

DESIGN AND APPLICATIONS OF RAPID IMAGE TILE PRODUCING SOFTWARE BASED ON MOSAIC DATASET

ZHA Zhuhua^{a,*}, HUANG Wei^a, WANG Cong^a, TANG Dejin^a, ZHU Lulu^b

^aNGCC, No.28 Lianhuachi West Road, Haidian District, Beijing, (zhazh, huangwei, wangcong, tangdejin)@ngcc.cn

^b ESRI China Information Technology Co. Ltd, No.1 Xibahe South Road, Beijing, zhull@esrichina.com.cn

Commission III, WG IV/3

KEY WORDS: Image Tile, Mosaic Dataset, Cluster, Tile Producing, Quality Checking

ABSTRACT:

Map tile technology is widely used in web geographic information services. How to efficiently produce map tiles is key technology for rapid service of images on web. In this paper, a rapid producing software for image tile data based on mosaic dataset is designed, meanwhile, the flow of tile producing is given. Key technologies such as cluster processing, map representation, tile checking, tile conversion and compression in memory are discussed. Accomplished by software development and tested by actual image data, the results show that this software has a high degree of automation, would be able to effectively reducing the number of IO and improve the tile producing efficiency. Moreover, the manual operations would be reduced significantly.

1. INTRODUCTION

At present, the map tile technology is widely used in web geographic information services, such as Internet maps. It cuts the original image into tiles with the same size according to certain rules, ensures efficient transmission and display under limited bandwidth by loading a small amount of image tiles. Therefore, how to generate tiles efficiently is the key technology to get fast service of image on web.

There are many kinds of software can provide the image to tile, but there are still some problems exists, mainly including:

- 1) With the rapid development of satellite sensor technology and UAV technology, the space and temporal resolution of remote sensing images have been greatly increased, the data volume has been rapidly increasing, therefore the amount of data that needs to be tiled is getting larger and larger.
- 2) Because of different data sources, spatial resolutions and data quality, software needs the capability to identify and be compatible with a lot of different image types.
- 3) The tile quality always has problems in mass production, so the quality of production needs to be improved. When the amount of data is very large, such as more than 10 million square kilometers of coverage each year, tiles of some area are often lost because of software, network and so on. Usually, these problems need to be examined manually, and the software cannot check the tiles automatically.
- 4) Productivity needs to be improved. More production mode need to be supported by the software such as the cluster technology, using more computing resources, improving concurrency efficiency and reducing processing time for single task.
- 5) The tile generating process has lot of steps operations are complicated and need large human interaction. The tile generating function of common software is relatively single, and the subsequent processing is not well supported, such as the adjustment of the tile name, watermark and the fusion of tiles. Usually, it needs some small tools to deal with the subsequent

processing and these tools need to be started manually. So the waiting time of machine happens and delays the producing schedule. These tools also make it harder for people to operate. Therefore, the existing methods and software of tile producing cannot meet the needs. We need to improve the producing capacity of image tiles, the existing tile generating process, the quality and efficiency of the image tiles, and adapt to the rapid publishing of the latest image.

2. SOFTWARE DESIGN

2.1 Background

The web geospatial services publishing require the use of large-scale image data updated from various sources regularly or irregularly. In order to effectively improve the utilization of data and enhance the producing capabilities of map tiles, it is necessary to improve and upgrade the existing tile data producing process, improve the quality of tile data and efficiency of producing.

The main deficiencies of the current process are:

- 1) The current process of tile producing is relatively complicated, which makes it more difficult to operate.
- 2) With involving lots of independent tools it is more difficult for operators, and it takes more time for waiting.
- 3) The efficiency of tile producing needs to be improved.
- 4) The quality of tile producing needs to be improved.

2.2 Design Goal

The goal of this software is to strengthen the back-end data producing and management capabilities, improve the overall efficiency from data collection to service publishing, accomplish rapid data integration, automated data processing, and rapid service publishing. The specific design goals for this software's functional module are: simplifying the operation

* National Key R&D Program of China(NO. 2017YFB0503700).
zhazh@nsdi.gov.cn

process, improving work efficiency, guaranteeing data quality and making it convenient to promote the using.

2.3 Design Solution

Based on the ArcGIS software technology, this paper designs a technology architecture suitable for large-scale image tile data generating. Based on this architecture, we develop customized software, integrate multiple generating links, and achieve high efficiency of multi-source and different image data with high-quality and stable performance. The architecture uses a three-tier model, e.g. Figure 1: the data layer, the service layer and the business layer. The data layer uses ArcGIS mosaic data set technology to finish the identification, management and sharing of multi-source image data. The service layer uses ArcGIS server image processing service, map service and image service to publish the image data into service. It uses cluster technology for invocation by business layer and tunes the number of process of tile generating with the number of CPU cores. The business layer uses mosaic data set to achieve data collection, storage and update. Moreover with using ArcGIS cartography for tile caching and other technologies to achieve tile producing, using memory conversion technology to achieve the tile format conversion and data compression, the tiles and quality reports would be finally completed.

