Research and Implementation of FPGA Image Compression Parallel Algorithm Based on CCSDS

Leqi Yang

Chongqing University Of Posts And Telecommunications, Chongqing, 400065, China

1225789332@qq.com

Abstract. High-pixel, high-resolution spectral images are an important source of key graphical information obtained from space remote sensing, however, with the dramatically expanding amount of graphical information, the limited transmission bandwidth and hardware resources are difficult to meet the demand for massive data, which makes real-time image lossless compression and transmission technology an important solution. For this reason, this paper is based on the CCSDS image lossless compression algorithm, using FPGA for the construction of image compression system, and proposes a method strategy based on parallel acceleration algorithm to realize the real time processing of large-capacity hyperspectral images. The simulation results show that hyperspectral river and mountain range images can be compressed with greatly improved compression efficiency while maintaining real time performance and accuracy.

Keywords: Hyperspectral images, Lossless compression, CCSDS, FPGA, Parallel acceleration algorithm.

1. Introduction

With the development of modern space technology, satellite-based image processing technology has been widely used in land use, monitoring environmental changes, planning urban development, forest fire prevention and geological survey. The spatial and spectral resolution of hyperspectral images is constantly increasing, and the data volume of images is expanding, bringing enormous pressure on channel transmission and image storage and requiring the compression of hyperspectral images. The ability of Earth-observing space vehicles, such as satellites, spacecraft and space stations, to transmit data from space to the ground in real time is one of the measures of success or failure of these space programs. However, the data volume of on-board images is generally large and the channel transmission bandwidth is limited, so how to compress the data efficiently and transmit it in real time is particularly important [1].

Although lossy compression of images saves more transmission bandwidth, lossy compression does not recover the original image 1:1 and may cause loss of useful information. This is unacceptable for hyperspectral image data which are difficult to acquire [2]. Hyperspectral images require high image quality and need to maintain the original spectral information as distortion-free as possible, so lossless compression becomes the preferred solution. Among the current hyperspectral image compression methods, the CCSDS 123.0-B-1 algorithm is the international standard for lossless compression of multi/hyperspectral images developed by the CCSDS organization. At this stage, the real-time hyperspectral image compression system is almost exclusively based on hardware implementation, so this paper is a comprehensive consideration of the performance of the algorithm and the development difficulty of hardware implementation and other factors to select the lossless compression algorithm [3].

For the lossless compression algorithm of hyperspectral images, many scholars at home and abroad have carried out a lot of research work. At present, the compression techniques of hyperspectral images can be divided into three kinds, namely, prediction-based [4][5][6], transform-based [7][8] and vector quantization-based [9][10][11]. The prediction-based compression technique is based on the similarity relationship between the pixel and the pixel at the adjacent position, using the known values in the neighborhood of the pixel to infer the current pixel value, which reduces the information entropy of the original image data, and achieves the compression effect. In 2019, Jiqiang Luo and Jiaji Wu et al. proposed the C-DPCM-RNN (C-DPCM-Recurrent Neural Networks) method, which replaces the linear predictor in C-DPCM with a recurrent neural network RNN, which achieves the highest compression ratio currently in the field on the corrected AVIRIS open hyperspectral test set. Transform-based compression techniques map the original data into the transform domain through a transform algorithm, which requires that the entropy of the transformed data is reduced and there is less redundant information.2019 Beijing Jiaotong University, Beijing, ChinaLijing Wu, Lijing, based on the MAlTLAB on the integer wavelet transform algorithm to study the method reduces the entropy of the information of the sun's magnetic field image and improves the limiting compression ratio After simulation and validation, the method can meet the sun's magnetic field image's Simulation and verification, can meet the demand of lossless compression of solar magnetic field image, compression ratio and compression time meet the actual requirements [12]. The lossless compression technique of Vector Quantization (VQ) represents all the pixel values under different spectral bands at any spatial position of the image as a pixel vector and quantizes the data directly. Bingyu Qu gives a code word search algorithm based on mean, variance and triangular inequality by analyzing the search range determined by two characteristic quantities, mean and variance. The experimental results show that the algorithm can effectively exclude the code words that cannot be excluded by the exclusion criterion determined by the two characteristic quantities of mean and variance, and the computation required for coding is low [13].

