Research on telemetry data compression technology based on inter frame differential adaptive run length encoding

Shi Fenglei*
Aircraft Flight Test Technology Institute of China, Xi’an, China
*E-mail: shifl002@avic.com

Abstract: Aiming at the problem of massive historical telemetry data storage in flight test, a lossless compression method based on adaptive interval run length encoding is proposed. Aiming at the problem of low compression efficiency of traditional run length encoding algorithm for word data, by studying the storage characteristics of telemetry data, this algorithm automatically identifies the frame format of telemetry data, and carries out longitudinal run length adaptive interval encoding for inter frame differential data to improve the compression efficiency. The test results show that the compression ratio of the improved algorithm is improved by 58.1% and 1.5% compared with the traditional run length encoding algorithm and the inter frame differential lateral run length encoding algorithm.

1. Introduction
In order to ensure the safety of the flight test process, a large number of data are telemetered and transmitted down for flight test safety monitoring[1]. The bit rate of telemetry data generally ranges from 2Mbit / s to 16mbit / s, and the maximum amount of data generated in one hour will reach about 7.2gb. With the increase of flight test sorties, nearly 100 TB telemetry data will be generated in a year. A large number of historical telemetry data has caused a great burden on the storage system. At present, only about half a year's data is stored, and the data half a year ago is stored by offline backup, which makes it difficult to analyze the historical data quickly when necessary. In order to reduce the requirement of telemetry data storage space, data compression is one of the key technologies. In order to ensure that the data can be restored accurately after compression, lossless compression encoding scheme is needed. Due to the huge amount of telemetry data stored, the performance of data compression and decompression should also be considered in the design of compression algorithm.

At present, lossless compression algorithms mainly include entropy encoding and dictionary encoding [2][3][4][5]. As a lossless compression encoding technology, entropy encoding replaces each letter in the data stream with a bit of different length. Huffman encoding, arithmetic encoding, Shannon van Nuel encoding and other compression methods belong to the entropy encoding algorithm; Dictionary coding uses the correlation between source symbols to build a dictionary for the information in the data stream. According to the dictionary, the repeated information is replaced by encoding, so as to achieve the purpose of compressing data. The typical dictionary encoding algorithms are RLE, LZ77, LZ78, LZO, LZW, etc. As a classical compression algorithm, the complexity of Huffman encoding mainly focuses on the probability statistics and the generation of probability tree. When the probability of each symbol in the information is not different, Huffman encoding algorithm can’t achieve good data compression effect, or even get a larger volume of compressed data; LZ series algorithm has a higher compression rate for most information, but the algorithm is complex. The computational complexity of arithmetic coding is focused on floating-point number.
The principle of run length encoding is simple, and the compressed data can be compressed by one traversal. The algorithm has small complexity and high time response. It is suitable for data compression with more continuous and repeated bytes [6]. The traditional run length encoding algorithm mainly encodes in bytes. For the flight test telemetry data, the data is in frames. The traditional run length encoding algorithm can’t compress it well. Therefore, this paper improves the traditional run length encoding algorithm and proposes a differential longitudinal adaptive interval run length encoding algorithm which is suitable for telemetry data of flight test.

2. Analysis of telemetry data storage characteristics
The telemetry data of flight test is generally based on the main frame, and each main frame is composed of several sub frames with fixed number and same length. For any specific testing machine, the sub frame length is fixed, which is mainly composed of synchronization head, sub frame count and data parameters. The telemetry data mainly collects the sensor parameters installed on the test machine, as well as the analog quantity, switching value and bus parameters, etc. especially, the bus parameters extracted account for the majority of all test parameters. The bus parameters of this part are generally updated when the aircraft opens the corresponding recording system, and the values of the previous time remain unchanged at other times. In terms of space and time, the telemetry data of flight test can be regarded as a combination of several two-dimensional matrices. Each main frame is the change of data in space, and the combination of several main frames is the change of data in time. Because there are many different parameters in the main frame, when the parameter sampling rate is low, the parameters have no repeatability in the main frame, and the correlation between the data is relatively small. But for the adjacent main frames, the parameters in each main frame are the same, so the telemetry data of flight test has correlation between frames.

