A Cluster-based Remotely Sensed Image Mosaic Algorithm with Parallel IO

ShiYong Liu[1], QiuYun Wu[2], Luo Chen[3], Jun Li[4], Ning Jing[5]

College of Electronic Science and Technology, National University of Defense Technology, China

Keywords: MPI, Parallel, Image Mosaic.

Introduction

This document aims to use the cluster system to solve the low efficiency problem of traditional remote sensing image mosaic method in the face of large-scale data sets.

Remote sensing image mosaic is very common in remote sensing image processing (Chen et al. 2015; Yang et al. 2015; Chen et al. 2014), since a single scene usually cannot cover large spatial extents of interest. The whole process is data- and computation-intensive and the IO operation accounted for the vast majority of time cost (Benkelman et al. 2004). With the improvement of the accuracy of image sensors and development of satellite and unmanned aerial vehicle technology, the acquisition interval of remote sensing image become shorter and shorter, single remote sensing image data is also growing (Wang et al. 2010; Chen et al. 2011; Berriman et al. 2007), meanwhile our environment undergoing rapid changes, there has been a relevantly urgent demand for high resolution large area information for environmental monitoring, hazard assessment, large range change detection, this has highly promoted a high request of rapid processing of image data. Traditional remote sensing image mosaic algorithm and commercial software only use single machine resource, can hardly quickly deal with large data sets. Therefore, parallel computing methods that can fully leverage state-of-the-art hardware platforms and that can be considered to adapt to mosaic algorithms. Hence we put forward a parallel mosaic method which named ParallelMosaic, the parallel IO mechanism is implemented in a parallel file system. The experiments were tested in a cluster environment and the results show that this method greatly improves the efficiency of image mosaic.

Experimental data for a large data set is split into several sub data sets of different sizes. Each data in the dataset are come from the GF-2 satellite images. Detailed information on datasets please see TABLE I.

TABLE I. Information about each image in datasets

Dataset name	Image information				
	Size	***Extend (Pixel)***	***Band count***	***Pixel type***	***Resolution***
Dataset1	5.4GB 5.0GB	54241×29702 38265×38376	3	Byte	1m
Dataset3	5.4GB 5.0GB 4.5GB	54241×29702 38265×38376 34364×39058	3	Byte	1m
Dataset3	5.4GB 5.0GB 4.5GB 5.2GB	54241×29702 38265×38376 34364×39058 43989×34851	3	Byte	1m

In this paper the whole process of parallel image mosaic can be divided into 5 parts. First, access to the folder with all the mosaic images. Second, obtain the geographical scope of all the mosaic image, calculate the minimum bounding box that all the included images formed. Third, the master node creates an empty set of results file based on the size of minimum bounding box, then the master node writes the resolution, bands count and other metadata information into the empty result file, the resolution of the resulting file is selected as the minimum value in all the mosaic images. Forth, Task partitioning, using MPI multi process parallel method to assign tasks for each process. Each process independently and separately reads the data from the image to the memory according to task partitioning rule before. Meanwhile the whole reading course is parallel, if the amount of each image data in dataset is very small, the parallel strategy is that one process is responsible for one image (Figure I), else if each of the image data in dataset is very big, the method is that all processes mapping to one image, all the processes read and write in parallel to one file at the same time (Figure II). Fifth, each process resampling the data it reads if it's resolution don't equal to the result file, then handle the no data value. Finally calculate the writing position, each process writes the effective value region into the empty result file.

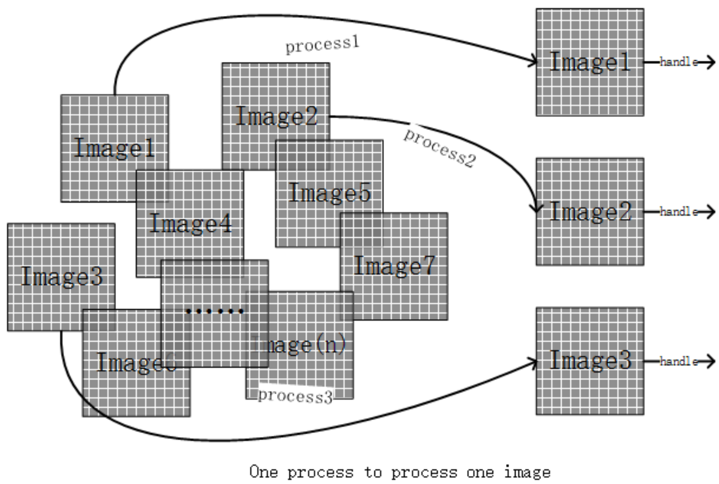

Figure I. Example of task partition when image data is small.

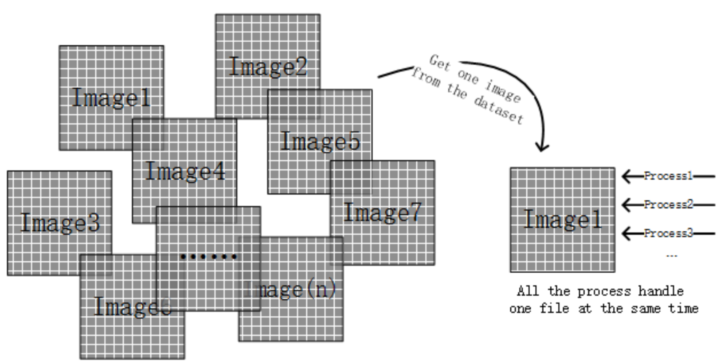

Figure II. Example of task partition when image data is very large.