With the dramatic expansion of the amount of graphic information, the limited transmission bandwidth and hardware resources limit the high quality of image information, how to minimize the cost of time, the minimum space overhead and the lowest information loss is the key to image processing, the current lack of real-time high-speed lossless compression of hyperspectral image compression method.

Based on the above, this paper proposes a FPGA parallel acceleration algorithm model based on CCSDS lossless compression algorithm for the realization and realtime transmission of hyperspectral image lossless compression algorithm. Firstly, the principle of hyperspectral image compression is analyzed, and based on the mathematical principle and physical architecture of lossless compression, C++ and Verilog programming language are used to realize the construction of compression system encoder and predictor as well as the parallel characteristics of the CCSDS 123.0-B-1 algorithm, and based on which, a parallel acceleration algorithm is proposed to realize real-time lossless compression and processing of hyperspectral images under the framework of FPGA. processing, which can greatly improve the compression efficiency while maintaining real-time and accuracy.

2. Algorithms And Fpga Fundamentals

2.1 CCSDS image lossless compression algorithm

2.1.1 CCSDS Basic Concepts

CCSDS (Consultative Committee for Space Data Systems), as the international consultative committee for aerospace data systems, is responsible for the development and promotion of data processing standards for use in aerospace missions. Image compression is a key technology in aerospace missions because the amount of data acquired by spacecraft is usually very large and must be transmitted and processed with limited bandwidth and storage resources. CCSDS defines a series of image compression standards, such as CCSDS 121.0-B-2, CCSDS 122.0-B-1, CCSDS 123.0-B-1, etc., which are designed to provide efficient image compression algorithms for space missions to provide efficient image compression algorithms to meet transmission and storage requirements. CCSDS 123.0-B-1 was completed in May 2015 by the CCSDS group to establish a recommended standard for lossless compression algorithms for 3D multi/hyperspectral image data and to indicate the compressed data format in order to reduce the pressure on the transmission channel bandwidth, to reduce the need for caching and storage as well as reduce data transmission time at a given rate, the standard is applicable to data compression applications for packetized remote sensing cross-supported space missions [3]. Both the correlation in the spatial dimension of the image data and the correlation in the spectral dimension are considered. So the compression performance of the compression algorithm is better than the traditional compression algorithm [14-15].

The biggest difference between the CCSDS lossless compression algorithm and the traditional lossless compression algorithm is that it makes full use of the property that the pixels of adjacent spectral segments of hyperspectral images are correlated, and when calculating the prediction value, it uses the pixels of the current spectral segment as well as those of the adjacent spectral segments. The data compression is realized by adaptive Rice coding after the prediction value is obtained [16,17]. The algorithm consists of two parts, the predictor and the encoder. The role of the predictor is to remove the correlation in the spatial dimension and spectral segment dimension of the hyperspectral image data, and the role of the encoder is to compress and encode the de-correlated hyperspectral image data, and the commonly used encoding method is Golomb Rice coding.

2.1.2 Fundamentals of the CCSDS lossless compression algorithm

The CCSDS 123.0-B-1 algorithm consists of two parts, the predictor and the encoder, where the input data is an image (a three-dimensional array with integer pixel values), and the output is a compressed codestream of the input image, from which the input image can be accurately recovered from this bitstream.The CCSDS 123.0-B-1 algorithm is based on the Fast Lossless (FL) [18,19] compression algorithm.The FL algorithm makes a prediction of the current pixel value based on the values of adjacent encoded pixels in a smaller neighborhood of the current pixel point.The computation process uses the pixel values of the current spectral segment and of several (usually 3) previous spectral segments. The prediction process uses an adaptive linear prediction algorithm and the prediction weights are updated each time, and the entropy coding stage performs lossless compression of the mapped prediction residuals, which is performed by scanning the image pixels sequentially. Three sequences of imaging hyperspectral image data are shown in Fig.

