

日冕暗化图像检测算法的并行设计与实现

杨宇航 彭博 李天瑞

(西南交通大学信息科学与技术学院, 成都, 611756)

摘要: 日冕物质抛射(Coronal mass ejection, CME)是空间灾害天气最主要的驱动源。在CME的众多伴生现象中,日冕暗化与之最为紧密相关。因此,对日冕暗化现象进行检测,有助于预报CME的爆发。随着观测数据量的增加,现有日冕暗化检测算法的效率亟待提升。本文基于一种半自动暗化检测算法,提出并实现该算法的并行方案,以提高算法效率,为实现日冕暗化的实时检测奠定基础。首先介绍了日冕暗化的现有工作,接着详细分析了一种半自动暗化检测方法,它在一定程度上提高了人工识别暗化区域的效率,但其效率并不能满足检测的实时性要求。然后,基于Matlab R2014a平台并行机制的特点,从数据、计算量和图像分块角度对半自动暗化检测算法提出了3种不同的并行方案,实验对比分析结果表明图像分块并行方案的效果最优。

关键词: 暗化检测;日冕物质抛射;图像处理;并行计算

中图分类号: TP391.41 **文献标志码:** A

Parallel Design and Implement of Coronal Dimming Detection Algorithm

Yang Yuhang, Peng Bo, Li Tianrui

(School of Information Science and Technology, Southwest Jiaotong University, Chengdu, 611756, China)

Abstract: Coronal mass ejection (CME) is considered to be the main driving force of space weather disasters. CME often appears with many other solar activities. Since coronal dimming is associated most closely with CME, the detection of coronal dimming can help to forecast CME. Continuous development of CME observations brings increasingly large amount of data, whereas the efficiency of the detection algorithms of coronal dimming needs to be improved. Three parallel algorithms are presented for a semi-automatic dimming detection algorithm. They make the foundation for real-time detection of coronal dimming. We firstly introduce the existing work about the topic, then analyze a semi-automatic dimming detection algorithm, which to some degree, improves the efficiency of artificial recognition of darkened area, but does not meet the requirements of real-time detection. Based on the principle of parallelism of Matlab R2014a, three different parallel algorithms are presented from different aspects, i. e. data, distance-calculation and image-divide. Experimental results show that the parallelization based on image-divide is the most efficient one among the three algorithms.

Key words: dimming detection; coronal mass ejection (CME); image processing; parallel computing

引言

日冕物质抛射(Coronal mass ejection, CME)^[1-3]是空间中场面最壮观、发生最频繁且爆发规模最大的太阳活动现象,是空间灾害天气(如太阳爆事件引起的卫星破坏、通讯中断和导航失灵等)最主要的驱动源。目前,直接预报 CME 还存在极大的难度。CME 有很多伴生现象,通过检测与 CME 爆发最为紧密相关的日冕暗化现象可以间接预测其发生。

近年来,关于暗化的研究图像主要分为两种:(1)基于差分图像的研究^[4-7],又分基准图差分(Base difference, BD)图像和时变差分(Running difference, RD)图像两种;(2)基于原始图像的研究^[5,8]。Podladchikova 等^[4]提出了新的 EIT 波观察(Novel EIT wave machine observing, NEMO)算法,基于 RD 图像的统计特征进行日冕暗化检测,并且基于 BD 图像用区域生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^[5]改进了 NEMO,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。Reinard 等^[7]基于 BD 图像采用阈值方法在规定的区域(Region of interest, ROI)内检测暗化像素,并计算了暗化尺寸。

随着日冕观测卫星的发展,从 1995 年的太阳和日球层探测器(Solar and heliospheric observatory, SOHO),2006 年的日地关系观测台(Solar terrestrial relations observatory, STEREO),一直到 2010 年的太阳动力学观测站(Solar dynamics observatory, SDO),观测数据的时间和空间分辨率越来越高,数据量越来越大,传统的暗化检测方式在效率方面已经不能满足检测的实时性要求。

本文基于一种半自动暗化检测算法^[9],提出并实现了该算法的并行化,以提升算法计算效率,更高效地进行日冕暗化的检测,为实现日冕暗化的实时检测奠定基础。实验结果表明,并行后的算法效率比原算法提升了近 58.20 倍。

