Data-Parallel Method for Georeferencing of MODIS Level 1B Data Using Grid Computing

Yincui Hu[1], Yong Xue[1,2,*], Jiakui Tang[1], Shaobo Zhong[1], and Guoyin Cai[1]

[1] 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

[2] Department of Computing, London Metropolitan University, 166-220 Holloway Road, London N7 8DB, UK

{huyincui@163.com, y.xue@londonmet.ac.uk}

Abstract. Georeference is a basic function of remote sensing data processing. Geo-corrected remote sensing data is an important source data for Geographic Information Systems (GIS) and other location services. Large quantity remote sensing data were produced daily by satellites and other sensors. Georeferenceing of these data is time consumable and computationally intensive. To improve efficiency of processing, Grid technologies are applied. This paper focuses on the parallelization of the remote sensing data on a grid platform. According to the features of the algorithm, backwards-decomposition technique is applied to partition MODIS level 1B data. Firstly, partition the output array into evenly sized blocks using regular domain decomposition. Secondly, compute the geographical range of every block. Thirdly, find the GCPs triangulations contained in or intersect with the geographic range. Then extract block from original data in accordance with these triangulations. The extracted block is the data distributed to producer on Grid pool.

1 Introduction

Large quantity imagery data of remote sensing produced daily by variable satellites and other sensors. The processing of remote sensing data is computationally intensive. It requires parallel and high-performance computing techniques to achieve good performance. G eo-corrected remote sensing data are source data for geographic information systems and other location services. When the correction formulation is complicate, the large image georeference will require a mass of time. Over the past decade, Grid has become a powerful computing environment for data intensive and computing intensive applications. Cannataro (2000) proposed to develop data mining services within a Grid infrastructure as the deployment platform for high performance distributed data mining and knowledge discovery.

Researchers have aimed to develop Grid platforms for remote sensing data processing. Aloisio *et al.* (2004) proposed Grid architecture for remote sensing data

* Corresponding author.

V.S. Sunderam et al. (Eds.): ICCS 2005, LNCS 3516, pp. 883 – 886, 2005.

processing and developed a Grid-enabled platform, SARA/Digital Puglia with his research group. SARA/Digital Puglia (Aloisio *et al.* 2003) is a remote sensing environment developed in a joint research project. Our research group has developed a grid-based remote sensing environment, which is the High-Throughput Spatial Information Processing Prototype System in Institute of Remote Sensing Applications, Chinese Academy of Sciences (Cai *et al.* 2004, Hu *et al.* 2004, Wang et al. 2003).

This paper focuses on the parallelization of the remote sensing data on a grid platform. First, we discussed the algorithm of rectifying remote sensing image. Second, data partition for georeference on grid is introduced. Finally, we analyzed the result of georeference on Grid platform.

2 Georeference Implementation on the Grid

2.1 MODIS Level 1B Data

MODIS level 1B products are obtained from **Mod**erate Resolution **I**maging **S**pectroradiometer (MODIS) carried on EOS satellite and have been calibrated. MODIS includes 36 spectral bands extending from the visible to the thermal infrared wavelengths (Running *et al.,* 1994). The MODIS Level 1B products contain longitude and latitude coordinates. These coordinates can be used as Ground Control Points (GCP) to rectify the image.

2.2 Parallel Rectification on Grid

There are three important components to rectify an image, i.e. transformation model selection, coordinate transformation and resample. In this paper, triangle warping model is selected to transform the coordinates. Moreover, cubic resample method is used. The triangular is build from the longitude and latitude coordinates.

The image rectification steps for interpolating, transforming and resampling can be integrated into one routine. This rectify routine can be repeated to correct every part of the large image. Now that the rectify routine can be data parallel processing in Grid, how to partition the data and how to merge the results are the main questions confront us.

2.3 Data Partition Strategy

The partition strategy influences the process efficiency and determines the merge strategy. So to select an efficient partition method is very important. The value of the output pixel is interpolated by value of pixels around its location in original image. The original location is obtained by triangle transformation from GCPs triangulations. According to the features of the algorithm, backwards-decomposition technique is applied to partition MODIS level 1B data. It comprises four steps as follows:

Firstly, partition the output array into evenly sized blocks using regular domain decomposition. Secondly, compute the geographical range of every block. Thirdly, find the GCPs triangulations contained in or intersect with the geographical range.

Then extract block from original data in accordance with these triangulations. The extracted block is the data that will be distributed to producer on Grid pool.