The traditional run length encoding is mainly based on bytes. When there are a large number of consecutive identical bytes in the data stream, run length encoding can obtain a good compression ratio. Because the run length code is encoded in bits or bytes, when the parameter data in the main frame is represented by one byte and the approximate values of adjacent parameters are the same, the algorithm can achieve good compression effect. But in flight test, telemetry parameters are mainly stored in 2-byte or 4-byte units. When the data stream is directly inspected by byte, the data repetition rate is greatly reduced. Taking the analog quantity acquisition as an example, the collected analog quantity is stored in 2 bytes, and the five consecutive analog quantity values in a certain subframe is as follows:

\{32768, 32770, 32771, 32769, 32775\}

The value stored in the above array is

\{0, 128, 2, 128, 3, 128, 1, 128, 7, 128\}

Since there are no consecutive repeated bytes in the above byte stream, the traditional run length encoding algorithm can only get a larger compressed data stream. In order to ensure the effective data compression results of telemetry data recorded in flight test of different aircraft, a longitudinal run length adaptive encoding algorithm based on inter frame difference is proposed according to the structure characteristics of telemetry frame in flight test.

3. An improved method of run length encoding
When the traditional run length encoding algorithm compresses the data stream, it directly encodes the compressed data stream according to the sequence of the data stream. For the data stream with less correlation between the connected data, the compression effect is generally poor. For the telemetry data frames recorded in flight test, the high sampling rate parameters exist in the same position of multiple sub frames, and the sub frames have large longitudinal local correlation in the main frame. The position of each parameter in the main frame is fixed, and the parameters between the connected main frames have temporal correlation, so it can be seen that the inter frame data has the characteristics of global repeatability. Therefore, lossless compression algorithm can be used to compress the inter frame redundant data. The idea of MPEG compression between frame which is widely used in the field of video compression is used for reference. MPEG compression algorithm takes 10 to 15 frames as a
compressed image group, and takes the first image as the intra encoding image (I image). It can adopt discrete cosine transform (DCT). The other images in the group are regarded as P images, and the different points of P images relative to the adjacent images are compressed by using predictive encoding technology, so as to achieve the optimal compression effect. For flight test telemetry data with inter frame correlation, the flight test telemetry data frame is divided into groups, and the reference frame data and inter frame difference value are compressed. Considering the particularity of flight test telemetry data, in order to obtain complete and accurate decompression data during decompression, lossless compression algorithm is needed for reference frame data or inter frame difference value.

According to the characteristics of telemetry data in flight test, the telemetry data frames of flight test are divided into groups periodically, that is, the telemetry data frames are divided into a group every fixed number, and the first frame of the group is defined as I frame (key frame), and other frames are defined as P frame. Compression algorithm needs to compress the difference between I frame and P frame and adjacent frames. Run length encoding algorithm is mainly considered when compressing telemetry data. Run length encoding is suitable for compressing data with local repeatability, and has very fast compression and decompression speed. The flight test telemetry data has the characteristics of intra frame local repetition and inter frame global repetition, so run length encoding is used to compress the difference between P frame and the previous frame.

4. Adaptive interval run length encoding algorithm
Most of the parameters in the telemetry data of flight test are stored in 2 or 4 bytes. Taking a test machine as an example, the bit rate of telemetry data is 4194304bit / s; the length of the subframe is 512 bytes (4096bit), the start word of the subframe is the synchronization header of the subframe (including 4 bytes, generally 0xfe6b2840), and the subframe count of main frame after the synchronization header is used to block the molecular frame; each main frame contains 32 subframes, and the main frame repeats 32 times per second. The frame format is shown in the figure below:

![Figure 1 The main frame structure of telemetry data](image)

It can be seen that the sub frame synchronization head has no repeatability with other parameters in the horizontal direction, but with other sub frames in the vertical direction. There is no correlation between the number of sub frames in the main frame in both horizontal and vertical directions, but the number of sub frames after difference is 0 in the main frame longitudinally, which has good repeatability. Therefore, after the difference of adjacent frames, the difference data has strong correlation and repeatability in the longitudinal direction, and the run length encoding algorithm can get better compression effect.

