

文章编号:1001-9081(2009)06-1665-04

## 基于并行计算的气溶胶定量遥感反演模型实现

郭建平<sup>1</sup>, 肖华东<sup>2</sup>, 刘昭华<sup>3</sup>, 曹春香<sup>4</sup>, 张 颢<sup>4</sup>, 光 洁<sup>4</sup>

(1. 中国气象科学研究院 大气成分观测与服务中心, 北京 100081; 2. 国家气象信息中心 计算机室, 北京 100081;

3. 江西理工大学 建筑与测绘工程学院, 江西 赣州 341000; 4. 中国科学院 遥感应用研究所, 北京 100101)

(xiaohd@cma.gov.cn; jgguo@cma.gov.cn)

**摘 要:**为了加快气溶胶光学厚度(AOD)反演计算速度,基于SYNTAM串行算法,提出了循环分块划分和聚合通信的策略,利用消息传递模型,在中国气象局的IBM Cluster 1600高性能计算机系统上,并行实现了从MODIS双星(TERRA和AQUA)卫星数据反演AOD。试验结果表明该方法大大减少了计算时间,与地面太阳光度计实测AOD数据进行对比验证,发现所有站点处的AOD反演相对误差小于22%,表明这种并行方法可以满足高精度监测空气质量要求。

**关键词:**气溶胶光学厚度; 并行计算; 循环分块; 消息传递接口

**中图分类号:** TP722.4 **文献标志码:** A

## Parallel implementation of aerosol optical depth retrieval algorithm

GUO Jian-ping<sup>1</sup>, XIAO Hua-dong<sup>2</sup>, LIU Zhao-hua<sup>3</sup>, CAO Chun-xiang<sup>4</sup>, ZHANG Hao<sup>4</sup>, GUANG Jie<sup>4</sup>

(1. Center for Atmosphere Watch and Service, Chinese Academy of Meteorological Sciences, Beijing 10081, China;

2. National Meteorological Information Center, Beijing 10081, China;

3. School of Construction and Surveying Engineering, Jiangxi University of Science and Technology, Ganzhou Jiangxi 341000, China;

4. Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing 100101, China)

**Abstract:** To speed up the computation of the retrieval of Aerosol Optical Depth (AOD), a cyclic partition and collective communication strategy based on serial computation algorithm of Synergy of TERRA/MODIS and AQUA/MODIS (SYNTAM) was proposed to archive parallel computation of AOD retrieval from the both Moderate Resolution Imaging Spectroradiometer (MODIS) satellite (TERRA, AQUA) data by message passing on the IBM cluster 1600 high performance compute of Chinese Meteorological Administration. Experimental results demonstrate that the parallel SYNTAM can decrease computation time greatly. Compared with measurements from ground-based sun-photometers, the relative error of experiment results from the parallel SYNTAM is less than 22% in call cases, indicating that it can meet the needs of high accuracy such as air quality monitoring.

**Key words:** Aerosol Optical Depth (AOD); parallel computation; cyclic partitioning; Message Passing Interface (MPI)

## 0 引言

气溶胶光学厚度(Aerosol Optical Depth, AOD)是一个重要的地球物理参数,在很大程度上可以指示空气质量。由于具有全球覆盖和固定的重访周期的特性,卫星遥感可为AOD的遥感反演提供了一个可靠支撑。

在反演陆地下垫面气溶胶的过程中,被动式卫星遥感数据的使用在很大程度上依赖于精确的地面反射率先验知识的获取。NASA(美国国家航天局)AOD业务算法所采用的DDV算法即从MODIS卫星图像上反演10 km分辨率的AOD产品,该方法具有较高精度<sup>[1]</sup>。但是,DDV往往要求下垫面具有较低的反射率,如水体、植被等均符合该要求。而沙漠、海岸带、干旱地区和城市等人为源气溶胶源区,该方法就不能很好地发挥作用了,可以说DDV算法对高反射率下垫面或雪覆盖地区无能为力<sup>[2]</sup>。现在通常利用文献[3]提出的DEEP BLUE算法来反演亮目标区比如城市、沙漠的气溶胶光学属

