

基于监督学习的日冕暗化检测与提取算法

田红梅¹ 彭博¹ 李天瑞¹ 谢宗霞²

(西南交通大学信息科学与技术学院 成都 610031)¹ (天津大学软件学院 天津 300072)²

摘要 日冕物质抛射(CME)是空间灾害天气的重要驱动源,而日冕暗化(dimming)被认为是 CME 初发的主要表征,对理解和预测 CME 具有重要作用。基于极紫外成像望远镜(EIT)和大气成像仪(AIA)的观测数据,实现了图像中日冕暗化现象的检测与提取。通过分析差分图中与暗化现象相关的图像统计特征,采用 Adaboost 分类算法检测暗化现象的发生,进而分割出日冕暗化区域。实验表明,提出的算法较现有算法能更准确有效地检测和提取日冕暗化区域,为分析日冕暗化特性提供了研究基础。

关键词 日冕物质抛射,日冕暗化,Adaboost 分类,图像分割

中图分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.5.009

Coronal Dimming Detecting and Extracting Algorithm Based on Supervised Learning

TIAN Hong-mei¹ PENG Bo¹ LI Tian-rui¹ XIE Zong-xia²

(School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, China)¹

(School of Computer Software, Tianjin University, Tianjin 300072, China)²

Abstract Coronal mass ejections (CMEs), which release huge quantities of matter and electromagnetic radiation into space above the sun's surface, are considered as one of the driven sources of space weather. Coronal dimming is now viewed as the important characteristic of CME. Dimming can help understand, predict and locate the occurrence of CME. Based on the observed data from extreme ultra-violet imaging telescope (EIT) and atmospheric imaging assembly (AIA), this paper implemented the coronal dimming detection and extraction. By analyzing the statistical characteristics of the difference images related to dimming, we applied Adaboost classification algorithm into dimming detection, and then segmented the coronal dimming region. The experiment results show that the proposed algorithm can effectively detect and extract the coronal dimming areas. Our work establishes the basis for analysis of coronal dimming features.

Keywords Coronal mass ejections, Coronal dimming, Adaboost classification, Image segmentation

1 引言

地球周围的空间环境状况主要由太阳的辐射状态决定,而且受到太阳辐射活动变化的制约。日冕物质抛射是太阳系中最壮观、最频繁的大尺度活动现象。CME 通常伴随很多物理过程,如耀斑、爆发日珥、射电爆发、高能粒子事件和日冕暗化等。对这些伴随物理过程开展细致深入的研究,可以帮助我们更好地理解 CME 的起源、传播和演化规律。日冕暗化是指日冕在一定区域随时间的推移而辐射下降的一种变化。日冕暗化由于与 CME 的足点紧密相关,而且还能提供关于 CME 源区和演化的一些信息,近些年来受到广泛的关注^[1-4]。

随着对暗化与 CME 关系研究的深入,目前的研究工作已经从最初两种事件简单的位置和时间相关分析,逐步深入到研究两种事件的定量物理关系。这就要求我们能够更加准确地定位天文图像中的暗化事件的区域。因此,设计准确的

检测和提取暗化区域是定量研究的基础。

1995 年,搭载在 SOHO 卫星上的极紫外成像望远镜实现了低日冕的观测,视场延伸到日心距大约 1.4R 处,它有 4 个观测波段:171、195、284 和 304Å^[5]。2010 年发射成功的 SDO 卫星上搭载的太阳大气成像仪将视场延伸到日面边缘以外 0.3R 处。它有 10 个观测波段,其中 6 个波段观测日冕等离子体:131、94、335、211、193 和 171Å,为日冕暗化的研究提供了更高分辨率的观测数据。本文基于 195Å 的 EIT 及 211、193 和 171Å 的 AIA 图像数据,研究了日冕暗化的自动检测和提取算法。

本文第 2 节介绍问题的研究背景和相关工作;第 3 节描述暗化检测的算法与实验;第 4 节利用随机游走算法自动提取暗化区域,并且与区域增长、阈值方法进行对比;最后总结全文并展望下一步工作。