3 Experiments and Analysis

Our experiments were performed on a Grid-computing environment, which is in the High-Throughput Spatial Information Processing Prototype System (HIT-SIP) based on Grid platform in Institute of Remote Sensing Applications, Chinese Academy of Sciences. Test data are MODIS level 1B products. The data format is HDF. The data is partitioned into many parts by backwards-decomposition techniques. The experiment result is shown in table1.

Table 1. Experiment results

Test data size	Sequential Time (sec)	No. of parts	Execution time (sec)
768 MB	Out of memory	10	1208
		80	856
83 MB	404	8	358
		80	281
20 MB	282	4	130
		8	110

The experiment shows that data-parallel georeference is efficient especially for those large-size data. The computer shows errors “out of memory, unable to allocate memory” when rectifying the large data (768 MB) in a computer with 512MB memory. This situation could be solved with the help of the resources in Grid pool. The large data is decomposed into small parts and distributed to the Grid.

4 Conclusions

As an important new field in the distributed computing arena, Grid computing focuses on intensive resource sharing, innovative applications, and, in some cases, high-performance orientation. We implemented data-parallel georeference in HIT-SIP platform. The experiments indicate that Grid is efficient for data-parallel georeference. The efficiency could be improved especially for those large data. For those processing that need large memory, Grid can also provide enough resources to solve the problem. Ongoing work on HIT-SIP includes developing middleware of remote sensing processing and providing remote sensing processing services for Internet consumers.

Acknowledgement

This publication is an output from the research projects "CAS Hundred Talents Program" and “Monitoring of Beijing Olympic Environment” (2002BA904B07-2) and “Remote Sensing Information Processing and Service Node” funded by the MOST, China and "Aerosol fast monitoring modeling using MODIS data and middlewares development" (40471091) funded by NSFC, China.

References

1. Aloisio, G., Cafaro, M.: A Dynamic Earth Observation System. Parallel Computing. 29 (2003), 1357-1362
2. Aloisio, G., Cafaro, M., Epicoco, I., Quarta, G.: A Problem Solving Environment for Remote Sensing Data Processing. Proceeding of ITCC 2004: International Conference on Information Technology: Coding and Computing. Vol.2. IEEE Computer Society, 56-61.
3. Cai GY, Xue Y, Tang JK, Wang JQ, Wang YG, Luo Y, Hu YC, Zhong SB, Sun XS, 2004. Experience of Remote Sensing Information Modelling with Grid Computing. Lecture Notes in Computer Science, Vol. 3039. Springer-Verlag, 989-996.
4. Cannataro, M: Clusters and Grids for Distributed and Parallel Knowledge Discovery. Lecture Notes in Computer Science. 1823(2000), 708-716
5. Hu YC, Xue Y, Wang JQ, Sun XS, Cai GY, Tang JK, Luo Y, Zhong SB, Wang YG, Zhang AJ.: Feasibility Study of Geo-spatial Analysis Using Grid Computing. Lecture Notes in Computer Science, Vol. 3039. Springer-Verlag (2004) 956-963.
6. Lanthier, M., and Nussbaurm, D: Parallel Implementation of Geometric Shortest Path Algorithms. Parallel Computing. 29(2003), 1445-1479.
7. Pouchard, L; Cinquini, L; Drach, B; Middleton, D; Bernholdt, D; Chanchio, K; Foster, I; Nefedova, V; Brown, D; Fox, P; Garcia, J; Strand, G; Williams, D; Chervenak, A; Kesselman, C; Shoshani, A; Sim, A.: An Ontology for Scientific Information in a Grid Environment: The Earth System grid. 3rd IEEE/ACM International Symposium on Cluster Computing and the GRID. IEEE Computer Society (2003) 626-632.
8. Roros, D. -K. D., Armstrong, M. P: Using Linda to Compute Spatial Autocorrelation in Parallel. Computers & Geosciences. 22(1996), 425-432
9. Roros, D. -K. D., Armstrong, M. P.: Experiments in the Identification of Terrain Features Using a PC-Based Parallel Computer. Photogrammetric Engineering & Remote Sensing. 64(1998), 135-142
10. Wang, S. and Armstrong, M. P.: A Quadtree Approach to Domain Decomposition for Spatial Interpolation in Grid Computing Environments. Parallel Computing. 29(2003) 1481-1504.
11. Jianqin Wang, Yong Xue, and Huadong Guo, 2003, A Spatial Information Grid Supported Prototype Telegeoprocessing System. In Proceedings of 2003 IEEE International Geoscience and Remote Sensing Symposium (IGARSS’2003) held in Toulouse, France on 21-25 July 2003, v 1, p 345-347.