

Parallel Compression Based on Prediction Algorithm of Hyper-spectral Imagery

Wenbin Wu^{1,*}, Yue Wu¹, and Xu Qiao²

¹Beijing University of Posts and Telecommunications, Beijing, China

²School of mechanical electronic & information engineering, China University of Mining and Technology (Beijing), Beijing, China

Abstract. Along with the development of the spectral imaging technology, the precision of the hyper-spectral imagery becomes very high, and the size of the hyper-spectral imagery becomes very large. In order to solve the problem of the transmission and the storage, it is necessary to research the compression algorithm. The traditional prediction algorithm is adopted in the serial processing mode, and the processing time is long. In this paper, we improve the efficiency of the parallel prediction compression algorithm, to meet the needs of the rapid compression. We select bands along the direction of spectral or the direction of space, so that the hyper-spectral imagery can be divided into sub images. We number the sub images, then send them to different processing units. Each unit does compression tasks at the same time. This paper also compares the relationship between the processing unit number and the compression time. The experiment shows that, the parallel predictive compression algorithm can improve the efficiency of compression effectively.

1 INTRODUCTION

THE parallel is a kind of way to improve the processing efficiency, and it uses multiple processing units to deal with the problem. Parallel means that can be calculated or operated at the same time. The parallel computing needs parallel data processors, which can divide an application into multiple sub tasks. They are sent to different processors, and processors work together to accomplish tasks. That will speed up the computation, or that will expand the size of problems.

The hyper-spectral imagery is a three-dimensional image taken by the imaging spectrometer. Because the large size of the hyper-spectral imagery, the traditional compression method is hard to meet the requirements of the high-speed encoding and decoding. Research on the parallel compression technology is not only to satisfy the demands of the engineering development, but also to resolve practical problems.

Traditional methods use serial processing mode to do the hyper-spectral imagery compression. According to the spectral correlation, we calculate the prediction coefficients. Then get the prediction residual based on the difference between the restructured imagery and the fore imagery. Finally, we encode the prediction residual.

The whole work is company with the complicated procedure, the long computing time, and the low efficiency. This paper will segment a hyper-spectral imagery, and break up the whole work into parts. Thus we will achieve the goal of the parallel compression of the hyper-spectral imagery.

2 Parallel Computing Theory

In 1972, Professor Michael J. Flynn proposed the Flynn classification method, and the classification method has far-reaching influence in developing the parallel processing technology. According to the classification, computer system is consists of the instruction streams and data streams. A computer process can be regarded as a series of the instructions dealing with a series of the information. The emphases of the classification is the diversity of the hardware, which manages instruction streams and data streams. Among them, multiple instruction multiple data (MIMD) computer can execute the multiple instruction streams simultaneously, and the instruction streams can operate different data streams separately.

In recent years, relying on the rapid development of parallel processors, MIMD computers are becoming more and more popular, such as multi core processor from Intel and AMD, the British ARM company and its partner's ARM series processor etc... The work of this paper will be based on the Intel Core i7 central processing unit computer.

Parallel computing is consists of the parallel method between time and space. The parallel of time refers to the computer pipeline, and the parallel of space refers to calculation using multiple processors concurrently. Parallel computing research in science is a parallel problem of space. On 2004 October, MathWorks released MATLAB parallel computing toolbox. Parallel programming interface based on the famous Message

* Wenbin Wu: zxjun@cumtb.edu.cn

Passing Interface (MPI), and the interface standard is discussing and passing by Message Passing Interface Forum. The standardization of MPI involves people in about 60 countries, including most of the major manufacturers of parallel computers, and the researchers from the university, government, laboratories and companies. In May 1994 the MPI standard was released. With the establishment of the MPI standard, the open source MPICH was developed by engineers from the Argonne National Laboratory and the Mississippi State University, and it has good portability. Another non open source version of MPI is Intel MPI, and Intel MPI can make full use of multi-core, multiprocessor and network communication at different levels.

MATLAB packaging MPI into parallel computing toolbox. To the use of the parallel computing toolbox, people do not need to consider the detail of MPI. On the one hand, it is convenient for the user to transplant an existing MATLAB program, on the other hand, advanced user can also develop further program through MPI interface. Through continuous improvement, the parallel computing toolbox can solve the problem of complex computation and the large size of data processing on multi-core and multiprocessor computer.

3 Hyper-spectral Imagery

The hyper-spectral remote sensing technology is a field of geophysical methods in advanced. In recent years, with the development of semiconductor technology, the hyper-spectral imagery caught by imaging spectrometers, as shown in Fig. 1, expanded in the scale of spatial resolution and spectral resolution. Thus, the data size expands constantly, and one scene data is from hundreds of MB to several GB or even a dozen of GB. It makes a lot of challenges to manage and transmit data.