性,作为DDV算法的有力补充。

与单角度卫星传感器不同,多角度成像辐射计大大增强了我们从遥感数字信号中分离地气相应贡献的能力,这将极大地增加AOD反演精度。自20世纪90年代开始,人们就试图利用这种辐射计来反演AOD,文献[4]对多角度卫星遥感反演AOD算法做了大量尝试。后来,顺航迹辐射计(ATSR)的双角度遥感影像数据亦被用来成功反演AOD<sup>[5]</sup>,作为ATSR的延续,ATSR-2以及先进沿轨扫描辐射计(AATSR)都获得了成功应用。一些新型的多角度传感器,比如:多角度成像分光辐射计(MISR)<sup>[6]</sup>和地球反射率偏振和方向观测仪(POLDER)传感器也得到了广泛地应用<sup>[7]</sup>。

上述算法大多依赖于先验知识建立的查找表(Look-Up Table, LUT)来加速AOD的数值求解。为了说明AOD中不确定部分,LUT通常假设几种主要大气成分如水汽或者臭氧在给定时间内不变。LUT通常利用6S(Second Simulation of the Satellite Signal in the Solar Spectrum)或MODTRAN

收稿日期:2008-12-25;修回日期:2009-03-03。

基金项目:国家973计划项目(2007CB714407;07S00502CX);中央公益性基本科研业务专项项目(2007Y001)。

**作者简介:**郭建平(1978-),男,江西吉安人,助理研究员,博士,主要研究方向:气溶胶定量反演研究、并行计算;肖华东(1980-),男,湖北黄冈人,工程师,主要研究方向:并行计算;刘昭华(1977-),男,山西运城人,讲师,硕士,主要研究方向:定量遥感、高性能地学计算;曹春香(1964-),女,内蒙古呼和浩特人,研究员,博士,主要研究方向:环境遥感;张颢(1974-),男,陕西西安人,博士,主要研究方向:定量遥感;光洁(1982-),女,安徽安庆人,博士研究生,主要研究方向:大气遥感。

(MODerate resolution atmospheric TRANsmittance and radiance) 等辐射传输模型建立, 这往往涉及到上几何量级的数值计算, 这种计算量在普通的 PC 机上是无法及时完成的。这对于空气质量监测(主要通过卫星遥感 AOD)的实时性要求提出了极大的挑战。幸运的是, 在众多的性能计算方法中, 并行计算是一种比较有前途的方法, 它具有大大减少计算时间的潜力。

并行计算通常被认为是计算的高端, 主要应用于复杂系统和巨大挑战问题的数值模拟, 比如: 气候建模, 空气质量监测与建模等。

在遥感领域科学家们已经开发了众多的并行算法, 并得到了成功而广泛的应用。比如: 众多的高光谱图像并行处理算法, 其中涉及到大量的空间和光谱信息的计算<sup>[8-11]</sup>。尽管遥感应用中的大气校正日益得到关注, 但讨论其并行算法的实现的文献却很少。比如: 文献[12]开发的一种利用 TM (Thematic Mapper) 遥感数据估算 AOD 的快速并行算法。

PC 版本 SYNTAM 串行算法的一个主要不足就是计算时间长, 如利用一台配置了 2.0 GHz CPU 和 512 MB 内存的 PC 机从一幅 1 km 分辨率的 5 000 × 5 000 图像上反演 AOD 需要几天的时间。考虑到并行计算的效率和快速处理能力, AOD 的并行反演算法可能成为气溶胶遥感监测的一种通用技术。在文献[13]中已经初步讨论了通过分块技术实现 SYNTAM<sup>[14]</sup>的并行改造。本文基于 PC 版 SYNTAM 串行算法, 将提出一种新的 AOD 并行反演算法, 它利用循环分块划分和聚合通信技术, 并在中国气象局的 IBM System 1600 Cluster 集群系统上得到了成功应用。

