Grid-based Aerosol Retrieval over Land Using Dark Target Algorithm from MODIS Data

Jiakui Tang[1], Yong Xue[2,3], Yanning Guan[1], Yincui Hu[1], Dong Shuo[4]

[1]National Oceanic Technology Center , Tianjin, 300111, China
[2]State Key Laboratory of Remote Sensing Science, Jointly Sponsored by the Institute of Remote Sensing Applications of Chinese Academy of Sciences and Beijing Normal University, Institute of Remote Sensing Applications, Chinese Academy of Sciences, P.O. Box 9718, Beijing 100101, China
[3]Department of Computing, London Metropolitan University, 166-220 Holloway Road, London N7 8DB, UK
[4]Department of Resource & Environment Science, Shijiazhuang College, Shijiazhuang, 050035, China
{tangjiakui@263.net}

Abstract. The grid refers to an infrastructure that dynamically aggregates computing resources which are geographically distributed or heterogeneous, and leverages on resources one don't own for oneself computational intensive applications. The technology has been used in applications in many disciplines. However, applications in aerosol retrieval from remote sensing data are just beginning tried. Dark Target algorithm for MODIS data has showed excellent competence at the aerosol distribution and properties retrieval, however, the implementation of the algorithm for aerosol remote sensing retrieval is very computational intensive even though LUT method is used for speeding up the retrieval computation. This paper mainly focuses on realization of aerosol retrieval using Dark Target algorithm on GCP-ARS (Grid Computation Platform for Aerosol Remote Sensing), which is one grid middleware developed in Institute of Remote Sensing Applications, Chinese Academy of Sciences. Issues of parameterization, task partitioning and submitting, dark target algorithm used for aerosol retrieval, and the collection of results are described in detail. Experimental results obtained using Condor-pool consisted of commodity PCs are finally discussed.
Keywords: Aerosol Retrieval, Grid Computing, AOT, Dark Target Algorithm.

1. Introduction

Grid computing is a promising technology that seeks to easily and efficiently coordinate the sharing of geographically distributed computing resources, thereby bring supercomputing power to its users [1](Foster *et al.* 2001). Grid computing distinguished from conventional distributed computing, cluster computing, which are both more constrained to computation on a local area networked of processors, by its focus on large-scale computing resources sharing, user's applications to utilize resources that spread across wide area networks. Much inspiriting progresses have been achieved since the term "Grid" was coined in the mid 1990s [2] (Foster and Kesselman 1998). Globus[3] (Foster and Kesselman 1996), consisting of a set of services and software libraries, is a grid computing software toolkit addressing the key technical problems in the development of grid-enabled tools, services, applications and systems. Built on top of Globus, Nimrod/G [4](Buyya *et al.* 2000) is a resource management system for scheduling of grid applications. Integrated with Globus toolkits, EuroGrid aims to create tech to for remote access to supercomputer resources and simulation codes in GRIP. Condor, which is one Grid computing developed by Condor team in university of Wisconsin- Madison US, creates one High-Throughput computing (HTC) environment by its ability to effectively harness of idle cycles of non-dedicated, preexisting resources under distributed ownership. Other grid systems/projects can be refered to Tang *et al.* (2004) [5].

Grid-solvers or middlewares have also been developed on these systems aforementioned. A web-based problem solving environment to simplify the submission, monitoring, and steering of Master-Worker[6] (Goux *et al.* 2000) for Grid computing applications is introduced by Good *et al*. (2000) [7], A architecture of matching grid application requirements to a set of heterogeneous grid resources for resources allocation in computational grid is proposed by Czajkowski *et al*. (1999) [8]. A definition of Dynamic Earth Observation System (DEOS) and the evolution of SARA middleware that integrated Web and Grid technologies were presented by Aloisio *et al.* (2000) [9], SARA provides the users with a high level web interface to search for remote sensing images and start an on-demand post –processing on them. Our grid middleware, GCP-ARS (Tang *et al.* 2004) [5] (Grid computing Platform for Aerosol Remote Sensing), based on the Condor Grid system, which aims to bridge the users and grid computing platform, is one portable and scalable grid middleware to evaluate the aerosol remote sensing retrieval on grid computing platform, and to get the real-time or near real-time aerosol information such as distribution for environment monitoring etc. GCP-ARS can dynamically aggregates computer resources through the Internet/Intranet into one computing environment, then harness their idle resources for aerosol retrieval computation.