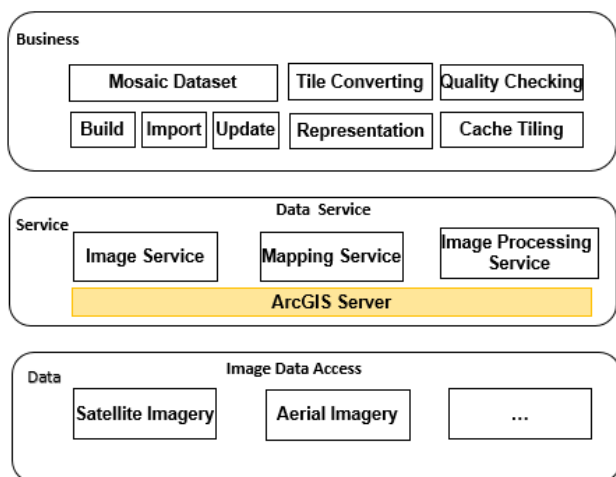


Figure 1. Software architecture

The designed data processing flow is shown in Figure 2. Among these, Check Tile 1 checks whether the tiles produced by ArcGIS Server are consistent with the original image coverage, while Check Tile 2 checks whether the number of final converted tiles are same with the number of tiles in the bundle and confirm whether the conversion process is complete.

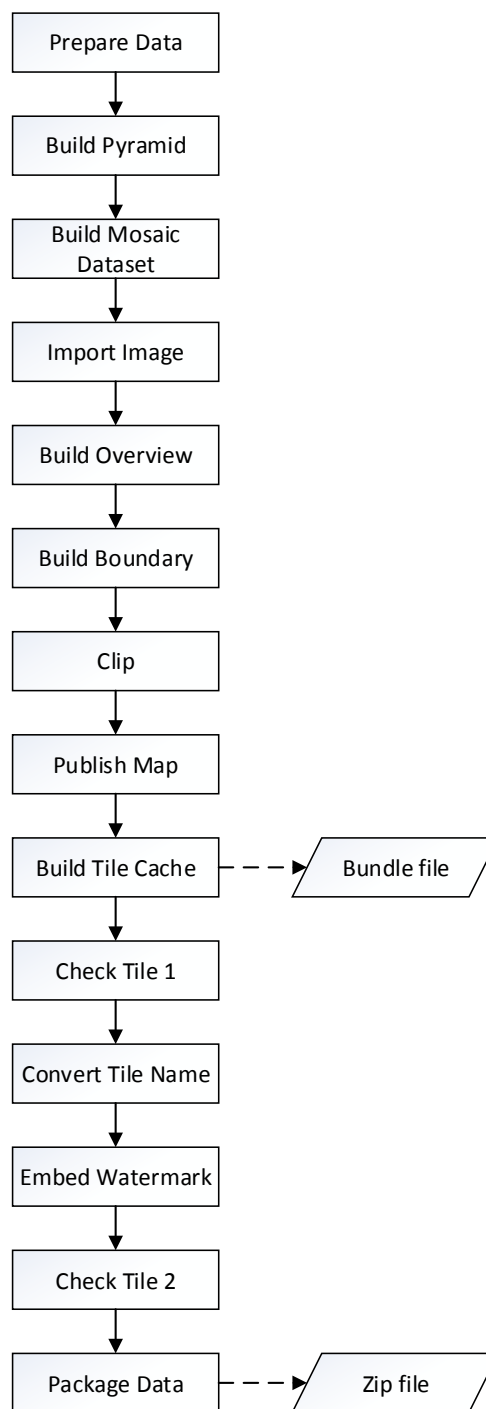


Figure 2. Processing flow

3. KEY TECHNOLOGY

3.1 Mosaic Dataset

The mosaic dataset can be used to solve the problem in identification, management and sharing large-scale multi-source image data. Mosaic dataset is ArcGIS 10's new imaging technology for managing raster data. It is a hybrid technology that combines raster datasets with catalogs and manages raster data in the way which is consistent with unmanaged raster catalogs. Therefore, it can index the data set, and queries can be performed on the set. Its storage method is similar to raster catalog and same with regular raster dataset when using. Mosaic

datasets are used to manage and publish massively multi-resolution, multi-sensor images. The datasets provide dynamic mosaic and real-time processing of raster data. Its greatest advantage is the advanced raster query function and real-time processing functions and can be used as a source for providing image services. In the practical process, a mosaic dataset allows addition of all types of raster data is mainly used. In this readable and writable dataset, any type of raster data can be added, modify the properties or functions that are applied to each raster or mosaic dataset is allowed, the pyramid can be constructed, the size of the pixel can be calculated, and so on.

3.2 Map Representation

While adding the map representation, the watermark would be directly generated when the image is tiled.

With this technology, the cached tiles are directly provided with watermark, which eliminates the secondary reading for tile data specifically for watermarking, reduces IO, and improves efficiency.

3.3 Cluster Technology

ArcGIS cluster processing technology can be used to achieve the efficient generating of tiles. The ArcGIS Server clustering can break down slice caching tasks, dramatically increasing cache efficiency and improving machine utilization.