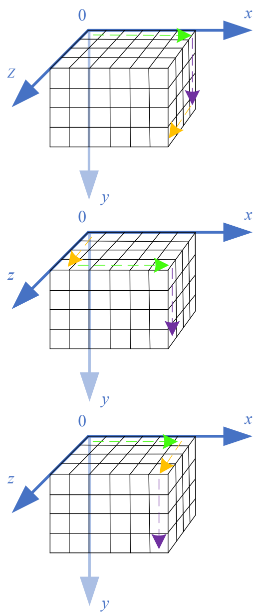

Figure 1. Imaging sequence for hyperspectral image data

Prediction is a modeling operation on the current sample values based on the sample values at previous moments [20]. To ensure lossless compression, the prediction must be reversible. Figure 2 represents the 3D prediction model. Where z denotes the band, x denotes the image width and y denotes the image height. The predictor first calculates the local sum of the point to be predicted Sz，y，x from which it calculates the local difference and direction local difference to obtain the local difference vector of the current spectral segment. Then using the local difference vectors of the corresponding pixel points of the previous p spectral segments and the weight vectors are linearly combined to obtain the predicted sampling value and the scaled sampling prediction value [21]. For each point that is predicted, the current weight vector is updated in order to compute the predicted value for the next point. Finally the prediction residuals are mapped to unsigned types by a mapping function.

The workflow of the predictor is as follows: first, local sums of the surrounding sampled values are calculated from the image data of the current image. Then, local differences of the peripheral sampled values are calculated using these local sums. Next, the peripheral local differences of the current spectral segment are combined with the central local differences of other spectral segments to form a difference vector. The predicted value of the center local difference is obtained by performing an inner product of the difference vector and the weight vector. This predicted value is then used to calculate the predicted value of the current sampled value. Finally, the mapping result of the predicted values is calculated. The specific formulas of local sum σx,y,z , local difference dx,y,z are pushed to the process can be found in the paper [22], which will not be repeated in this paper, the formulas are as follows:

$$U_z(t)=\begin{bmatrix} d_z^N(t) \\ d_z^W(t) \\ d_z^{NW}(t) \\ d_{z-1}(t) \\ d_{z-2}(t) \\ \vdots \\ d_{z-P_z^*}(t) \end{bmatrix},\quad W_z(t)=\begin{bmatrix} w_z^N(t) \\ w_z^W(t) \\ w_z^{NW}(t) \\ w_z^{(1)}(t) \\ w_z^{(2)}(t) \\ \vdots \\ w_z^{(N)}(t) \end{bmatrix} \tag{1}$$

The prediction of information is achieved when the predicted value of the central localized difference is equal to the inner product of the weight vector and the difference vector at when $t > 0$.

The focus of the CCSDS algorithm to realize adaptive coding is the update of the weight vector. After the prediction of each pixel point to be predicted, the individual weight values of the weight vector are updated once, thus ensuring that the prediction of the next pixel point to be predicted can be more accurate. The update is based on the difference between the gray value of the pixel to be predicted and the predicted value. The entropy coding in this algorithm refers to the length-constrained Golomb coding, which has better coding effect on the non-negative integer signals that conform to the geometrical symmetry and whose distribution is characterized by the exponentially decaying probability distribution, which is beneficial to the hardware coding and decoding.

2.2 Introduction to FPGA Development Environment

2.2.1 FPGA Development Platform

FPGA (Field Programmable Gate Array) is a flexible and programmable hardware platform, which is widely used for real-time image processing and compression in aerospace missions due to its highly parallel characteristics and low power consumption. The FPGA board used in this paper is the AX7020 development board officially published by ALINX, equipped with the xc7z020clg400 chip of the Zynq7 series, which contains 53,200 LUTs, 106,400 FFs, and 220 DSP resources, with 1 Gb of DDR3 memory mounted off-chip, as well as various types of PCIe interfaces.The Matlab version is 2017b. The simulation hardware configuration is: CPU Intel Core i7-9750H at 2.6GHz, RAM: 16GB.

2.2.2 FPGA Design Methodology

(1) Top-level design: Top-down design starts with top level design based on the functional division of the system, and after the top-level design is completed, the functions of each module are designed in detail. The advantage of the top-down design method is that it makes the design hierarchical and easy to manage. The top-level design includes: all the global logic, input and output ports, all the modules exemplified as black boxes, and so on. These tasks should be completed before the modularized design, and then the design of each module is carried out in parallel. In practice, the inputs of the encoder implemented in the FPGA will be connected to a line scan camera or indirectly through an A/D converter, while the outputs will be connected to the channel that transmits the stream. Figure 3 shows the block diagram of the FPGA implementation of the CCSDS encoder.