Aerosol retrieval using satellite images is the procedure to separate the atmospheric molecular contribution from the TOA (Top Of Atmosphere) reflectance to abstract the aerosol information such as aerosol optical thickness, size distribution, which are key parameters for climate modeling, satellite image atmospheric correction, environment pollution monitoring, etc. At the present time, the typical operational aerosol retrieval model for sea surface or land surface such as DDV [10](Kaufman *et al*. 1997) (Dark Dense Vegetation) pixels or Dark target algorithm usually include such steps as spectrum channel (band) TOA Reflectance (Apparent Reflectance) estimation from satellite image, Ground Surface Reflectance estimation based directly on satellite image or others such in situ measurements, Look-up Table (LUT) generation, Table Look-up and interpolation to get the retrieval result image. Dark target algorithm is heavily computationally intensive for the last two steps involving into large amount of RTE (Radiative Transfer Equations) calculations in LUT generation and interpolation computation when looking up the LUT. This paper mainly focuses on the application of grid computing technology for aerosol retrieval over land dark target pixel from MODIS data based GCP-ARS, which is one grid middleware we are developing.

The rest of paper is as follows. In section 2, we introduce the scheme of GCP-ARS, one grid computing middle based Condor system. In section 3, Dark target algorithm used for aerosol retrieval over land is described. Section 4 describes the distributed aerosol retrieval computing using GCP-ARS. In section 5, the experiments results and analysis are presented. Section 6 contains the concluding remarks.

2. The Scheme of GCP-ARS

GCP-ARS, developed by Telegeoprocessing Research group in Institute of Remote Sensing Applications (IRSA), Chinese Academy of Sciences (CAS), is a advanced high throughput computing grid middleware that supports the execution of aerosol remote sensing retrieval over a geographically distributed, heterogeneous collection of resources. In GCP-ARS, Condor system provides the technologies for resource discovery, dynamic task-resource matching, communication, and result collection, etc.

LUTGEN module [5] provides the users of the power to generate the LUT on grid computing platform, LOOK-UP module can be used to look up and interpolate the LUT generated by LUTGEN module or user self-offered to determine aerosol optical thickness. GCP-ARS provides the users friendly Windows-based GUI (Graphical User Interface), computation tasks auto-partitioned, tasks auto-submission, the execution progress monitoring, the sub-results collection and the final result piecing, which results in screening the complexities of grid applications programming and the Grid platform for users. GCP-ARS architecture mainly consists of three entities: clients, a resource broke, and producers. Aerosol retrieval is launched though a client GUI for execution at producers that share its idle cycles through the resource broker. The resource broke manages tasks application execution, resource management and result collection by means of Condor system

3. Dark target algorithm for aerosol retrieval description

Since the original application to aerosol remote sensing retrieval over land from Landsat Multispectral Scanner (MSS) data [11](Kaufman and Sendra 1988), Dark target algorithm has been a popular approach used for aerosol retrieval over land surface. After the launch of MODIS [12](Kaufman *et al.* 1997b), it has been adapted to the operational algorithm for the remote sensing of aerosol over land and further for the distribution of aerosol concentration, for example, biomass in the tropic or urban industrial aerosol in the midlatitudes continuously monitoring. This algorithm takes advantage of the MODIS wide spectral range and high spatial resolution and the strong spectral dependence of the aerosol opacity for most aerosol types that result in low optical thickness in the mid-IR channel. The main steps of dark target algorithm from MODIS data used in this paper are (1) identification of dark target pixel in the mid-IR; (2) estimation of the surface reflectance at visible band; (3) determination of the aerosol optical thickness at the visible band. In step (1), we take the pixels with reflectance at 2.1um, less that 0.15 as Dark target pixel for aerosol optical thickness determination according to proposed by Kaufman *et al.* (1997). In step (2), the visible band reflectance at 0.47um and 0.65um are estimated by the relationship between surface reflectance in the visible and mid-IR as follows:

$$\rho_{0.47um}=0.5*\rho_{2.13um} \qquad (1)$$

$$\rho_{0.65um}=0.25*\rho_{2.13um} \qquad (2)$$

In step (3), LUT method is used for aerosol optical thickness retrieval computation. LUT method is popular used in aerosol remote sensing retrieval [10,13](Kaufman *et al*. 1997; Mobley *et al*. 2002), LUT is a pre-generated table which denotes a matrix of dependency variable corresponding a set of combinations of independent variable values. For aerosol retrieval, we refer to LUT as apparent reflectance matrix corresponding to Sensor-Target-Solar geometric conditions, ground surface reflectance, aerosol and atmosphere conditions. In this paper, the LUT used was pre-generated by LUTGEN module of GCP-ARS. The LUT consists of dependency Apparent Reflectance variable corresponding to independent variables: solar zenith, sensor zenith, relative zenith, ground surface reflectance and aerosol optical thickness (AOT).

4. Distributed Aerosol Retrieval using GCP-ARS

Aerosol retrieval using dark target algorithm is a pixel by pixel computing procedure, which make it possible for distributed aerosol retrieval computation. Although the speedup by using LUT method, AOT computation of a swath MODIS image data requires more than ten hours on a single computer, which make grid computing well matching to addressing this problem. Through analyzing the procedure of Aerosol retrieval using dark target algorithm, in GCP-ARS, we can implement the procedure by three main sequential phases: data files input, execution and termination. Data files refer to LUT file that pre-generated by LUTGEN module of GCP-ARS, MODIS image data file refer to separated solar zenith angle file, sensor zenith angle file, relative azimuth angle file, surface reflectance file and apparent reflectance file in ASCII format pre-processed by other software such as ENVI (http://www.ResearchSystems.com/envi) or processed immediately by GCP-ARS from MODIS 1B lever data in HDF format. The second LUT look-up and interpolation execution phase can be separated as repetitive tasks for the entire satellite image, which can be submitted to Grid computing platform for simultaneously execution, thereby reducing the total execution time. The LOOKUP module of GCP-ARS facilitates the table look-up and interpolation computation for aerosol retrieval on grid computing platform consists of a pool or pools of networked computing resources. LOOKUP mainly consists of seven sub-modules as shown in Fig. 1: Select LUT file (Fig. 1.1), Select MODIS file (Fig. 1.2), tasks partition (Fig. 1.3), tasks submission (Fig. 1.0), AOT merging (Fig. 1.4) and GCP console (Fig. 1.5). After the first 3 sub-module is done well, if click the fourth highlighted sub-module command menu, the tasks will be auto-submitted and execute on grid computing platform, the progress bar window (Fig. 1.6) will start up to monitor the progress of execution of all tasks. When one of all the tasks submitted is executed completed, the corresponding AOT result image will automatically be displayed in a new window. Sub-module of GCP console is always highlighted waiting for user clicked to dynamically check or change the state of grid pool and manage the executing tasks. By clicking the fifth module AOT Merging command menu after all tasks execution termination, the sub-AOTs of all the tasks will be auto-collected and merged into one final whole AOT data file, which is saved in ASCII format.

The most challenge of using grid computing techniques for aerosol retrieval mainly includes the partitioning of tasks and sub-AOTs merging, especially automatization of this procedures. Partitioning of tasks is very crucial since available grid resources are dynamic and their performance maybe diverse. To realize automatization of tasks partitioning and balance the producers' tasks load, the strategy for tasks partition used in LOOKUP is to partitions the number tasks according user's expectation as equally as possible.

