 基于微波图像的辽东湾海冰典型要素信息提取[D]. 中国科学院海洋研究所,2006,1-4.

[3] 罗亚威,张蕴斐,孙从容,等. “海洋 1 号”卫星在海冰监测和预报中的应用[J]. 海洋学报,2005,27(1):7-18.
[4] 石剑飞,闫怀志,牛占云. 基于凝聚的层次聚类算法的改进[J]. 北京理工大学学报,2008,1(28):66-69.
[5] 王宁. 基于 MODIS 数据的渤海海冰重要参数提取技术与探测系统[D]. 中国海洋大学,2009.
[6] Liu Rongjie, Zhang Jie, Song Pingjian, Shao Fengjing, Liu Guanfeng. An Agglomerative Hierarchical Clustering based High-Resolution Remote Sensing Image Segmentation Algorithm[C]. International Conference on Computer Science and Software Engineering,2008,403-406.
[7] Hall D K, Riggs G A, Salomonson V V. Algorithm Theoretical Basis Document (ATBD) for the MODIS Snow and Sea Ice-Mapping Algorithms[ED/OL]. 2001-09-01.
[8] 郭衍游,焦明连. 利用 MODIS 数据反演渤海海冰分布[J]. 淮海工学院学报(自然科学版),2010,1(19):84-87.
[9] 吴龙涛,吴辉碇,孙兰涛,等. MODIS 渤海海冰遥感资料反演[J]. 中国海洋大学学报,2006,36(2):173-179.
[10] 王芳,李国胜. 海洋悬浮泥沙二元特征参数 MODIS 遥感反演模型研究[J]. 地理研究,2007,6(26):1186-1197.
[11] 靳鹏飞. 一种改进的 Sobel 图像边缘检测算法[J]. 应用光学,2008,4(29):625-628.

A Method of MODIS Sea Ice Edge Extraction Based on Agglomerative Hierarchical Clustering Method

WANG Ning^{1,2,3}, ZHANG Xi³, JI Yonggang³, LU Tongzhen³, YU Bo³

(1. *Shandong Provincial Key Laboratory of Marine Ecology and Environment & Disaster Prevention and Mitigation, Qingdao 266033, China*; 2. *North China Sea Marine Forecasting Center of SOA, Qingdao 266033, China*; 3. *The First Institute of Oceanography, SOA, Qingdao 266061, China*)

Abstract: In this paper we suggested the importance of sea ice edge in our country's sea ice monitoring and forecasting. The sea ice edge is the boundary between water and ice. Monitoring of sea ice edge's position and the change is vital to sea water exchange, marine fishery, nearshore marine operation, offshore oil-gas exploration, meteorological observation and so on. The conventional method of sea ice edge extraction often has the shortcomings of massive mixed spots existing and the slow computation speed. Clustering is a way to group objects into more than one cluster, so that the same cluster objects have a higher similarity, and an object in a different cluster is larger. Agglomerative hierarchical clustering algorithm is a bottom-up policy. It first considers each object as a cluster, and then combines these clusters into larger cluster, until all the objects in a cluster, or one condition is met. Based on the agglomerative hierarchical clustering algorithm, combined with MODIS images and the characteristics of Bohai suspended sediment distribution, we discussed the MODIS image segmentation algorithm. The algorithm regards spectral characteristics and shape properties as the rule, joining image meshing, breaking grouper and noise removal and other functions, and optimizing the ice-water identification parameters. This algorithm can effectively improve the availability and accuracy of image segmentation results. We combined this algorithm and edge detection algorithms to extract sea ice edge of the image segmentation results. The algorithm has a better removal of broken spot and noise. Comparing the result of interpretation of MODIS data and the result from the algorithm, we can conclude that the algorithm can better complete the MODIS image segmentation and on this basis we can have a better result of the sea ice edge.

Key words: agglomerative hierarchical clustering method; sea ice edge; MODIS; ice-water identification