In order to meet the needs of hyper-spectral imagery data management and transmission, hyper-spectral imagery data compression as an effective way has got a wide range of applications. The research of the compression of hyper-spectral imagery has made some progress, but there is still some room for improvement. The main deficiencies is that the better effect it compresses, the longer time it often takes. Therefore, how to improve the compression efficiency is an important part of engineering application. Development of parallel computing provides a useful way to improve the efficiency of hyper-spectral imagery compression. This paper will focus on the technology and methods of hyper-spectral imagery parallel compression, and gives a parallel algorithm for hyper-spectral imagery compression based on prediction.

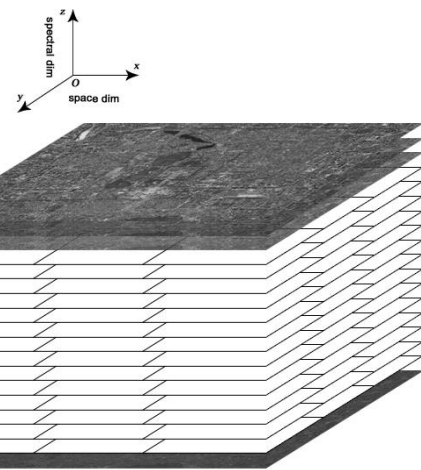


Fig. 1. The structure of hyperspectral image.

4 Parallel Prediction Algorithm

In early twentieth Century, with successive efforts of Kolmogorov, Turing, Shannon et al., probability theory had made great achievements in source coding. The branch of probability theory has cleared the obstacle for future research on data compression. Linear prediction ideas appeared in the nineteenth Century. In 1966, S. Saito and F. Itakura first proposed the linear predictive coding. They put forward a method for automatic phoneme recognition, and this method uses the maximum likelihood estimation to code the speech. Prediction method is one of the oldest and most effective compression method, which can be used to remove the correlation of hyper-spectral image. It receives errors by coding the restructured values and fore values to achieve compression. Differential pulse code modulation (DPCM) and its improved algorithm are used constantly. Through the use of parallel computing method presented in this paper, we propose the two order linear parallel prediction algorithm.

The traditional two order linear predictive compression is presented in Fig. 2. First of all, we use the band 1 and band 2 to predict band 3, similarly, then we use the band 2 and band 3 to predict band 4.

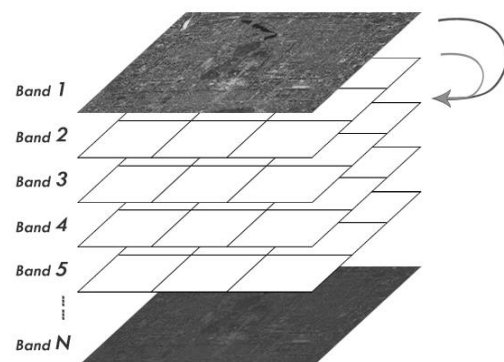


Fig. 2. Hyperspectral image prediction compression sequence.

Because the traditional method needs to predict bands from A to Z, with the increasing data of hyper-

spectral imagery, the time required often cannot meet the needs of rapid processing. We improve the algorithm, and group the bands. As an example of Fig. 3 shows, 5 bands are in one group. Without considering the correlation between the groups, we use the parallel compression. Thus, parallel computing can make a rapid compression.

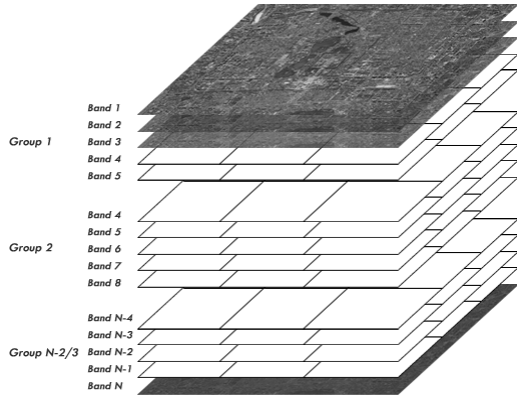


Fig. 3. Parallel compression of hyperspectral image.

Two order linear predictive compression algorithm takes advantage of spectral correlation of hyper-spectral imagery, and compression results are well in the spectral dimension. The prediction algorithm considers the minimum mean square error, and the function of two-dimensional mean square error is as follows:

$$E^2 = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (g(i, j) - a \cdot f_1(i, j) - b \cdot f_2(i, j))^2 \quad (1)$$

In the function $g(i, j)$ is the elements of restructured bands, and $f_1(i, j)$, $f_2(i, j)$ is the element of fore bands. When square error is minimal, it satisfies function 2.

$$\begin{cases} \frac{\partial E^2}{\partial a} = 0 \\ \frac{\partial E^2}{\partial b} = 0 \end{cases} \quad (2)$$

From function 2, we can calculate a and b values:

$$\begin{cases} a = \frac{r(g, f_2)r(f_1, f_2) - r(g, f_1)r(f_2, f_2)}{r(f_1, f_2)r(f_1, f_2) - r(f_1, f_1)r(f_2, f_2)} \\ b = \frac{r(g, f_1)r(f_1, f_2) - r(g, f_2)r(f_1, f_1)}{r(f_1, f_2)r(f_1, f_2) - r(f_1, f_1)r(f_2, f_2)} \end{cases} \quad (3)$$