5. Experiments and Analysis

Our experiments were carried out on a grid computing pool consists of 5 low commodity PCs nodes connected using 100Mbps Ethernet switch. The information of each node is shown in Table 1.

Table 1. The structure information of LUT generated in our test

IP	Hostname	Arch-OS	Role
192.168.0.5	HU	Intel/WIN50	Manager/Producer/Client
192.168.0.3	ZHONG	Intel/WIN50	Producer/Client
192.168.0.109	WANG	Intel/WIN50	Producer/Client
192.168.0.120	JENNYJORDAN	Intel/WIN50	Producer/Client
192.168.0.100	IRSA-CGY	Intel/WIN50	Producer/Client

The date and time of MODIS data we have chosen for our test on GCP-ARS were acquired at 03:20 UTC on August 11, 2003. The image size is 512 × 512 with spatial resolution 500*m*, which covers most part of plain of north China and

Bohai gulf. The apparent reflectance image of MODIS band 3 at 470 *nm* is showed in Fig. 2, Fig. 3 shows the final AOT image retrieved using GCP-ARS.

There are 5 producers available for our aerosol retrieval experiment using GCP-ARS, To get higher performance according to the suggestion by Tang *et al.* (2004)[5], the whole image computation was partitioned into 5 tasks as equally as possible. To check the performance of GCP-ARS, contrarily, traditional computational test was carried on a single Intel CPU P4 1.6 GHz computer with 512Mb RAM, we define that the execution time Tg for aerosol retrieval on grid platform was the duration from the time of tasks submission to the time of all tasks executed successfully, which includes the time consumed for the generation of tasks at the Condor manager, communication time in shipping tasks to producers, producer's execution time, and the results collection time from producers, by contrary, the sequential execution time (Ts) for a single computer.

Table 2 shows the performance for 5 tasks partitioned and 5 available producers in our preliminary experiment. The execution time on GCP-ARS Tg is much less than on a single computer, which demonstrates the good performance of GCP-ARS for improving the efficiency of aerosol retrieval. Rationally, we can infer better performance if more and more producers are available. Typically, more and more satellites are or will be available for aerosol retrieval and atmosphere remote sensing to acquire the simultaneous RTM parameters, which can be exploited for GCP-ARS, if many producers are available, to compute as quick as real-time or near-real-time aerosol information. Of course, we maybe have to make a tradeoff due to more time needed for Condor manager to handle more available producers and more tasks, further experiments are undergoing.

Table 2. Aerosol retrieval performance using GCP-ARS contrasted to single computer

	Submission time (hour:min:sec)	Termination time (hour:min:sec)	Time consumed (Tg & Ts) (hour:min:sec)
Single computer	19:07:37	19:31:42	00:24:05
GCP-ARS	18:45:02	18:51:13	00:06:11

6. Conclusion and further development

The above experiments results of distributed aerosol retrieval computing are very encouraging, which show the good potential of grid computing technology for getting aerosol information retrieval from satellite data. By means of LOOKUP module of our GCP-ARS middleware, Aerosol optical thickness images could be quickly determined from MODIS data using grid computing technique, which, by pooling and aggregating the idle CPU circle together, has high throughput computing to meet the needs of intensive interpolation computations. Using grid computing platform, real time or near real time aerosol information maybe retrieved, which will meet the needs such as fast environment monitoring and decision making, climate modeling and forecasting, etc.

Our work for implementation of other aerosol retrieval algorithms or models is ongoing well, other further work directions cover the integration of GCP-ARS with web service which make user possible remote access over the Internet, tasks un-equally partitioned according the actual performance of available producers to make producers' load balanced.