The algorithms are tested in the same high performance cluster, communication between nodes through the Gigabit network, parallel file system using IBM GPFS (General parallel File System). The Figure III shows the efficient of ParallelMosaic with the number of processes in dealing with Dataset1. Comparing with GDAL mosaic tool and ArcGIS in the process number of 16, the result shows that our parallel mosaic algorithm has a great improvement in performance and scalability (Figure IV). According to these experiments, the factors influencing parallel performance on a cluster are discussed.

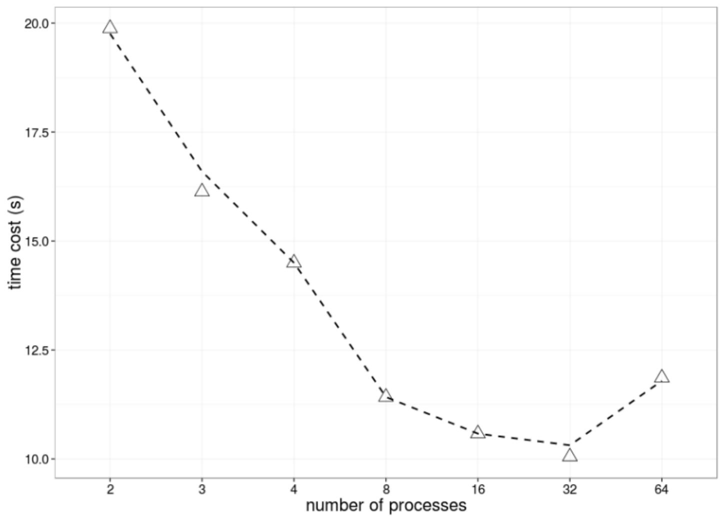

Figure III.   Time cost with different of number of processes

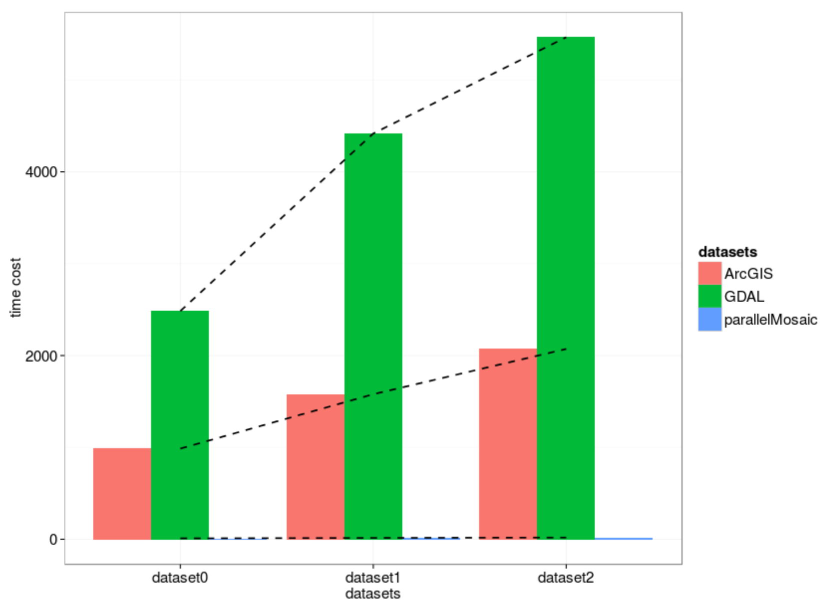

Figure IV.   Performance compared between ArcGIS, GDAL, ParallelMosaic.

This paper provides a method on the image mosaicking parallelization for large scale application based on cluster system. We firstly introduce the motivation of image mosaicking parallel for large scale application, then explain the programming with multiple-step procedure and the parallel I/O operation. So far parallelization of image mosaicking for large scale application is still on its early stage, perspective for future work may try to realize the parallel image mosaic algorithm using spark distributed cluster system, compare to MPI, finally get the performance analysis results.

Acknowledgements

This study is supported in part through HTRDP(863) grant(2015AA123901).

References:

Chen, L., Ma, Y., Liu, P., Wei, J., Jie, W., & He, J. (2015). A review of parallel computing for large-scale remote sensing image mosaicking. Cluster Computing, 18(2), 517-529.

Yang, J., & Zhang, J. (2015). Parallel performance of typical algorithms in remote sensing-based mapping on a multi-core computer. Photogrammetric Engineering & Remote Sensing, 81(5), 373-385.

Chen, C., Chen, Z., Li, M., Liu, Y., Cheng, L., & Ren, Y. (2014). Parallel relative radiometric normalisation for remote sensing image mosaics. Computers & Geosciences, 73(C), 28-36.

Benkelman, C. (2004). Digital aerial image mosaic method and apparatus. US, US6694064.

Wang, Y., Ma, Y., Liu, P., Liu, D., & Xie, J. (2010). An Optimized Image Mosaic Algorithm with Parallel IO and Dynamic Grouped Parallel Strategy Based on Minimal Spanning Tree. International Conference on Grid & Cooperative Computing (pp.501-506).

Chen, C., Tan, Y. H., Hai-Tao, L. I., & Hai-Yan, G. U. (2011). A fast and automatic parallel algorithm of remote sensing image mosaic. Microelectronics & Computer, 28(3), 59-62.

Berriman, G. B., Good, J. C., Laity, A. C., Katz, D. S., Jacob, J. C., & Brieger, L., et al. (2007). An all-sky 2mass mosaic constructed on the teragrid: processing steps for generation of a 20-terabyte 2mass all-sky mosaic. Highlights of Astronomy, 14, 625.