3.4 Tile Quality Checking

When producing tiles through ArcGIS Server, lost or blank tiles are often happens. The current solution is caching again with problem areas after finding out the problem manually. It can be done in small-scale image data. On the contrary, it cannot be completed when the errors happened in large-scale image data, such as a whole province.

Therefore, how to check the loss of the tiles and how to control the time of tile checking becomes very important for ensuring the quality of the image tiles. Finally, we find out the way of using ArcGIS bundle files to check tile. Multi-process technology is used to achieve a high-speed tile checking, and then we get the bound of the problem tiles. After that, we can reproduce the tiles. The key technology is to be able to quickly read and write bundle files.

We can use the characteristics of bundle files to perform grading checks and hence the efficiency would be improved. The rule for grade division is:

- 1) A complete Bundle file as the first grade.
- 2) The 16 blocks (4 rows and 4 columns) divided by the bundle file as the second grade.
- 3) The 1024 tiles included in one block as the third grade.

To sum up, a complete bundle file should contains 16384 tiles. The checking steps are:

3.4.1 First Grade Checking: it is whole bundle checking.

The Bundle file name is used to calculate the starting row/column number and ending row/column number of the file, so it can calculate the theoretical coverage of the file, and use this coverage and the tile producing coverage to perform spatial query. If the tile producing coverage completely includes the theoretical coverage, and the number of file tiles is 16,384, the Bundle file is qualified. Then, we check the number of converted tiles. If the number is 16,384, the conversion process qualified. If not, we run the next step.

3.4.2 Second Grade Checking: it is block checking and tile checking.

We can check the 16 blocks in the bundle one by one, if qualified, check the next block. If not, check the specific tile.

3.5 Tile Conversion and Compression in Memory

With this technology, you can avoid generating loose tiles, saving disk space, avoiding the problems after removing loose tiles, and also facilitating the maintenance of the hard disk.

The number of converted tiles from the Bundle is 16384, if the tiles are output into picture one by one, there are 16384 reads and writes IO. So the pressure on the IO is large. To reduce IO, the memory compression technology can be used to create picture path index for 16384 tiles, then the picture path divided into 16 packages, and the pictures compressed into corresponding zip archives directly in the memory to avoid loosening tiles, reduce disk IO and increase conversion efficiency.

4. EXPERIMENT

4.1 Environment

Based on the design of this paper, we completed the related software development.

The experimental method is by using two machines with same hardware and software to complete the same image tile generating task, then comparing the time-consuming to verify the effectiveness of this paper.

The experimental data is selected high-resolution image data from three provinces in North China, covering an area of 268,200 square kilometres, with resolution of 0.8 meters and the data volume is 1.5TB. According to the standard titled with "Platform for geoinformation common services data specification for electronic data", we need to produce the 17th and 18th levels tile data, and the format is JPG.

The experiment hardware and software environment are shown in Table 1.

Item	Configuration
Server	Dell FX2 (CPU:4*Xeon E5-4610v3, Memory:128GB, Hard drive:7.2TB)
OS	Windows Server 2008
runtime	ArcGIS Server 10.2.2

Table 1. Experiment environment

The experiment compares the time-consuming and manual steps by the integrated software developed in this software and the existing original flow. The time starts with the task and ends at the completion of the task, including the time for the computer waiting for the next manual operation.

4.2 Result

The experiment result is: the time consuming by original flow and waiting for operation is 342 hours, and the time consuming by software designed in this paper is 73.5 hours, the efficiency is 4.6 times faster as the original method, e.g. Table 2.

Item	Original flow	Designed Software
Time consuming	342 hours	73.5 hours
Manual steps	21	1

Table 2. Experiment result

The decrease in time is mainly due to lots of dependent tools in the original flow. Every next step needs to be manual operated for starting, so it wastes lots of time for waiting.

5. CONCLUSION

Through the software designed in this paper, the existing operation flow is improved, all the manual work between the links are omitted, the manual workload is greatly reduced, and the efficiency is also improved by means of clustering. Making use of the automatic checking, the data quality is guaranteed and the rework rate is further reduced. Therefore, the overall efficiency and quality of the tile producing are effectively improved. The software greatly reduces manual operation,, producing reliable tile quality with high degree of automation, and it is more conducive to promotion and using.

REFERENCES

DONG Jie, HUANG Yan, DONG Ping, 2012, Research on Management of Large Quantities of Image Data Based on Mosaic Dataset, *Infrared*, 33(10),pp:30-32.

LIU Shiyong, CHEN Luo, et. 2018, Surveying and Mapping, *Computer Engineering and Applications*, 54(1), pp:48-53

ArcGIS Resources, 2014,What is a mosaic dataset, <http://resources.arcgis.com/en/help/main/10.1/index.html#//009t0000003700000>

YU Zongqiu, LIU Xiaochun, 2017, Production and Application of ArcGIS Vector Tile Map in Zhuhai, *Surveying and Mapping*, 40(3),pp:122-124,

LIU Yi, CHEN Luo, JING Ning, XIONG Wei, 2013, Parallel Batch-Building Remote Sensing Image Tile Pyramid with MapReduce, *Geomatics and Information Science of Wuhan University*, 38(3), pp:278-282