## 1 SYNTAM 算法

卫星传感器所获得的辐射经常受到气体分子、大气气溶胶颗粒物的影响。这也使得确定气溶胶对大气辐射总量的贡献和利用遥感方法确定大气中的气溶胶含量皆成可能。但是利用卫星遥感影像反演 AOD 的一个关键问题是区分气溶胶散射与地表辐射对总辐射量的贡献, 目前已有多种数学物理方法来对二者进行区分<sup>[15]</sup>, SYNTAM 算法即是一个典型代表。

在 SYNTAM 算法中, 仅考虑大气中的短波辐射传输。文献[16]提出的算法主要解决了非偏振的吸收、散射和辐射传输, 而文献[17]为了从 AVHRR 数据中获取表面反射率, 进一步简化了该辐射传输方程。新型传感器如 MODIS 具有多个多可见光波段观测能力和较短的重访周期特性, 使得实时反演气溶胶光学厚度成为可能。

在 TERRA 和 AQUA 卫星过境的约 3 个小时时间间隔内, 基于地表的双向反射率和气溶胶类型不变两点假设, 得到:

$$A_{j,\lambda_i} = \frac{(A_{j,\lambda_i}'b - a_j) + a_j(1 - A_{j,\lambda_i}')e^{(n_j-b)e^{(0.00879\lambda_i^{-4.09} + \beta_j\lambda_i^{-\alpha})\sec\theta_j'}}}{(A_{j,\lambda_i}'b - a_j) + b(1 - A_{j,\lambda_i}')e^{(a_j-b)e^{(0.00879\lambda_i^{-4.09} + \beta_j\lambda_i^{-\alpha})\sec\theta_j'}}} \quad (1)$$

式中各符号意义参见文献[14]。再利用 Broyden 法<sup>[18]</sup>进行数值拟合, 可以求解式(1)中相应的波长指数和浊度系数, 进而可算出 MODIS 470, 550 和 660 nm 等三个可见光波段处的 AOD。

基于 SYNTAM 串行算法和 1 km 分辨率的 MODIS 图像反演一次覆盖中国内地区域的 AOD, 首先需准备大约 2 GB 的

14 个预先计算的波段辐射文件, 且中间计算(包括辐射和几何校正)过程又产生大量的数据。因此, PC 机上反演 AOD 需要很长的时间, 只有通过并行方法来加速 AOD 反演, 才能使得该算法具有现实意义。在本文下一部分将讨论 SYNTAM 算法的并行化实现。

## 2 并行算法

SYNTAM 算法涉及到大量数据的数值计算, 计算密集且耗时, 但计算过程具有像素级并行能力, 即每个像素的 AOD 反演过程可以独立执行。因此, SYNTAM 可以很好的实现并行化改造。

为了从 MODIS 图像中反演 AOD, 必须满足下列条件: 首先, 线性的 SYNTAM 算法能够在不同的处理器上执行, 即对处理器个数具有很大的伸缩性。第二, 为了有效地实现并行, 必须考虑负载均衡问题, 这主要通过对 MODIS 图像的等大小分块来实现。最后, AOD 的反演过程是完全相互独立的, 即不同处理器之间没有通信发生(除数据广播外)。如果上述条件满足, 就可以在采用消息传递接口在单指令流多数据流模型的并行集群上实现 SYNTAM 算法的并行化。在这种结构中, 每个处理器上执行相同的 SYNTAM 代码, 但是处理不同的 MODIS 图像块。MPI 模型是一种最重要的并行编程工具, 具有处理器之间的消息传递功能。

考虑到在利用卫星影像反演 AOD 过程的系统内存限制, 该并行算法通过一种有效的数据划分和处理器间的任务映射, 实现最小化处理器间的通信和处理器的 I/O 通信。

通过空间域的循环分块可以有效解决内存限制, 并获得较好的负载均衡。如我们把  $n \times (1/P)$  的计算任务分配到每个计算节点, 其中  $n$  表示总的计算任务量,  $P$  表示计算节点个数。循环分块的目的是为了最小化处理器间的通信, 最大化负载均衡。具体过程如下:

首先, 把任务域的二维分布转化为一维分布。这种转化可以通过循环的形式来执行, 即为每个处理器分配相同的时间片来确保负载均衡。然后在一维空间内进行循环分块, 最后保证每个子块包含相同数目的计算任务, 除非在子块内存在云覆盖的像素, 通过一些阈值避免这些像素参与复杂的数值计算。这一问题将在本文的第 4 章进行详细论述。

图 1 为 SYNTAM 算法的并行实现过程, 表示将一个长度为 12 的向量分别映射到 4(图 1(a)) 和 5(图 1(b)) 个节点上执行 AOD 反演的情况。

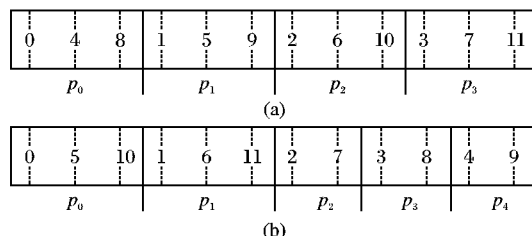


图 1 数据分割示意图

下面以对一个长度为  $M$  的向量  $V$  进行 SYNTAM 算法的并行设计为例, 一维循环分块的 AOD 反演并行实现过程如下(其中  $myid$  为计算节点,  $nprocs$  为节点个数):

- 1) 初始化;
- 2) 广播或分发数据给处理器 / 节点;
- 3) for  $myid = 1$  to  $nprocs$
- 4) 计算  $V(myid)$

- 5)  $myid = myid + 1$
- 6) 计算结束
- 7) 收集计算结果,并组合、转换到相同的单位
- 8) 以文本和图的形式输出 AOD

需要强调的是,在 SYNTAM 并行算法中,AOD 的计算不存在处理器之间的通信,通信只发生在输入数据的分发和计算结果的收集。

### 3 并行反演试验

我们在中国气象局的 IBM System Cluster 1600 集群系统上进行并行反演试验,该集群系统由 IBM SP Power4 +, eServer pSeries 655 (1.7 GHz Power4 +) /Federation 组成,每个 p655 节点有 8 个处理器,共 3 200 个处理器,为大规模计算问题提供了 AIX 或 Linux 的解决方案。该系统通常采用一种集群系统管理软件(CSM)来管理,对一维分块和 MPI,通常选择 AIX 5.3 操作系统和 LoadLeveler 来优化资源利用与输出。

为了证明 SYNTAM 并行算法的有效性,我们以 2002 年 12 月 12 日 10:30 获取的覆盖全国陆地下垫面的 TERRA/MODIS 和 AQUA/MODIS 影像遥感反演 AOD 为例。

依据本文第 1 章所描述的 SYNTAM 算法的要求,采用 TERRA/MODIS 和 AQUA/MODIS 的 1,2,3,4,7 波段的数据(其中第 1、2 波段空间分辨率为 250 m,第 3、4、7 波段为 500 m)的数据作为输入数据。并在处理前,重采样到 1 km。试验采用 Albers 等面积投影,大约有  $4\,500 \times 4\,000$  个像素需要并行计算。并在数值计算前提出被云覆盖的像素,因为这些像素大大的影响着负载平衡,最终影响并行算法的性能。

利用 SYNTAM 并行算法反演 AOD 的工作流程如图 2 所示。除 MODIS 影像几何和辐射校正外,所有步骤均在中国气象局的 IBM 1600 集群系统上执行。