1 半自动暗化检测算法

半自动暗化检测算法,给出可能发生暗化的若干个推荐区域,再由专家进行判定。该算法的作用在于缩小暗化发生区域的选择范围,减少了人为识别暗化现象的工作量,解决完全靠人工从卫星返回图像数据中识别暗化区域的效率低、准确率低的问题。文献[9]中检测算法的主要流程为:

算法流程 1:半自动暗化检测算法。

输入:图像观测序列 S

输出:暗化区域序列 C

for S 中每一幅图像

 迭代计算暗像素点焦

 标记连通区域

 合并距离相近的不同区域

 将区域中面积最大的 4 个区域放入 C

 更新区域序列 C

end for

return C

在半自动暗化检测算法中,根据式(1)计算两两区域之间的距离 $D(R_i, R_j)$ 。如果 $D(R_i, R_j) < T_{Dis}$,则合并区域 R_i 和 R_j ,其中 T_{Dis} 是事先设定的阈值,表示要合并的两个区域最大可能距离。直到不存在两区域之间距离小于 T_{Dis} 时,融合完毕。

$$D(R_i, R_j) = \min(|P_x P_y|) \quad P_x \in R_i, P_y \in R_j \quad (1)$$

式中: R_i 和 R_j 为两个不同的区域, P_x 和 P_y 分别为属于 R_i, R_j 的两个像素点, $|P_x P_y|$ 为两个点之间的

距离。本文的并行工作主要针对算法流程 1 中第(4)步“合并距离相近的不同区域”进行。

2 暗化检测算法并行方案

Matlab 程序执行过程中,遇到并行标志时,将启动 Client(客户)端和 Worker(工作)端模式^[10,11]。其中,“Client”端负责编写和启动并行代码,并行代码的具体执行在“Worker”端进行。不同“Worker”间并行工作,每个“Worker”内部串行工作。

经过对算法特点的分析及各部分耗时的测试,发现算法中“合并距离相近的不同区域”部分是整个过程中最耗时的部分,且耗时远远大于其他部分。结合 Matlab 并行机制,提出了 3 种并行方案,分别是数据并行、计算并行和图像分块并行。

2.1 数据并行

暗化检测算法的测试数据是一个由多幅图像组成的图像序列。原算法中按时间序列依次对每幅图像进行处理,最后获得暗化区域。

数据并行方案的主要思想是:将图像序列分为 n 组,尽量保证每组图片数量相同。每组交由不同的 Worker 进行处理,最后在 Client 端将结果汇总,以此实现多处理器同时处理多幅图像的并行目的。数据并行方案的算法过程为:

算法流程 2:数据并行检测暗化区域。

输入:图像检测序列 S 。

输出:暗化区域序列 C 。

将 S 分为 n 组,每组交给一个“Worker”

“Worker”端,for 每一幅图像

迭代计算暗像素点集

标记连通区域

合并距离相近的不同区域

将区域中面积最大的 4 个区域放入 C

更新区域序列 C

return C

end for

从“Worker”端返回“Client”端。

2.2 计算并行

计算并行针对“合并距离相近的不同区域”部分进行。原算法中计算不同区域间的距离时,对每一区域内的所有像素点都进行了计算。实际上,只要计算出有小于阈值的距离,就可以停止计算。基于此想法提出的并行思想是,并行计算一个区域和其余区域的距离。

2.3 图像分块并行

图像分块并行的思想是,对每一幅图像,在标记出所有暗像素点集后,将其横向分割为等宽的 n (n 满足能被 1 024 整除)块,交由不同的“Worker”进行“合并临近像素点集”处理。然后再将各块拼接起来,对边缘区域进行处理。图像分块并行方案的算法过程为:

算法流程 3:并行计算一幅图像中区域间距离。

输入:一幅图像中的连通区域 A ,共 n 个区域。

输出:一幅图像中的暗化区域队列 C 。

for A 中 $n-1$ 个区域中的每个区域 i

将剩下的 $n-1$ 个区域分组, 每组交给一个“Worker”

