

天气雷达资料实时并行处理方法

王志斌¹, 陈波¹, 万玉发¹, 吴涛², 罗兵³, 沃伟峰³

(1. 中国气象局武汉暴雨研究所, 武汉 430074; 2. 武汉中心气象台, 武汉 430074; 3. 国家气象中心, 北京 100081)

摘要: 利用共享存储多处理器的集群环境, 研究高频度实时多部天气雷达资料处理的并行计算方法, 根据单部天气雷达的计算特点和多部雷达混合处理的方法, 提出一种粗粒度消息传递接口分布式内存和细粒度 OpenMP 共享内存混合编程的 2 级并行方法。实验结果表明, 该方法使系统资料处理速度得到较大提高。

关键词: 天气雷达; 资料处理; 加速比

Weather Radar Data Real-time Parallel Processing Method

WANG Zhi-bin¹, CHEN Bo¹, WAN Yu-fa¹, WU Tao², LUO Bing³, WO Wei-feng³

(1. Wuhan Institute of Heavy Rain, China Meteorological Administration, Wuhan 430074;

2. Wuhan Central Meteorological Observatory, Wuhan 430074; 3. National Meteorological Center, Beijing 100081)

【Abstract】 Using cluster environment of shared memory multiprocessors, this paper studies parallel computing method processed by several high frequency real-time weather radars data. It presents two-stage parallel method which include a coarse grained message passing interface distributed memory and fine grained OpenMP shared memory hybrid programming, according to the single weather radar calculation characteristics and several radars mix processing method. Experimental results show that this method makes system radar data processing speed greatly enhanced.

【Key words】 weather radar; data processing; speed-up ratio

1 概述

随着我国现代化多谱勒天气雷达在全国广泛的建设, 以及多谱勒天气雷达在监测暴雨、大风、雷暴等强对流天气的巨大作用, 它的应用越来越广泛, 我国即将建成仅次于美国的第二大谱勒天气雷达监测网^[1]。因此, 为了提高对中尺度灾害性天气的研究以及预警能力, 以发挥多部雷达在联合监测天气中的作用, 就必须把来自多部雷达的资料进行组网拼图。由多个雷达重复取样获得的天气信息要比只由单个雷达取样获得的更加精确, 还可以很大程度解决因单部雷达观测的波束几何学原因(例如静锥区、波束展宽、波束高度、波束阻挡等)引起的很多问题。把单部雷达联合起来进行使用要解决的首要问题是资料实时处理的时效性应得到满足。单部多谱勒天气雷达资料处理的时间间隔一般情况下以 6 min 为 1 个周期, 因此, 某一区域要同时使用多部雷达资料必须在 6 min 内完成所有收集到的雷达资料。如果按照一部的顺序处理多部雷达资料, 则在现有单个 CPU 计算机的环境下处理的极限也就十几部雷达。处理 10 部以上需要丢弃资料, 或放宽处理周期时间, 这样对生命史非常短的强对流天气系统是可行的。并行处理多部雷达资料是一个有效的方法。并行计算的应用已遍布天气预报^[2-3]、石油勘探、航空航天、核能利用、生物工程等领域, 理论研究与应用普及均取得很大进展。但是在上述领域, 并行计算主要解决的是大规模计算问题, 对于多部天气雷达资料处理而言, 并不像大规模计算问题存在明显的集中耗时, 其计算粒度比较小, 由于要不断地进行单部雷达循环处理, 因此造成处理时间过长。在天气雷达资料应用中应用集群计算技术仍比较少, 国内还没有相关的研究, 美国强风暴实验室(NSSL)是美国处理天气雷达资料的中心^[4], 它的处理系统是由若干台 SUN 服务器组成,

分块处理雷达资料后进行拼接处理, 没有基于并行化方法。一些欧洲国家, 由于地域小, 一个国家就几部雷达, 因此没有必要进行并行化处理, 但对于国内, 这是非常有意义的研究课题。随着计算机技术的发展, 多核 CPU 技术的日益成熟, 利用普通的硬件就可以搭建小型集群系统, 使得天气雷达这样的小型计算问题可以利用集群计算技术加以解决。

本文利用天气雷达资料处理算法特性, 挖掘并行性, 并通过消息传递接口 MPI 和 OpenMP 的并行计算方法, 实现分布式内存和共享内存两级并行资料处理^[5], 从而有效地提高天气雷达处理的运行效率。