到稿日期:2014-02-18 返修日期:2014-04-16 本文受国家自然科学基金(61202190,61175047,61105054),中央高校基本科研业务费专项资金(2682013CX055)资助。

田红梅(1990-),女,硕士生,主要研究领域为图像处理、数据挖掘,E-mail:hm_tian@foxmail.com;彭博(1980-),女,博士,讲师,硕士生导师,主要研究领域为机器视觉、模式识别;李天瑞(1969-),男,博士后,教授,博士生导师,主要研究领域为智能信息处理、数据挖掘、粗糙集、粒计算、云计算等;谢宗霞(1981-),女,博士,副教授,主要研究领域为图像处理、机器学习。

2 相关工作

随着获取的观测数据量越来越大,人们开始关注日冕暗化的自动检测和提取方法。近年来,关于暗化的研究图像主要分为两种:(1)基于差分的图像^[6-9],其又分为基准图差分(BD)图像和时变差分(RD)图像两种。差分图能消除原始图上不必要的干扰,更加突显出暗化区域。BD图像由每一幅图像都减去同一幅暗化前的图像得到。RD图像是由每一幅图像减去与其相对应的前一时刻暗化图像得到。(2)基于原始图像的研究方式^[7,10]。

Podladchikova等^[6]提出了NEMO算法,其基于RD图像的统计特征进行日冕暗化检测,并且基于BD图像,用区域增长的方法提取暗化区域。该方法不能用于检测小规模暗化区域,受局部噪声干扰较大。Attrill等^[7]改进了NEMO算法,即基于原始图像的局部统计特征进行日冕暗化检测,减少了计算时间,提高了检测小规模暗化的能力;另外,在暗化提取时,计算了无暗化BD图像的均值并利用 1σ 阈值来判断暗化像素点。Attrill等^[8]利用水平集方法在BD图像上对暗化的扩展和恢复进行检测跟踪,其效果与区域增长算法较为接近,但需要初始化暗化区域轮廓线,无法实现自动检测功能。Reinard等^[9]基于BD图像,采用阈值方法在规定的ROI区域内检测暗化像素,计算了暗化尺寸。Krista等^[10]将CODIT应用于暗化的检测,他们还指出已经报道的各种暗化和CME检测算法都存在很大误差,导致不同的检测方法分析的结果有较大差异。由上可看出,目前对暗化区域的检测和提取技术还不够成熟,方法的普适性和推广性较差,检测精度不够高,这些都影响现有算法统计出来的物理结果的可信度。

相对于RD图像来说,BD图像能更好地展示暗化的形态和物理特征。RD图像能展示暗化的变化特征,但如果连续的两个时刻暗化没有变化,那么RD图像上就可能丢失暗化现象。所以,本文认为BD图像是研究暗化更好的选择。

本文主要以Podladchikova和Attrill等^[6,7]的工作为基础,提出一种新的自动检测和提取方法。由于NEMO中的统计方法在小规模暗化发生时并不能很好地从整张图像上的数据特征上体现出来,受Attrill等^[7]基于局部图像统计的启发,本文用滑动窗口方式对BD图像进行切分,分别计算每分块图像的统计特征,以增强暗化区域对统计数据的影响效果。为了实现自动检测和提取,本文将Adaboost分类方法应用到日冕暗化的检测中,利用随机游走算法提高暗化区域的分割精度,通过实验将本文方法与区域增长^[6]和阈值方法^[7]的结果进行了比较。

3 日冕暗化检测

Adaboost是一种简单有效的分类算法,其核心思想是针对同一个训练集训练不同的分类器(弱分类器),然后把这些弱分类器集合起来,构成一个更强的最终分类器(强分类器)。近年来,该算法广泛地应用于癫痫预测^[11]、噪声检测^[12]、火焰预测^[13]等实际问题中。本文使用Adaboost算法来实现日冕暗化的检测工作。