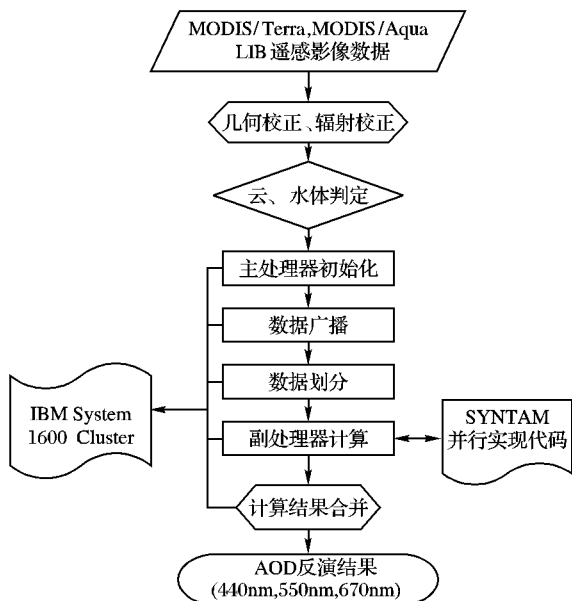


图2 利用 SYNTAM 并行算法反演 AOD 流程

考虑到系统的特殊性,首先需要明确几个相关概念。加速比(Speedup)是并行计算系统中一个经常使用的性能评价指标。加速比的定义如下:

$$S_p = T/T(p) \quad (3)$$

其中  $T$  表示在单个处理器上线性算法的处理时间,  $T(p)$  表示在  $p$  个处理器上并行算法的处理时间。

并行效率(Efficiency)  $E_p$  可以表示为:

$$E_p = S_p/p \quad (4)$$

### 4 试验结果与分析

图3显示了第3章所描述的不同节点数 AOD 并行反演算法的所需的计算时间。很明显,随着集群系统所使用的计算节点数的不同,计算时间会发生显著变化。图3还表明在计算节点从1增加到32个的过程中,处理时间是与计算节点个数成线性关系下降的。但节点个数超过16个,计算时间下降缓慢,这意味着16个计算节点可能最优。

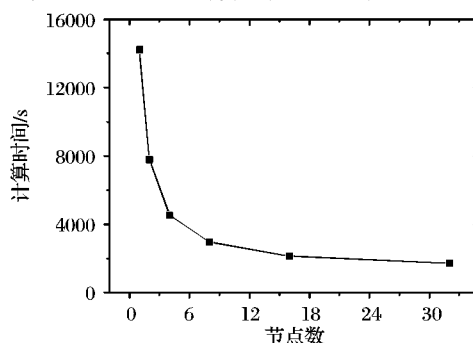


图3 利用并行 SYNTAM 反演算法全国陆地下垫面 AOD 计算时间

表1进一步证明了利用并行算法可以有效的降低计算时间,比如利用32个节点需要1738 s,而在单处理器上串行处理则需要39个小时。而且,随着处理器个数的增加,加速比成线性提高(假定处理节点不超过16个),这表明 SYNTAM 并行算法具有很好的伸缩性。但在32个节点是,加速比增长缓慢,这主要是由于硬件的限制,比如:CPU 带宽、网络通信等。或者最有可能的是由于包含云覆盖像素的不参与计算造成的负载不平衡。

表1 并行 SYNTAM 算法基于不同计算节点的计算性能

节点数	耗时/s	加速比	并行效率
1	14225	—	—
2	7791	1.83	0.92
4	4560	3.12	0.78
8	2973	4.78	0.60
16	2147	6.63	0.41
32	1738	8.18	0.26