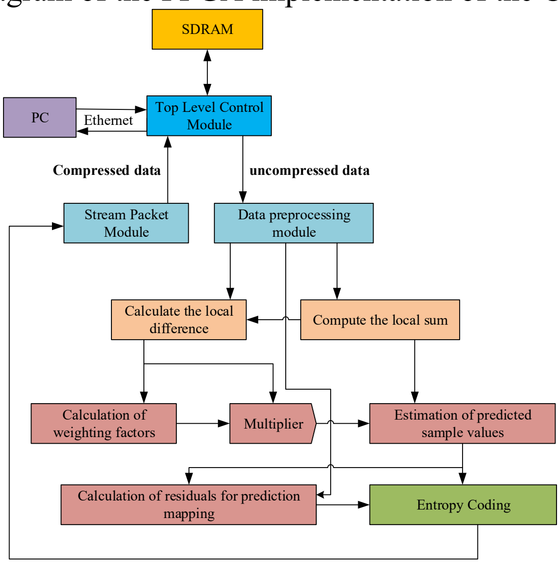

Figure 3. Block diagram of FPGA implementation of CCSDS encoder

(2) The number of system operation delay cycles and throughput: the number of system operation delay cycles refers to a system from the beginning of data processing, to the first output available 'a total of the number of operation cycles needed' is an important indicator of the system, together with the highest clock frequency determines the throughput of the system, to reduce the number of operation delay cycles 'It requires the designer to understand the system's algorithms on the basis of the design of the algorithm structure of the module level parallelism is relatively high. Due to the non-recyclable nature of the satellite system, the stability of the system requires a very high level of stability, so the clock frequency of the system has a certain upper limit, in this case, to improve the throughput of the system, it is necessary to improve the parallelism of the system.

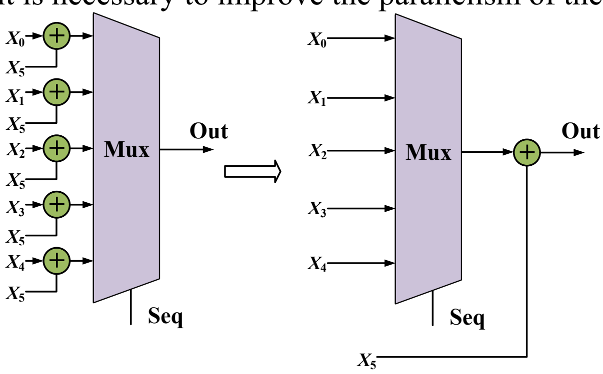

Figure 4. Conversion of unshared resources into shared resources

2.3 Principles of Parallel Acceleration Algorithms

2.3.1 Introduction to Parallel Algorithms

Parallel Computing refers to the process of using multiple computational resources simultaneously to solve computational problems, which is an effective means to improve the computational speed and processing power of computer systems [23]. In terms of parallel computing modes, it can be mainly categorized into two modes: temporal parallelism and spatial parallelism. Temporal parallelism refers to the division of the task into different parts, and each part can be processed simultaneously at the same time. Spatial parallelism refers to the simultaneous execution of computing tasks on multiple computing units, which is specifically divided into data parallelism and task parallelism. Data parallelism means that for the same task, the data to be processed is reasonably allocated to different computing units for execution. Task parallelism is the splitting of a computational problem into multiple subtasks that can be processed simultaneously. In general, the algorithms in task parallelism are more difficult to disassemble, whereas data parallelism is for numerical operations within a task, which only requires data to be independent of each other.

2.3.2 Parallel Algorithm Design

Solving a given problem using a parallel processor system requires the design of a parallel algorithm based on the type and characteristics of the system. There are usually three ways to do this: first, detecting the inherent parallelism in the serial algorithm found and directly serializing it, second, designing a new parallel algorithm from the characteristics of the problem itself, and third, modifying an existing parallel algorithm to make it solvable for another similar class of problems. Currently, the following design aspects of parallel algorithms are commonly used [24]:

(1) Pipelining Technique is an important design technique for parallel algorithms, which is particularly prominent in VLSI parallel algorithms. The basic idea is to divide a computational task t, into a series of sub-tasks t1, t2, …tm, so that once t1 is completed, the succeeding sub-tasks can start immediately and perform the computation at the same rate.