Acknowledgement

This work is support by the research projects "Grid platform based aerosol fast monitoring modeling using MODIS data and middlewares development" (Authorization Number 40471091) funded by National Natural Science Foundation of China (NSFC), "CAS Hundred Talents Program" (BRJH0101) funded by Chinese Academy of Sciences, "Aerosol retrieval using MODIS data " (HX020032) funded by Beijing Environment Monitor Center, China, and "Dynamic Monitoring of Beijing Olympic Environment Using Remote Sensing" (2002BA904B07-2) funded by the Ministry of Science and Technology, China. The advice and assistance of Prof. Guo Shan is gratefully acknowledged.

References

[1] Foster, I., Kesselman, C. and Tuecke, S., 2001, the Anatomy of the Grid, *International Journal on Supercomputing applications*. www.globus.org/research/papres/anatomu.pdf.

[2] Foster, I., and Kesselman, C., (Editors). 1998, The Grid: Blueprint for a new Computing Infrastructure. (*Morgan Kaufmann Publishers*, Inc.).

[3] Foster, I., and Kesselman, C., 1996, Globus: A Metacomputing Infrastructure Toolkits, *Proceedings of the Workshop on Environment and Tools for Parallel Scientific Computing,* SIAM Lyon, France.

[4] Buyya, R., Aramson, D., and Giddy, J., 2000, Nimrod/G: An Architecture for a Resource Management and Scheduling System in a Global Computational Grid, *Proceedings of the* 4^{th} *International Conference and Exhibition on High-Performance computing in the Asia-Pacific Region (HPCASIA' 2000),* China, IEEE CS Press, USA.

[5] Tang, J.K., Xue, Y., Guan, Y.N., Liang, L.X., Hu, Y.C, Luo, Y., Cai, G.Y., Zhang, A.J., Wang, J.Q., Zhong, S.B., and Wang, Y.G., 2004, *A New Approach to Generate the Look-up Table Using Grid Computing Platform for Aerosol Remote Sensing, 2004 IEEE International Geoscience and Remote Sensing Symposium (IGARSS'04)*

[6] Goux, J.P., Kulkani, S., Linderoth, J., and Yoder, M., 2000, An enabling framework for master-worker application on the computational grid., *Proceedings of 9 IEEE International Symposium on High Performance Distributed Computing (HDPC-9).*

[7] Good, M., and Goux, J.P., 2000, iMW: Aweb-based problem solving environment for Grid computing applications. *Technical report,* Department of Electrical and Computer Engineering, Northwestern University.

[8] Czajkowski, K., Foster, I., and Kesselman, C., 1999, Resource coallocation in computational grids, *Proceedings of the 8 IEEE International Symposium on High Performance Distributed Computing (HDPC-8).*

[9] Aloisio, G., Cafaro, M., Falabella, F.P., Kesselman, C., Williams, R., 2000, Web Access to Supercomputing Using the Grid, available from http://www.cacr.caltech.edu/

[10] Kaufman, Y.J., Tanre, D., Remer, L., Vermote, E., Chu, A., Holben, B., 1997, Operational remote sensing of tropospheric aerosol over land from EOS moderate resolution imaging spectroradiometer, *Journal of Geophisics Research,* Vol, 102, No, D14, 17051-17067.

[11] Kaufman,Y.J., and Sendra, C., 1988, algorithm for automatic atmospheric corrections to visible and near-IR satellite imagery. *International Journal of Remote Sensing,* 9, pp1357-1381.

[12] Kaufman, Y. J., Wald, A. E., Remer, L. A., Gao, B. C., Li, R.R, & Luke F., (1997b). The MODIS 2.1-um Channel-Correlation with Visible Reflectance for Use in Remote Sensing of Aerosol. *IEEE Transactions on Geoscience and Remote Sensing,* Vol.35, pp1286-1298

[13] Mobley, C., Sundman, L. K., Davis, C. O., Montes, M., and Bissett, W. P., 2002, A Look-up Table Approach to Inverting Remotely Sensing Ocean Color Data, *Ocean Optics XVI*, Santa Fe, New Mexico.