3.1 预处理

从图像序列中挑选一幅图像为基准图像(Base Image),用于计算BD图。在BD图中划分子块,计算相关的统计信息,最后将数据归一化。该预处理流程如图1所示。

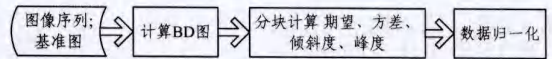


图1 数据预处理

计算BD图时需要计算当前图像与基准图像之间的时间差,使用Solarsoft^[14]中的drot_map.pro技术校正太阳自转的角度。图2所示为1997年5月12日以4:34时刻图像为Base Image所得到的4幅BD图,其中暗化区域表现为低灰度值部分。

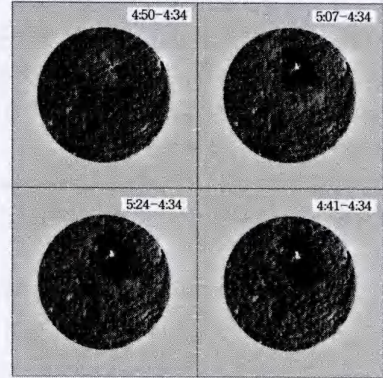


图2 BD图像,基准图为4:34

为进一步分析暗化区域,BD图中日冕区域分为8个分块,每个分块为 $1/4$ 圆,按45度旋转,如图3所示。计算每个分块的像素在整个过程中均值、方差、倾斜度和峰度的变化。

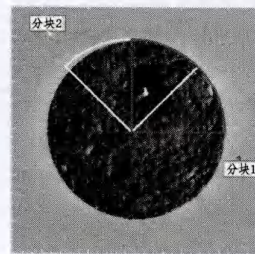
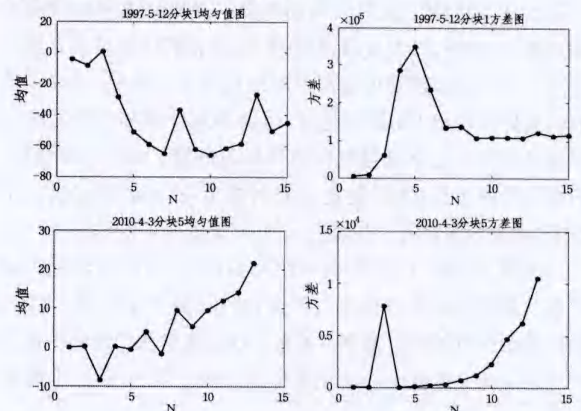


图3 分块示意图

每次暗化发生时,均值、方差、倾斜度和峰度的幅度不同,如图4所示,1997年5月12日和2010年4月3日的两次暗化现象的幅值就有较大差异。



注:图中N表示图片编号

图4 均值、方差对比图

利用式(1)对每个分块的均值做归一化处理。

$$x_i' = \frac{x_i - x_1}{|x_i - x_1|_{\max} + 1} \quad (1)$$

其中, x_i' 表示归一化之后的均值; x_i 表示第 i 个点的均值; \bar{x} 表示该区域在此时间段内的所有均值; x_1 表示第 1 个点的均值; $|x - x_1|_{\max}$ 表示该区域在此时间段内均值相对于 x_1 的最大绝对值。式(1)的分母加上 1, 是为了防止 $|x - x_1|_{\max}$ 为零而造成数据丢失。方差、倾斜度和峰度的归一化与均值同理。

3.2 利用 Adaboost 实现检测

通过预处理, 可得每幅图中 8 个分块的均值、方差、倾斜度和峰度的变化轨迹(后文简称轨迹)。检测分为训练和测试两个步骤, 均采用 Adaboost 方法对轨迹上的点进行分点。测试时如发现暗化点, 则表示该时间内该分块上有暗化现象发生, 也即整个日冕上有暗化现象发生; 否则, 表示未能检测出暗化现象。训练过程和测试过程的描述如图 5 和图 6 所示。