(2) Divide-and-Conquer Technique, the basic idea is to decompose the original problem into a number of sub-problems with the same characteristics and divide-and-conquer, and share the computational load on each computational node in order to reduce the amount of computation of individual nodes, so as to accelerate the overall operational efficiency of the purpose. If the size of the resulting subproblem is still too large, the partitioning strategy can be used repeatedly until the subproblems are easily solved. There are two common partitioning methods: data partitioning is where each node basically performs the same task, but with different data. Task partitioning breaks down the total task into several subtasks and spreads the task across the nodes.

(3) The Balancing-tree Technique, which involves constructing a balanced binary tree with the input elements as leaf nodes and then traversing back and forth from leaf to root. The success of this method is partly due to the fast access to the required information in the tree. The balanced binary tree method can be generalized to arbitrary balanced trees where the number of children of the inner nodes is not just two. This method is very effective for data propagation, compression, extraction, and prefix computation.

(4) Double Technique, also known as Pointer Jumping Technique, is particularly suitable for dealing with the data structure represented by chained lists, directed graphs, or rooted trees, and has a wide range of applications in graph theory and chained list algorithms. The distance between the data to be processed is progressively doubled with each recursive call, and after k steps all data with a distance of 2k can be computed. The doubling technique is used in communication algorithms such as global broadcast and global diffusion algorithms only.

2.3.3 Applications of Parallel Algorithms

The reason for choosing parallel computing is that serial computing does not meet the demands of solving the problem, either in terms of time or in terms of computational accuracy. Nowadays, the performance of computers or the method of dealing with problems are progressing, but the requirements for timeliness are also getting higher and higher. So parallel computing is an effective way to obtain higher performance of computing, and the current application of image parallel processing technology is multifaceted. From the urgency of the application of image parallel processing technology, its application areas are mainly concentrated in the military, industrial automation and monitoring alarms as well as the criminal investigation of public security, etc., with the strong impetus of these application areas, the image parallel processing technology has been rapidly developed [25].

3. Simulation Modeling And Result Analysis

In order to verify the correctness and validity of the real-time lossless compression processing algorithm based on FPGA proposed in this paper, this section builds a verification system using the AX7020 development board FPGA chip officially published by ALINX as the design verification platform. The entire design of the image lossless compression algorithm is programmed using the Verilog HDL hardware description language, and the data source used for verification is a multispectral image with different widths and heights and number of bands. The width and height of the input images are not more than 4095 pixels, the number of bands is not more than 15 and the bit depth is not more than 14 bits. To verify the correctness of the system, a series of multispectral images are tested in this paper. The source of the images are multispectral images of rivers and mountains taken by satellites. The size of the former river image is 1024*1024, the number of bands is 8, and the bit depth is 10 bits. The size of the mountain range image is 1024*1024, the number of bands is 4, and the bit depth is 8 bits. The pixel storage format of these multispectral images are in BIP format. Decompressing the uploaded data after FPGA compression and comparing it with the original data, the results are identical. It proves that this system can realize the lossless compression of multispectral images accurately and correctly.

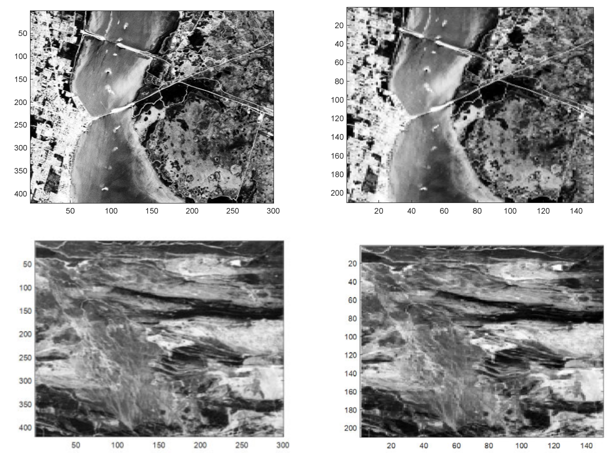

Figure 5. Compression comparison of rivers and mountains (original image on the left, compressed image on the right)

Simulation results show that the CCSDS algorithm has good compression performance in dealing with rivers and mountains, etc., and its compression rate is up to 99.8%, which can realize the spectral image without any compression. In addition, porting the CCSDS algorithm to the FPGA device improves the compression speed of the algorithm.