“worker”端, for 每一个区域 j

计算区域 i 和 j 的距离 D_{ij}

如果 D_{ij} 小于阈值, 合并区域 i, j

end for

end for

合并后的区域放入区域队列 C

return C

算法流程 4: 图像分块并行计算图中暗化区域。

输入: 图像序列 S 。

输出: 每一幅图像中的暗化区域 C 。

for 每一幅图像

迭代计算暗像素点集

标记连通区域

将图像横向分为 n 块

n 块分为 m 组, 每组交由一个“Worker”

“Worker”端, for 每一块

合并距离相近的不同区域

end for

从“Worker”返回“Client”, 合并图像

for 每一个拼接边缘

计算边界区域内不同点集间的距离

合并距离相近的不同点集

end for

将合并后的区域放入暗化区域队列 C

return C

end for

3 并行方案对比实验

本文中提出的 3 种并行方案, 对暗化区域检测的准确率与原算法一致。为验证效率提升效果, 在实验部分设置了 6 个实验, 分别对 3 种并行方案进行测试和对比。

进行对比实验的硬件和软件环境如

表 1 所示。

3.1 实验数据

用于实验的数据是由太阳动力学天文台 (Solar dynamics observatory, SDO) 上搭载的大气成像组件 (Atmospheric imaging assembly, AIA) 拍摄的时间与空间高分辨率的完整太阳盘面的影

像。实验中用到的数据是 2010 年 7 月 5 日发生的一次日冕暗化现象, 在 2:40~4:00 这个时间段内的

表 1 硬件和软件环境

Tab. 1 Hardware and software

硬件环境	CPU	AMD A8-4500M APU with Radeon (tm) HD Graphics 1.90 GHz4 核
	RAM	4.00 GB, 可用 3.45 GB, 空闲 1.82 GB
软件环境	OS	Windows 7 64 位旗舰版
	实验平台	Matlab R2014a

图像序列,共有 301 幅图像。实验时,根据需要从中选取不同图像组成序列进行相关分析。

3.2 实验结果分析

图像分块并行暗化检测算法的效率与图像分成的块数相关。在启动 4 核处理器的情况下,分别测试块数为 2,4,8,16,32,64 时的效率。当一幅图像分为 32 块的时候,算法效率最佳。

选取 1,2,3,4,5,10 幅图像,分别测试原半自动暗化检测算法和数据并行方案在 CPU 核数为 1,2,4 核情况下的效率;计算并行方案和图像分块并行方案(分为 32 块)在 CPU 核数为 2,4 核情况下的效率。

表 2 所示为原半自动暗化检测算法、数据并行暗化检测算法、计算并行暗化检测算法和图像分块并行暗化检测算法的耗时情况。从表 2 看出,随着图像幅数的增多,算法消耗的时间也越多;CPU 核数的改变对串行算法的耗时情况几乎没有影响;在数据并行、图像分块并行方案中,CPU 核数越多,算法执行效率越高。

表 2 3 种并行方案在不同 CPU 核数下的耗时

Tab.2 Time consuming of three parallel programs with different CPU core numbers

算法	CPU 核数	图像幅数					
		1	2	3	4	5	10
半自动暗化检测算法	1	101.39	303.60	455.50	637.12	814.39	1618.30
	2	101.39	303.77	455.99	637.10	816.25	1621.79
	4	102.51	305.35	456.21	637.56	818.00	1626.99
数据并行暗化检测算法	1	154.73	287.95	323.23	423.07	580.15	1065.83
	2	134.88	237.73	409.82	484.07	701.68	1285.42
	4	134.03	237.34	301.22	367.61	550.07	982.30
计算并行暗化检测算法	2	54.10	147.96	221.21	306.42	386.18	776.62
	4	68.47	170.05	254.31	352.27	443.19	896.92
图像分块并行暗化检测算法	2	2.58	4.80	6.79	8.67	10.69	21.42
	4	2.57	4.59	6.57	8.54	10.81	20.81

从表 2 中选出各个方案在最佳 CPU 核数(表中粗体标出),分别是原算法单核、数据并行 4 核、计算并行 2 核和图像分块 4 核情况下的耗时,绘制出如图 1 所示时间曲线,进行对比分析。从图 1 中的 4 条曲线可以看出,3 种并行方案对原算法的效率都有所提升,其中,图像分块并行的加速效果大大优于数据并行和计算并行。