图4显示了稳定状态下,随着处理节点的增加,并行执行效率和加速比的趋势。加速比增幅趋缓,而并行效率则逐渐下降。

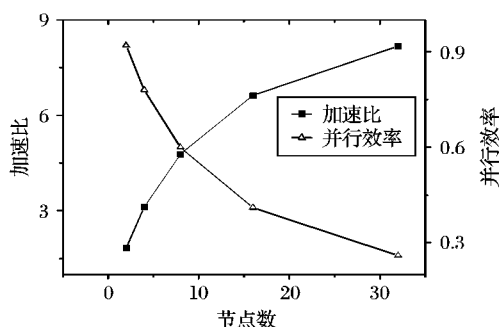


图4 SYNTAM 并行算法的计算效率和加速比

如图5所示为利用 SYNTAM 并行算法得到的 AOD 计算结果。由图可知 AOD 的空间分布,并可作为初始场用于空气质量建模预报空气污染。

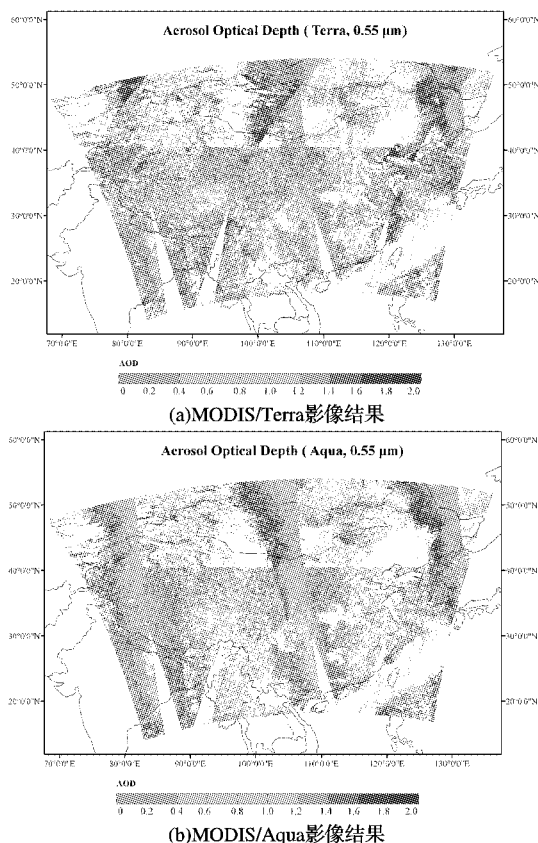


图 5 基于 MODIS 影像反演 AOD 的结果

为了验证 SYNTAM 算法的并行计算精度,我们收集了与 MODIS 过境时匹配的中国气象局气溶胶监测网多个观测站获取的太阳光度计实测 AOD,并进行了对比(详见表 2)。由于中国气象局气溶胶监测网仅装备了三个常用波段的太阳光度计,因此 470 nm 和 550 nm 的 AOD 值需要利用文献[19]所提出的方法进行插值计算得到。波长为 550 nm 的 AOD 反演相对误差如表 2 所示,最大误差值为 17%。初步反演结果表明这种并行算法在气溶胶遥感监测可行性非常强,但是在今后的研究工作中需要进一步提高 AOD 反演的自动化水平和精度。

表 2 利用地面观测 AOD 对 SYNTAM 并行算法反演结果验证

站点(经度°E; 纬度°N)	$\tau_{550}$	$\tau'_{550}$	$\Delta\tau_{550}/\%$
塔中(83.67, 39.00)	0.22	0.26	2
民勤(103.08, 38.63)	0.54	0.58	8
兰州(103.88, 36.05)	1.32	1.40	6
龙凤山(127.60, 44.73)	0.12	0.14	17
榆社(112.98, 37.07)	0.72	0.75	4
上甸子(117.01, 40.65)	1.02	0.99	3
北京(116.47, 39.80)	2.53	2.63	4

## 5 结语

本文所描述的 SYNTAM 并行算法基于 MPI 和循环分块实现,通过一个个例反演试验,证明该算法是一种有效的陆地下垫面 AOD 反演方法。

不同数目计算节点的试验表明,该算法具有很强的伸缩性,而且能够最小化处理器间通信,同时最大化负载平衡。同时,与串行算法相比,SYNTAM 并行算法能够达到降低 AOD 反演时间,具有近似线性的加速比的目标。循环分块极大地提高了 AOD 计算的并行效率,即意味着在试验中达到了最优的负载平衡。

结果表明 SYNTAM 并行算法具有较高的性能,因此可以说将来 AOD 的近实时自动反演具有很好的前景,如其实时 AOD 产品可以广泛地应用空气质量模型和气候模式。