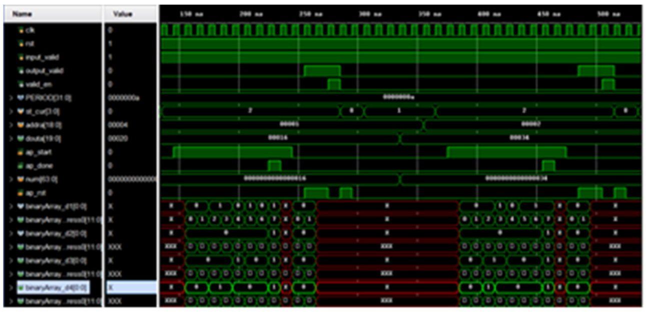

Figure 6. vivado module simulation diagram

douta is ram outputting the data to be encoded, and binaryarrayd_dx is outputting the RICE encoded data.

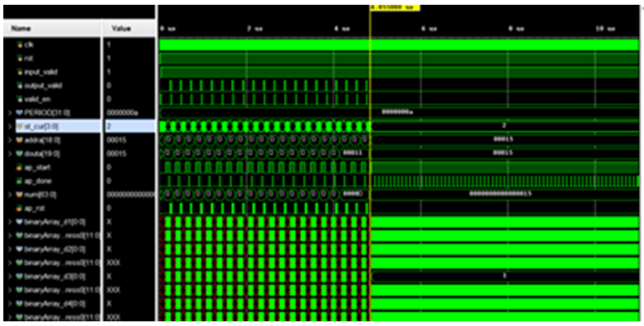

Figure 7. vivado simulation results

After compiling the system and checking the Quartus II static timing analysis tool, it is known that the imin.mat to ima.mat matrix dimensions are reduced by 3/4 and the compression ratio is 646kb/89kb=7.26 times. The compression speed is increased by 3 times. CCSDS compression algorithm only needs 4.85us to encode and complete, while Matlab needs 1583us to complete, the compression time speed is increased by 326 times. Through vivado simulation, it can be clearly seen that the lossless compression of hyperspectral image combined with parallel algorithm has significant efficiency and can meet the high-speed processing of image information.

4. Conclusion

In this paper, based on the CCSDS hyperspectral image lossless compression algorithm standard, the algorithm is analyzed and combined with parallel acceleration algorithm on FPGA chip to realize real-time lossless compression processing of hyperspectral images. The image compression system makes full use of the advantages of FPGA parallel computing, which not only can correctly realize the function of the algorithm, but also can meet the system performance requirements. Finally, the image compression system is simulated and verified using simulation tools and host computer software to verify the feasibility of the hardware system implementation, and the following conclusions are obtained:

(1) Through the compression of CCSDS-based FPGA hyperspectral images, CCSDS has achieved a better balance between compression performance and hardware implementation complexity, suitable for high-speed real-time processing of spatial data, and basically realized the lossless compression algorithm analysis results show.

(2) The parallel algorithm is proposed with FPGA chip as the design verification platform, and simulation experiments are carried out. The experimental results show that the parallel algorithm improves the compression speed by 326 times in comparison with the ordinary Matlab compression speed, and greatly reduces the fractal compression time under the premise of guaranteeing the image quality. It shows the important application value of parallel computing in the field of image compression

References

[1] Gao Chao. Research and implementation of CCSDS image compression algorithm based on FPGA[D]. Dalian University of Technology,2007.

[2] Liu T. Research and implementation of lossless compression algorithm for CCSDS of hyperspectral images[D]. Shenyang University of Aeronautics and Astronautics,2020.

[3] Shuxian He. Implementation of lossless compression algorithm for CCSDS hyperspectral images based on Tilera platform[D]. Xi'an University of Electronic Science and Technology,2015.