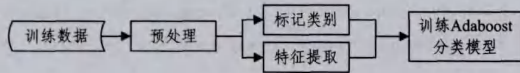


图 5 训练过程

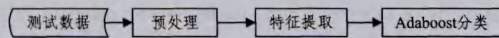


图 6 测试过程

3.2.1 训练

在标记的暗化图像序列上提取训练数据, 经过预处理后, 提取每条轨迹上每个点的特征。

(1) 类别

基于文献[6]提出的暗化区域的统计值特征, 将这些轨迹点分为两个类别, 记作 $C = \{-1, 1\}$, -1 表示可检测出暗化现象; 1 表示未能检测出暗化现象。通过观测图像和轨迹上 4 个值的变化情况, 对每个轨迹点的类别进行标记。如图 4 中 2010 年 4 月 3 日分块 5 的轨迹上, 将第 3 个点标记为 -1 。

(2) 特征描述

对轨迹上的每个点 i , 提取其特征向量 $I(i)$ 。

$$I(i) = \{x_i', y_i', s_i', k_i'\} \quad (2)$$

其中, x_i', y_i', s_i', k_i' 分别表示轨迹点的归一化均值、方差、峰度和倾斜度。然后, 将获取到的特征和类别作为训练过程的输入数据, 即可得到训练好的分类器模型。

3.2.2 测试

将某个新时间段的图像序列进行预处理, 再如 3.2.1 节中所述提取测试数据的轨迹上每个点的特征数据。然后将得到的数据输入到训练好的分类器中, 即可得到每个轨迹点的分类结果。若有分类结果为 -1 的点, 则表示检测出该段时间内有暗化现象发生; 否则, 即未能检测出暗化现象。

3.3 实验

本文采集 SOHO EIT¹⁾ 和 SDO AIA²⁾ 的暗化现象数据, 经过预处理后, 挑选了 351 个轨迹点数据, 其中负样本(未能检测出暗化现象, 类标为 1)187 个, 正样本(能检测出暗化现象, 类标为 -1)164 个。将数据平均分成 3 份, 每份 117 个特征点数据, 用于算法的交叉验证。每次拿两份做模型训练, 剩下的一份用于测试, 表 1 给出了检测的准确度结果。

表 1 测试准确度统计

| 训练数据 | 测试数据 | Adaboost 准确度 | SVM 准确度 |
|---------|------|--------------|---------|
| (1) (2) | (3) | 82.05% | 76.07% |
| (1) (3) | (2) | 82.05% | 76.92% |
| (2) (3) | (1) | 78.63% | 76.92% |

注: 其中(1) (2) (3)依次代表 3 份数据。

由表 1 可以得出, Adaboost 准确度的平均值为 80.91%, SVM 的平均准确度为 76.64%, 即这两种算法均能检测出大多数暗化现象, 而 Adaboost 效果略优于 SVM 算法。

4 日冕暗化提取

利用第 3 节中的方法实现日冕暗化的检测以后, 通过图像分割实现日冕暗化区域的自动提取。随机游走分割算法需要用户指定一组分别代表每一特定区域的种子点, 比较随机游走者从每个像素点游走到种子点的概率, 以最大概率为准则对各像素点赋予相应的类别, 从而实现图像分割^[15]。该算法在许多图像上都表现出很好的性能, 本文将其应用到天文领域, 实现暗化区域的提取, 并与文献[6]的区域增长、文献[7]的阈值方法进行对比。由于目前获得暗化图像数量有限, 且差分图像噪声较大, 本实验在可人为观测的 6 张暗化现象差分图像上进行分割, 对实验结果进行对比。

首先在 BD 图像中找到像素值最小的一部分像素点, 并从中选择最大连通区域作为暗化区域的种子点。同时, 挑选像素值最大的一部分像素点作为非暗化区域的种子点, 如图 7 第二行所示, 其中白色代表暗化区域的种子点, 黑色表示非暗化区域的种子点。然后用随机游走的方法直接提取暗化区域。整个过程无需人工干预, 实现了暗化区域的自动提取。本文得到的结果如图 7 第 3 行所示, 其它两种方法的结果如图 4、5 行所示。