最后,进行计算加速比的实验。从描述的数据中,依次取时间间隔为 6 min 的图像组成包含 14 幅图像的图像序列进行测试。4 种算法耗时情况及相对于串行算法的加速比见表 3。从表 3 中可以看出,图像分块并行暗化检测算法的加速效果最佳。

表 3 3 种并行方案加速比

Tab.3 Speed up of three parallel programs

参数	串行 1 核	数据并行 4 核	计算并行 2 核	图像分块 4 核
耗时/s	1 516.70	865.40	730.98	26.06
加速比		1.75	2.07	58.20

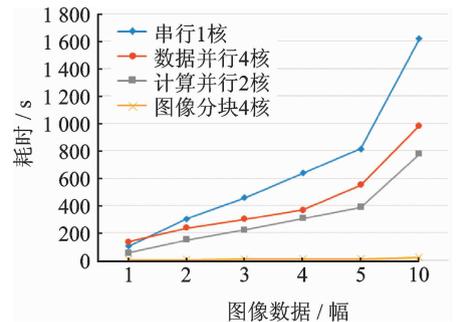


图 1 不同算法最优情况耗时曲线

Fig. 1 Time consuming curves of different algorithms under optimal circumstances

4 结束语

CME 是太阳系中最壮观、最频繁的大尺度活动现象,是空间灾害天气最主要的驱动源。通过检测 CME 的伴生现象日冕暗化现象的发生可对其进行预报。随着日冕观测卫星的发展,观测数据的时间和空间分辨率越来越高,数据量越来越大,传统的暗化检测算法的效率亟待提升。本文为提高现有暗化检测算法效率,基于一种半自动暗化检测算法提出了 3 种并行方案:数据并行、计算并行和图像分块并行。其中,基于图像分块的并行方案对原算法的加速效果最明显,加速比高于 58%,为实现暗化现象的实时性检测奠定一定基础。下一步工作将针对大规模数据集,借助更好的并行平台,进一步提升暗化检测算法的效率,以满足实时性的要求。

参考文献:

- [1] 周桂萍,曹卓良,汪景琇. 日冕物质抛射和太阳表面磁活动的关系[J]. 天文学进展, 2003, 21(1):41-54.
Zhou Guiping, Cao Zhuoliang, Wang Jingxiu. Relationship between coronal mass ejections and solar surface magnetic activity [J]. *Progress in Astronomy*, 2003, 21(1):41-54.
- [2] 申远灯. 日冕物质抛射及相关太阳活动的观测研究[D]. 云南:中国科学院国家天文台云南天文台, 2009:38-52.
Shen Yuandeng. Observational studies and associated coronal mass ejections of solar activity [D]. Yunnan: National Astronomical Observatories, Chinese Academy of Yunnan Observatory, 2009:38-52.
- [3] Zhang Y, Wang J, Attrill G D R, et al. Coronal magnetic connectivity and EUV dimmings[J]. *Solar Physics*, 2007, 241(2):329-349.
- [4] Podladchikova O, Berghmans D. Automated detection of eit waves and dimmings [J]. *Solar Physics*, 2005, 228(1/2):265-284.
- [5] 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.
- [6] 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.
- [7] Reinard A A, Biesecker D A. Coronal mass ejection-associated coronal dimmings[J]. *The Astrophysical Journal*, 2008, 674(1): 576.
- [8] 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.
- [9] 田红梅. 日冕图像中暗化现象的检测与提取技术研究[D]. 成都:西南交通大学,2015:27-37.
Tian Hongmei. Detection and extraction of coronal dimming phenomenon of coronal images [D]. Chengdu: Southwest Jiaotong University, 2015:27-37.
- [10] Duin R P W. PRTools-Version 3.0-A matlab toolbox for pattern recognition[J]. *Proc of Spie*, 2000, 46:1331.
- [11] Choy R, Edelman A. Parallel MATLAB: Doing it right[J]. *Proceedings of the IEEE*, 2005, 93(2): 331-341.

作者简介:



杨宇航(1993-),女,硕士研究生,研究方向:计算机视觉、数据挖掘, E-mail: yuhang_yang@foxmail.com.



彭博(1980-),女,博士,副教授,研究方向:计算机视觉、离散最优化算法和模式识别。



李天瑞(1969-),男,教授,博士生导师,研究方向:大数据、云计算、数据挖掘、粒计算和粗糙集。