2 天气雷达资料处理流程

天气雷达资料处理流程主要包含: 原始资料收集, 资料名称的归整, 资料的质量控制, 坐标转换, 拼图处理和基于拼图产品的生成。

2.1 原始资料收集

每一个区域(或省)有多部雷达, 雷达站通过专线每 6 min 实时传输经过压缩的数据到区域(或省)的 FTP 服务器上。

2.2 资料名称规范化

在每一个 6 min 周期内, 每分钟都可能有不同的雷达资料到达区域的 FTP 服务器上, 此类资料的命名以分钟为单位, 将此类文件命名统一到 6 min, 即文件名以 6 min 间隔为单位进行命名。

基金项目: 国家“863”计划基金资助项目(2006AA01A123); 国家自然科学基金资助项目(40675070)

作者简介: 王志斌(1965-), 男, 副研究员, 主研方向: 计算机图像处理, 数据库开发; 陈波, 硕士研究生; 万玉发, 研究员; 吴涛, 高级工程师; 罗兵, 研究员; 沃伟峰, 工程师

收稿日期: 2009-06-17 **E-mail:** cemp@sohu.com

2.3 质量控制

天气雷达获取的并非都是降水回波,还包括非气象回波,主要包括:地物杂波,晴空回波以及异常传播(Anomalous Propagation, AP)下的超折射回波,沿海地区可能会探测到海浪回波。虽然一些非气象回波有一定的预报指示意义(如 AP 回波),但对于降水估算、自动临近预报及洪水警报而言,可能会引起较大的估算误差和错误,必须使用程序对其自动实时识别和剔除。

质量控制过程是对反射率场进行弱回波消除、噪声过滤等预处理后,提取出回波的三维结构特征量,运用模糊逻辑算法识别杂波(地物杂波、AP 等),并进行剔除。

2.4 坐标转换

坐标转换主要目的是把雷达资料按极坐标储存方式转换成按经纬网格坐标的存储方式,可以提供 GIS 系统显示和其他系统使用。大量的三角函数变换通过查表方式(即极坐标对应的经纬度表)减少计算量,每个雷达站只要在初始时计算一次,其他时候查表。

2.5 拼图处理和基于拼图的产品生成

通过质量控制后的单部雷达形成一个范围较小的经纬度格点场,将所有在区域范围内的单部雷达资料进行拼接形成大范围的经纬度格点场。并在此数据基础上进行产品的生成,再提供给用户。

上述处理过程大都可以用并行化方式进行,而且特别适合 Join-Fork 模式进行,雷达资料并行算法流程如图 1 所示。雷达资料并行处理系统可以在多台具有多核或多 CPU 的计算机上共同运行,并行算法主要分为:(1)基于分布式内存的多进程,并行运行在多机上;(2)基于共享内存的多线程,并行运行在多核或多 CPU 单机上^[6]。两者配合使用可以使雷达资料处理速度提高到最大化。

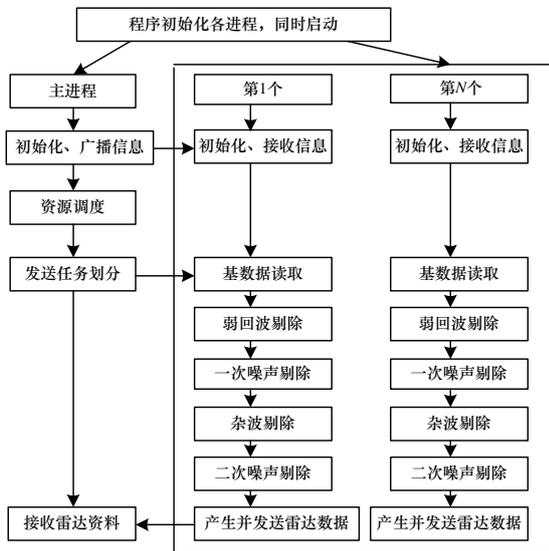


图 1 雷达资料并行算法流程

3 并行算法分析

3.1 顺序处理时间

当收到某一雷达资料后,按照顺序处理资料的方式,系统即可处理雷达资料,雷达资料要进行实时监控,如果在处理的过程中收到新的雷达资料,则系统还是按照同样的方式进行处理。该处理方式可以提高系统对雷达资料的处理速度,但对雷达资料的规整和后续的基于拼图处理会有很多不利影