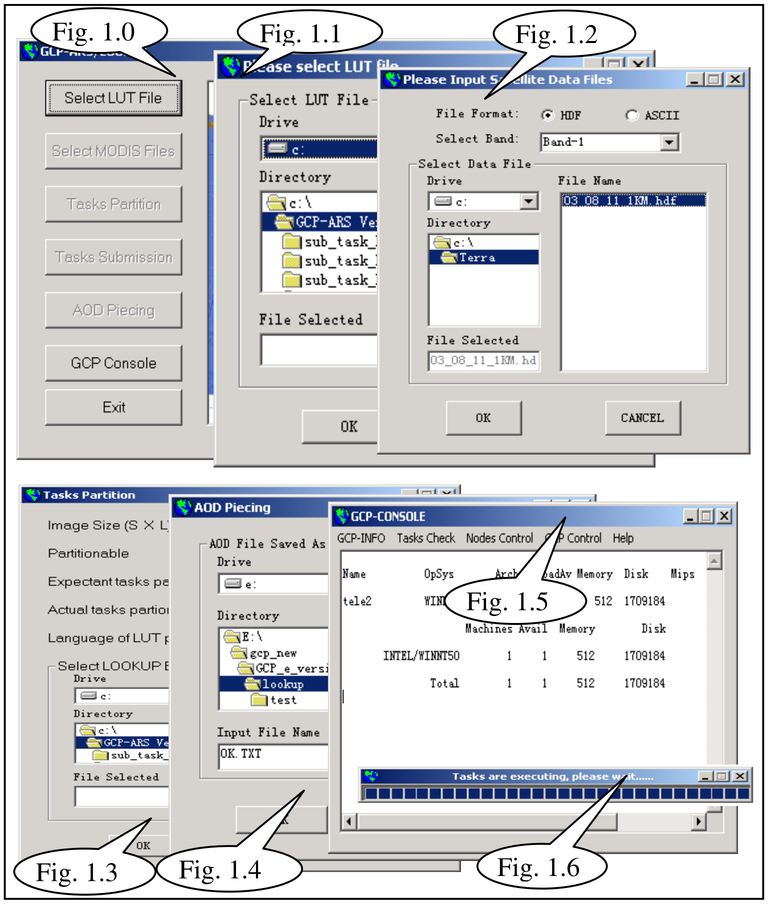

Fig. 1. Sub-modules of LOOKUP of GCP-ARS

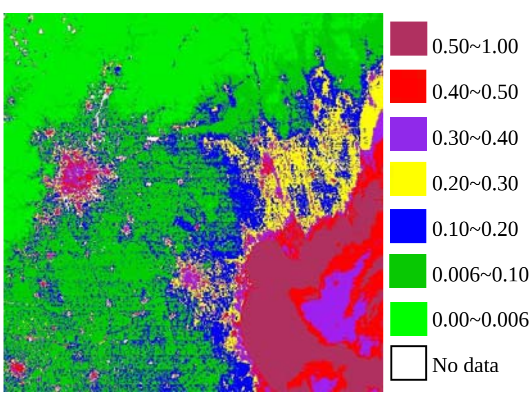

Fig. 3. AOT Image of MODIS Band 3 at 470*nm* retrieved using GCP-ARS

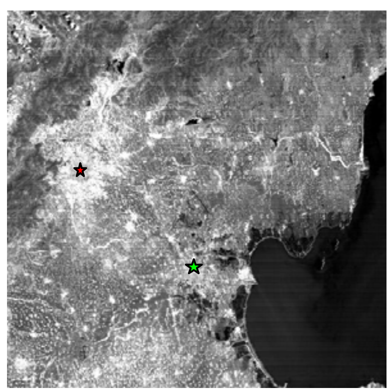

Fig. 2. Apparent Reflectance Image of MODIS Band 3 at 470*nm*. Red and green pentagons stand for Beijing city and Tianjin city location respectively.