[4] Lin Chengchen, Hwang Yintsung. An efficient lossless compression scheme for hyperspectral image using two-stage prediction[J]. IEEE Geosci Remote Sens Lett, 2010, 7(3):558-562.

[5] B. Aiazzi, P. Alba, L. Alparone, et.al. Lossless compression of multi/hyper-spectral imagery based on a 3-D fuzzy prediction[J]. IEEE Geosci Remote Sens Lett, 1999, 37(5):2287-2294.

[6] A. K. Rao, S. Bhargava. Multispectral data compression using bidimentional interband prediction[J]. IEEE Geosci Remote Sens, 1996, 34(2):228-240.

[7] Wang Lei, Wu Jiaji, Jiao Licheng, et.al. Lossy-to-lossless hyperspectral image compression based on multiplierless reversible integer TDLT/KLT[J]. IEEE Geosci Remote Sens Lett, 2006, 6(3):587-591.

[8] J. A. Saghri, A. G. Tescher, J. T. Reagan. Transform Coding of Multispectral Imagery[J]. IEEE Signal Process Mag, 1995, 12(1):32-43.

[9] Ma Jing, Wu Chengke, Li Yunsong, et al. Dual-direction prediction vector quantization for lossless compression for LASIS data[C]. Proceedings of IEEE data compression conference, Snowbird, Utah, USA, 2009:458-467.

[10] M. J. Ryan, J. F. Arnold. Lossy compression of hyperspectral data using vector quantization[J]. Remote Sens Environ, 1997, 61(3):419-436.

[11] G. R. Canta, G. Poggi. Kronecker-product gain-shape vector quantization for multispectral and hyperspectral image coding[J]. IEEE Trans Image Process, 1997, 7(5):668-678.

[12] Wu Lijing. Optimization research and FPGA design of lossless compression algorithm for solar magnetic field images[D]. Beijing Jiaotong University,2019.001403.

[13] Qu Bingyu. Research on vector quantization technique for image compression[D]. Chongqing University of Posts and Telecommunications,2018.

[14] Mano M. M, Ciletti M. D. Digital design and Verilog implementation [M]. Beijing: Electronic Industry Press, 2007: 249-250.

[15] JIANG Quanlu, WANG Lixin, Lv Yongjia, et al. Design of LMS-based adaptive filter [J]. Electronic Design Engineering, 2011, 19(14):67-69.

[16] Kefalas N, Theodoridis G. Low-memory and high-performance architectures for the CCSDS 122.0-B-1 compression standard [J].Integration, 2019, 69:85-97.

[17] Gabriele M., Riccardo C. y Matteo B., et al. CCSDS 131.2-B-l telemetry transmitter: a VHDL IP core and a validation architectu re on board RTG4 FPGA [J].Acta Astronautica, 2020, 176:484-493.

[18] IPN. Low-Complexity Lossless Compression of Hyperspectral Imagery via Adaptive Filtering: IPN Progress Report[R]. California Institute of Technology, 2005.

[19] Jose Enrique Sanchez, Estanislau Auge, Josep Santalo, et al. Review and implementation of the emerging CCSDS Recommended Standard for multispectral and hyperspectral lossless image coding:2011 First International Conference on Data Compression, Communications and Processing, 222-228, 2011[C]. CA: California Institute of Technology, c2011.

[20] Deng, H.-T. Research on adaptive lossless compression algorithm for multispectral images and FPGA realization [D]. Haikou: Hainan University, 2014.

[21] Wang Liping. Research on lossless compression technology of hyperspectral image based on prediction [D]. Xi'an: Xi'an University of Electronic Science and Technology, 2014.

[22] Yin Bin,Pei Jiajun,Meng Guangyuan et al.FPGA implementation of lossless compression algorithm for CCSDS multispectral images[J]. Journal of Wuhan University(Science Edition),2016,62(06):531-536.

[23] M. J. Quinn. Parallel computing: theory and practice[M]. New York: McGraw-Hill, 1994.

[24] Bao-Hua Cao. Research on parallel algorithm based on JPEG2000 image compression[D]. University of Electronic Science and Technology,2005.

[25] Wang Shaofei. Research on parallelization method of digital image compression algorithm[D]. Shandong Normal University,2008