响。最好的处理方法是在一个 6 min 周期内同时处理完成所有收集到的雷达资料,形成一个大面积的拼图数据。典型的工作示意图如图 2 所示。

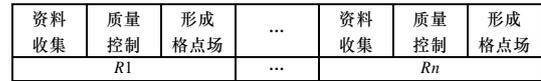


图 2 雷达资料处理工作示意图

由图 2 可知,一个周期的雷达处理时间是可以进行估算的。设有 N 部雷达每个雷达处理时间为 t_i ,这样处理一个周期的所有雷达时间总和为

$$t = tc + \sum_{i=1}^n ti + tm = tc + \sum_{i=1}^n (tr_i + te_i + tn_i + ta0_i + ta1_i + tg_i) + tm$$

其中, t 为一次处理的总运行时间; tc 为雷达资料处理时初始化运行时间; tm 为所有雷达处理完成后进行雷达资料拼图时所用的时间; te_i 为雷达弱回波消除时间; tn_i 为杂波消除处理时间; $ta0_i$ 和 $ta1_i$ 为噪声过滤时间; tr_i 为资料读取时间。

3.2 分布式内存并行算法

雷达资料处理过程就是资料收集规范、质量控制、形成格点场的循环过程,其中主要时间用在资料质量控制上,虽然在单部雷达的计算上时间是很短的,但是多部雷达资料进行处理就会消耗大量时间,而且要在 6 min 内完成时间就更宝贵了,缩短处理时间的理想的方法是对单部雷达质量控制过程进行全部的并行化处理,但是质量控制过程中有些是不能完全并行的,如弱回波消除和杂波消除处理是一个顺序的过程,他们之间存在因果关系。但对于单部雷达而言,多次噪声过滤是可以进行并行化处理的,即各次的噪声过滤处理可以独立分配到单个节点上运行,这些是粒度比较小的处理。另外对于单部雷达而言,它们之间的处理是互相独立的,这 2 个方面可以采用分布式内存结构的 MPI 主从结构模式进行编程^[7],把每一个雷达的资料处理分配到多台计算机上运行,每个雷达处理完成后,由从计算机系统上的结果数据传回到主计算机上进行处理形成大的拼图场。因此,在一个周期内运行的时间并行化后如下(系统申请计算机系统的资源条件得到全部满足):

$$t = tc + \sum_{i=1}^n t_i + tm = tc + \sum_{i=1}^n (tr_i + te_i + tn_i + tg_i + (ta0_i + ta1_i) / m) / n + tm$$

其中, m 为噪声过滤的次数,一般为 2,在该并行计算中,多机并行计算机处理的都是相似资料,实现负载均衡比较容易,多机之间的开销和数据交换量不是很大,执行的效率比较高。但单部雷达的处理过程如弱回波消除和杂波消除处理等不能进行并行化处理,限制了处理速度的进一步提高,因此,要对此类算法进行改进。

3.3 共享内存并行算法

在雷达资料处理过程中质量控制模块的计算是最耗时的,对于多部雷达可以采用共享内存的 OpenMP 进行并行计算,实现 MPI 和 OpenMP 的混合编程,OpenMP 是在多处理器上的共享内存并行开发标准,采用多线程的方法充分利用多个处理器实现效率的提高。由于 OpenMP 中生成多线程和最后线程的等待及产品结果的计算等都是并行计算产生的额外计算量,一次雷达处理模块的运行时间相对较少,因此必须保证额外计算量消耗的时间要比并行计算减少的时间少,这样 OpenMP 并行计算才有效。为了达到该目的,应尽量减少多线程的生成次数。

3.4 硬件系统的选择

由上文分析可以看出,分布式内存并行算法和共享内存并行算法都可以提高雷达资料的处理速度。当硬件为集群系统且每台计算机只有一个CPU时,适合单独采用分布式内存并行算法,系统中可并行的计算机最大数目为本地收到雷达的数目,此时使用该并行算法所能提高的速度达到极限值。当集群中的每台计算机有多个CPU构成共享内存的集群系统时,即SMP集群系统,适合应用共享内存和分布式内存结合的并行算法,多个计算机之间用分布式内存并行算法,1个计算机的多CPU之间用共享内存并行算法,这样可以最大限度地提高雷达处理速度。对于气象部门,选择一台多核的共享内存计算机系统是比较现实的,它的开销小且容易维护。

4 实验结果及分析

在雷达资料处理实验中,设计雷达数共有12个,所使用的处理平台为4台4核CPU的计算机组成的集群系统(共16个CPU),操作系统采用Linux,设计及实现并程序的目的是使并行代码比相应的串行代码运行快,为了评估并行计算对雷达资料处理效率的影响,通常用加速比作为衡量标准,加速比为串行程序执行时间/并行程序执行时间,在实验中,测试2种情况:(1)单纯MPI并行,每个CPU1个MPI进程,最多为16个MPI进程;(2)MPI和OpenMP混合并行,每台计算机运行1个MPI进程和4个OpenMP线程。结果如表1所示,串行处理时间为259.91s。其中,MPI+OpenMP混合并行中,每个MPI进程包含4个OpenMP线程。