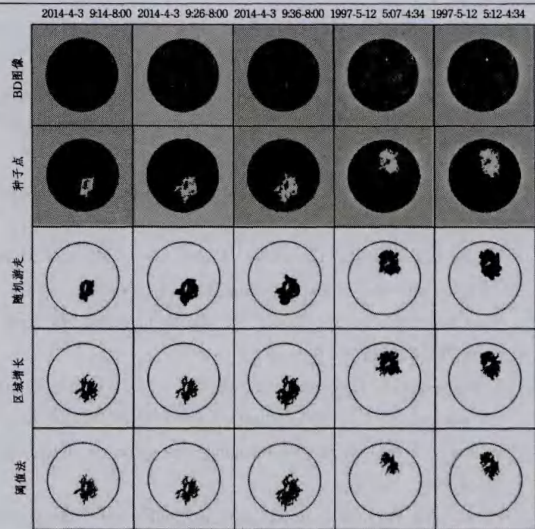


图 7 采用不同暗化提取算法得到的结果

随机游走方法相比于其它两种算法来说, 不需要中值过滤的过程, 而且考虑了图像的全局信息, 因此能得到比其它两种算法更完整、连续的暗化区域, 分割结果与人的主观视觉习

¹⁾ <http://umbra.nascom.nasa.gov/eit/eit-catalog.html>

²⁾ <http://jsoc.stanford.edu/data/aia/>

惯更一致;而区域增长和阈值方法更容易受周围噪声的干扰。

结束语 本文实现了日冕暗化的自动检测和提取工作。通过分块地进行数据分析,增强了暗化的统计特征,利用 Adaboost 分类方法对暗化进行检测,获得了一定效果。将随机游走方法和检测的结果相结合实现自动提取暗化区域,可以克服区域增长和阈值方法容易受周围噪声干扰的缺点,能获得更加完整、连续的日冕暗化区域。目前的数据集相对较小,后期将扩大数据测试算法效果。

由于本文的暗化检测和提取建立在 BD 图像的基础上,基准图的选择显得十分重要,否则容易受其它现象的干扰。如何合理选择基准图是今后需要完善的工作。

参 考 文 献

- [1] Harrison R A, Lyons M. A spectroscopic study of coronal dimming associated with a coronal mass ejection[J]. *Astronomy and Astrophysics*, 2000(358):1097-1108
- [2] Hudson H S, Lemen J R, Webb D F. Coronal X-Ray Dimming in Two Limb Flares[C]// *Proceedings of a Yokkoh Conference*. 1996;379-382
- [3] Gopalswamy N, Hanaoka Y, Hudson H S. Structure and dynamics of the corona surrounding an eruptive prominence[J]. *Advances in Space Research*, 2000, 25(9):1851-1854
- [4] Hudson H S, Lemen J R, Cyr O C S, et al. X-ray coronal changes during halo CMEs[J]. *Geophysical Research Letters*, 1998, 14(25):2481-2484
- [5] Delaboudinière J P, Artzner G E, Brunaud J, et al. EIT: Extreme-ultraviolet Imaging Telescope for the SOHO mission[J]. *Solar Physics*, 1995, 162(1/2):291-312
- [6] Podladchikova O, Berghmans D. Automated Detection of Eit