## 参考文献:

- [1] REMER L A, KAUFMAN Y J, TANRE D, *et al.* The MODIS aerosol algorithm, products, and validation [J]. *Journal of the Atmospheric Sciences*, 2005, 62 (4): 947–973.
- [2] ENGEL-COX J A, HOLLOMAN C H, COUTANT B W, *et al.* Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality [J]. *Atmospheric Environment*, 2004, 38 (16): 2495–2509.
- [3] HSU N C, TSAY S C, KING M D, *et al.* Aerosol properties over bright reflecting source regions [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2004, 42 (33): 557–569.
- [4] MARTONCHIK J, DINER D. Retrieval of aerosol optical properties from multi-angle satellite imagery [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 1992, 30 (22): 223–230.
- [5] GREY W M F, NORTH P R J, LOS S O. Computationally efficient method for retrieving aerosol optical depth from ATSR-2 and AATSR data [J]. *Applied Optics*, 2006, 45 (12): 2786–2795.
- [6] DINER D J, ABDOU W A, CONEL J E, *et al.* MISR aerosol retrievals over southern Africa during the SAFARI-2000 dry season campaign [J]. *Geophysical Research Letters*, 2001, 28 (16): 3127–3130.
- [7] DEUZE J L, HERMAN M, GOLOUB P, *et al.* Characterization of aerosols over ocean from POLDER/ADEOS-1 [J]. *Geophysical Research Letters*, 1999, 26 (10): 1421–1424.
- [8] CHANG C I. *Hyperspectral imaging: Techniques for spectral detection and classification* [M]. New York: Kluwer, 2003.
- [9] PLAZA A, PLAZA J, VALENCIA D, *et al.* AMEPPAR: Parallel morphological algorithm for hyperspectral image classification in heterogeneous networks of workstations [EB/OL]. [2008–10–10]. <http://www.umbc.edu/rssipl/people/aplaza/Papers/Journals/2006.LNCS.Ameepar.pdf>.
- [10] PLAZA A, VALENCIA D, PLAZA J, *et al.* Commodity cluster-based parallel processing of hyperspectral imagery [J]. *Journal of Parallel and Distributed Computing*, 2006, 66(3): 345–358.
- [11] VALENCIA D, PLAZA A. FPGA-based compression of hyperspectral imagery using spectral unmixing and the pixel purity index algorithm [M]. [S. l.]: Springer Berlin / Heidelberg, 2006: 24–31.
- [12] FALLAH-ADL H, JAJA J, LIANG S. Fast algorithms for estimating aerosol optical depth of thematic mapper imagery [J]. *Journal of Supercomputing*, 1997, 10(4): 315–330.
- [13] GUO J P, XIAO H D, CAO C X, *et al.* Fast aerosol optical depth retrieval from MODIS [EB/OL]. [2008–10–10]. <http://www.igars08.org/Abstracts/pdfs/1280.pdf>.
- [14] TANG J K, XUE Y, YU T, *et al.* AOD Determination by Exploiting the Synergy of TERRA and AQUA MODIS [J]. *Remote Sensing of Environment*, 2005, 94(3): 327–334.
- [15] KING M D, KAUFMAN Y J, TANRE D, *et al.* Remote sensing of tropospheric aerosols from space: Past, present and future [J]. *Bulletin of the American Meteorological Society*, 1999, 80 (11): 2229–2259.
- [16] CHANDRASEKHAR S. *Radiative transfer* [M]. New York: Dover Publication, 1960.
- [17] XUE Y, CRACKNELL A P. Operational bi-angle approach to retrieve the earth surface albedo from AVHRR data in the visible band [J]. *International Journal of Remote Sensing*, 1995, 16(3): 417–429.
- [18] BROYDEN C G. A class of methods for solving nonlinear simultaneous equations [J]. *Mathematics of Computation*, 1965, 19(92): 577–593.
- [19] CHE H, SHI G, UCHIYAMA A, *et al.* Intercomparison between aerosol optical properties by a PREDE skyradiometer and CIMEL sunphotometer over Beijing, China [J]. *Atmospheric Chemistry and Physics*, 2008, 8(12): 3199–3214.