表1 集群计算机并行计算雷达资料处理时间 s

| CPU 个数 | 时间 | |
|--------|--------|-----------------|
| | MPI 并行 | MPI+OpenMP 混合并行 |
| 2 | 177.11 | - |
| 3 | 152.61 | - |
| 4 | 137.41 | 189.31 |
| 5 | 130.83 | - |
| 6 | 123.91 | - |
| 7 | 121.43 | - |
| 8 | 120.71 | 103.13 |
| 9 | 119.78 | - |
| 10 | 117.77 | - |
| 11 | 116.79 | - |
| 12 | 115.61 | 98.57 |
| 13 | 115.61 | - |
| 14 | 115.61 | - |
| 15 | 115.61 | - |
| 16 | 115.61 | 78.56 |

从表1可以看出,无论采用哪种并行算法,都可以提高雷达资料处理速度,但采用MPI并行方法,当进程数超过收集到的雷达个数时,处理速度便不再提高。此时,使用MPI和OpenMP的混合并行,处理时间可进一步减少。由图3可以看出,在处理器数目比较少时,单纯MPI的并行加速比大

于MPI+OpenMP混合并行加速比,但随着处理器数目逐渐增大,两者加速比逐渐接近,加速比较低,程序的并行算法有待进一步完善。

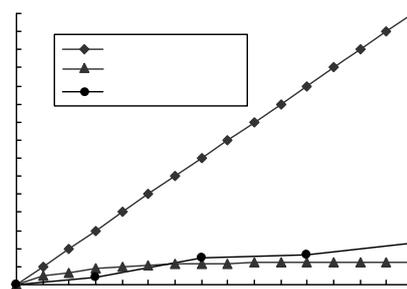


图3 CPU个数与加速比的关系

5 结束语

通过对天气雷达系统特点分析,本文研究了在SMP集群环境下,利用MPI和OpenMP对雷达系统进行两级并行数据处理的方法,为今后以区域为中心进行多部雷达拼图提出了很好的解决办法。特别是为在某一特定时间、周期内一定要完成任务的处理系统提出了解决问题的思路和方法。

参考文献

- [1] 杨洪平,张沛源,程明虎,等.多普勒天气雷达短时预报技术研究进展暴雨灾害[J].暴雨灾害,2007,25(2):185-189.
- [2] 伍湘君,金之雁,黄丽萍,等.GRAPES模式软件框架与实现[J].应用气象学报,2005,16(8):539-546.
- [3] 臧增亮,饶宜锐,潘晓滨,等.区域分解对气象模式并行计算速度的影响[J].计算机工程,2008,34(17):4-6.
- [4] Istok M J, Okulski P R, Saffle R E, et al. NWS Use of FAA Radar Data-progress and Plans[C]//Proc. of the 21st International Conference on Interactive Information Processing Systems for Meteorology, Oceanography and Hydrology. San Diego, USA: American Meteorological Society, 2007.
- [5] 王惠春,朱定局,曹学年,等.基于SMP集群的混合并行编程模型研究[J].计算机工程,2009,35(3):271-273.
- [6] 单莹,吴建平,王正华.基于SMP集群的多层次并行编程模型与并行优化技术[J].计算机应用研究,2006,23(10):254-258.
- [7] 蔡佳佳,李名世,郑锋.多核微机基于OpenMP的并行计算[J].计算机技术与发展,2007,17(10):87-91.

编辑 陆燕菲

(上接第254页)

参考文献

- [1] Ballard D H. Generalizing the Hough Transform to Detect Arbitrary Shapes[J]. Pattern Recognition, 1981, 13(2): 111-122.
- [2] Ma De, Chen Xing. Hough Transform Using Slope and Curvature as Local Properties to Detect Arbitrary 2D Shapes[C]//Proc. of the 9th Int'l Conf. on Pattern Recognition. Beijing, China: [s. n.], 1988: 511-513.

- [3] Ser P K, Siu W C. Non-analytic Object Recognition Using the Hough Transform with the Matching Technique[J]. Computers and Digital Techniques. 1994, 14(1): 11-16.
- [4] Thomas A D H. Compressing the Parameter Space of the Generalized Hough Transform[J]. Pattern Recognition Letters, 1993, 13(2): 107-112.

编辑 任吉慧