Waves And Dimmings[J]. *Solar Physics*, 2005, 228(1/2):265-284

- [7] Attrill G D R, Wills-Davey M J. Automatic Detection and Extraction of Coronal Dimmings from SDO/AIA Data[J]. *Solar Physics*, 2010, 262(2):461-480
- [8] Attrill G, Nakwacki M S, Harra L K, et al. Using the Evolution of Coronal Dimming Regions to Probe the Global Magnetic Field Topology[J]. *Solar Physics*, 2006, 238(1):117-139
- [9] Reinard A A, Biesecker D A. Coronal Mass Ejection-Associated Coronal Dimmings[J]. *The Astrophysical Journal*, 2008, 674(1):576
- [10] Krista L D, Reinard A. Study of the Recurring Dimming Region Detected at AR 11305 Using the Coronal Dimming Tracker (CoDiT) [J]. *The Astrophysical Journal*, 2013, 762(2):91
- [11] Ayinala M, Parhi K K. Low complexity algorithm for seizure prediction using Adaboost[C]// *Proceedings of the Engineering in Medicine and Biology Society (EMBC)*. San Diego, 2012:1061-1064
- [12] Cao J, Kwong S, Wang R. A noise-detection based AdaBoost algorithm for mislabeled data[J]. *Pattern Recognition*, 2012, 45(12):4451-4465
- [13] Lan R S, Jiang Y, Ding L G, et al. Automated flare prediction using the AdaBoost algorithm[J]. *Research in Astronomy and Astrophysics*, 2012, 12(9):1191-1196
- [14] Freeland S L, Handy B N. Data Analysis with the SolarSoft System[J]. *Solar Physics*, 1998, 182(2):497-500
- [15] Grady L. Multilabel random walker image segmentation using prior models[C]// *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Volume 1, San Diego, 2005:763-770

(上接第 27 页)

- [2] Armbrust M, Fox A, Griffith R, et al. A view of cloud computing [J]. *Communications of the ACM*, 2010, 53(4):50-58
- [3] Dean J, Ghemawat S. MapReduce: simplified data processing on large clusters[J]. *Communications of the ACM*, 2008, 51(1):107-113
- [4] White T. Hadoop: the definitive guide(2nd ed)[M]. O'Reilly, 2012
- [5] Talbot J, Yoo R M, Kozyrakis C. Phoenix++: modular MapReduce for shared-memory systems[C]// *Proceedings of the Second International Workshop on MapReduce and its Applications*. ACM, 2011:9-16
- [6] Ekanayake J, Li H, Zhang B, et al. Twister: a runtime for iterative mapreduce[C]// *Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing*. ACM, 2010:810-818
- [7] He B, Fang W, Luo Q, et al. Mars: a MapReduce framework on graphics processors[C]// *Proceedings of the 17th international conference on Parallel architectures and compilation techniques*. ACM, 2008:260-269
- [8] Zaharia M, Chowdhury M, Das T, et al. Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing[C]// *Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation*. USENIX Association, 2012:2

- [9] Kang L Y, Wang X Y, Bai R J. Analysis of MapReduce Principle and Its Main Implementation Platforms[J]. *New Technology of Library and Information Service*, 2012(02):60-67
- [10] Li J J, Li Q, Tian B. The Analysis and Comparison between MapReduce Implementations[J/OL]. <http://www.paper.edu.cn/html/releasepaper/2011/11/464>. 2011
- [11] Li J J, C J, W D, et al. Survey of MapReduce Parallel Programming Model[J]. *ACTA Electronica Sinica*, 2011(11):2635-2642
- [12] Borthakur D. The hadoop distributed file system: Architecture and design[OL]. http://hadoop.apache.org/docs/ro.18.1/hdfs_design.html
- [13] Thusoo A, Sarma J S, Jain N, et al. Hive: a warehousing solution over a map-reduce framework[J]. *Proceedings of the VLDB Endowment*, 2009, 2(2):1626-1629
- [14] Olston C, Reed B, Srivastava U, et al. Pig latin: a not-so-foreign language for data processing[C]// *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*. ACM, 2008:1099-1110
- [15] Abouzeid A, Bajda-Pawlikowski K, Abadi D, et al. HadoopDB: an architectural hybrid of MapReduce and DBMS technologies for analytical workloads[J]. *Proceedings of the VLDB Endowment*, 2009, 2(1):922-933
- [16] Anil R, Dunning T, Friedman E. Mahout in action[M]. Manning publications, 2011