Where,

$$r(f, g) = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} f(i, j)g(i, j) \quad (4)$$

From the function 3, we can calculate the corresponding coefficients a and b . Through using prediction coefficients to predict hyper-spectral imagery, we calculate a residual matrix by the restructured imagery and fore imagery. Then we use entropy coding to compress the residual matrix, to achieve two order linear prediction compression of hyper-spectral imagery. The parallel predictive compression process can refer to Fig. 4. We have grouped the hyper-spectral image, then we compress each group. We get the residual matrix and

the coefficient data, and we compress the residual matrix parallel. Because each group of the data is less than the fore data, the parallel compression can increase the efficiency of the traditional predictive compression.

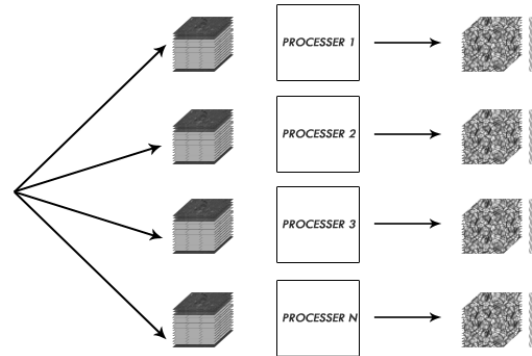


Fig. 4. Parallel compression process of hyperspectral image.

5 Experimental Results and Analysis

Through the two order linear parallel prediction compression algorithm, we can speed up the efficiency of the hyper-spectral imagery compression. In the experiment, we use Hyperion hyper-spectral imagery, as shown in table 1.

Table 1. HYPERION L1 DATA DESCRIPTION.

Data Name	Hyperion L1
Wavelength	356 ~ 2577 (nm)
Bands Number	242
Image Element	30 (m)
Image Size	256×6460
VNIR Band	1~70 (356 ~ 1058 nm)
SWIR Band	71~242 (852 ~ 2577 nm)
Data Type	2 Bytes
Image Format	BIL
Byte Order	Network (IEEE)
File Size	800,427,520 (bytes)

In the parallel predictive compression, the data were grouped according to the bands. Each group has $N/n+2$ elements, where N is the total number of bands and n is the number of processing units. Then the data in each group is sent to each processing unit. The process is shown in Fig. 4.

We will compare two order linear prediction methods and parallel prediction methods. From table 2, it can be seen. The two order parallel predictive compression

algorithm reduce the time of the compression significantly, without changing the other compression effects such as compression ratio.

Table 2. VELOCITY ESTIMATION VALUES.

Clustering Time	2 Order Prediction (s/scene)	Parallel 2 Order Prediction (s/scene)
1 Processer	29.2222	29.2222
2 Processers	31.3096	16.8164
3 Processers	35.9879	14.5569
4 Processers	51.1540	13.0479

6 Conclusion

From the results of the above analysis and comparison, the parallel predictive compression algorithm can make full use of system resources, and it can improve the efficiency of the compression. In practical engineering, it delivers a good application prospect.

7 Acknowledgements

This work was supported by China geological survey geological survey project: Remote sensing sensor optimization technology for mineral and energy exploration. (Grant No: 1212011120221) And it was supported by the National Science & Technology Pillar Program during the 12th Five-year Plan Period. (Grant No: 2011BAD04B05)

The authors are in school of mechanical electronic & information engineering, China University of Mining and Technology (Beijing), Beijing 100083, China (e-mail: zxjun@cumtb.edu.cn).

References

- Ingram, R. N., et al., International Journal of Remote Sensing 25(22), *An automatic nonlinear correlation approach for processing of hyperspectral images*. 4981-4998. (2004)
- Du, Q., et al., IEEE Geoscience and Remote Sensing Letters 6(4), *Segmented Principal Component Analysis for Parallel Compression of Hyperspectral Imagery*. 713-717. (2009)
- Plaza, A., et al., Remote Sensing of Environment 113, *Recent advances in techniques for hyperspectral image processing*. S110-S122. (2009)
- Plaza, A., et al., Ieee Signal Processing Magazine 28(3), *Parallel Hyperspectral Image and Signal Processing*. (2011)
- Zhao, X. and X. Qiao, Spectral Characteristics Research of the Hyperspectral Image Based on the Correlation Matrix. Information Science and

Engineering (ISISE), 2012 International Symposium on. (2012)

- Santos, L., et al. Journal of Applied Remote Sensing 7, *Lossy hyperspectral image compression on a graphics processing unit: parallelization strategy and performance evaluation*. (2013)
- Taher, A., et al., Hyperspectral image segmentation using a cooperative nonparametric approach. Image and Signal Processing for Remote Sensing Xix. L. Bruzzone. 8892. (2013)
- Santos, L., et al., Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing 6(2), *Highly-Parallel GPU Architecture for Lossy Hyperspectral Image Compression*. 670-681. (2013)