Because the frame format information of telemetry data is not known in advance, it is necessary to automatically identify the frame format information of telemetry data to be compressed before compression. Firstly, the stored telemetry data stream of flight test is read, and the sub frame
synchronization head (0xfe6b2840) is taken as the start recognition word of the sub frame, and the start and end of the first sub frame in the main frame is identified with the mark of the first sub frame count (generally 0); continue to read the sub frame with the next sub frame count of 0, then the basic format of the telemetry data main frame can be obtained. The frame format information is recognized automatically by the program, and the frame difference is carried out with the main frame as the unit. The differential data is encoded by the longitudinal run length encoding algorithm, and the frame format information and interval information are marked on the head of the compressed data. In terms of encoding scheme, the complexity of inter frame differential adaptive interval run length encoding algorithm is only three times of that of traditional RLE.

5. Algorithm validation
In order to verify the effectiveness of the algorithm, the telemetry data recorded in flight tests of different test aircraft are tested in the same environment. The test includes two processes: data compression and data decompression. The test platform is a win7 32 notebook computer with i5-4310u dual core 2.0GHz and 4GB memory. The comprehensive evaluation is based on the data compression rate, data compression speed and data decompression speed. TD-RLE is the inter frame differential lateral run length encoding algorithm, and LD-RLE is the inter frame differential longitudinal run length encoding algorithm.

![Figure 2. Comparison of data compression speed](image-url)
As can be seen from the figures, the complexity of the inter frame differential vertical run length encoding algorithm is increased by 2 times and 0.5 times compared with the traditional run length encoding algorithm and the inter frame differential horizontal encoding algorithm due to the
consideration of the correlation between the inter frame data and the vertical correlation of the frame data. The increase of the algorithm complexity affects the data compression speed and decompression speed, and the data compression speed decreases by 67.4% and 3.4% on average 8%, and the data decompression speed decreased by 44.1% and 21.3% on average. Considering the vertical correlation of intra frame data, the average compression rate of data is 58.1% higher than that of traditional RLE algorithm, and 1.5% higher than that of inter frame lateral differential run length encoding algorithm (TD-RLE). The average size of compressed data is about 10.6% of the original data.

6. Conclusion
Aiming at the problem of massive historical data storage in flight test telemetry, an adaptive interval run length encoding algorithm is proposed. By automatically identifying frame format, inter frame difference algorithm is used to remove inter frame data redundancy, and longitudinal run length encoding is used to compress the differential frame data. The results show that the average compression speed is about 66 MB/s, and the average size of the compressed data is 10.6% of the original data, the average data decompression speed is about 56 MB/s, which can meet the needs of mass flight test telemetry data storage and use.

The next step will continue to study multithreading parallel processing technology, make full use of multiple cores of CPU to further improve the speed of data compression and decompression, so as to reduce the data compression time and access delay.

Acknowledgments
The author would like to thank AVIC’s stable support project (WD-2019-2-2) for its support to this study.

References
[1] Bai Xiaoxian, Yang Tingwu, yuan Bingnan. The Telemetry Development Trends and Countermeasures in Aviation Flight Test [J]. Measurement and control technology, 2010,29 (11): 6-9.
[2] Ma Siwei. History and Recent Developments of AVS Video Coding Standards [J]. Journal of Computer Research and Development, 2015,52 (01): 27-37.
[3] Zhao Liping, Lin Tao, Zhou Kailun. An Efficient ISC Offset Parameter Coding Algorithm in Screen Content Coding [J]. Chinese Journal of Computers, 2017, 40(05):1218-1228.
[4] Wu Haifeng, Zhan Wenfa, Cheng Yifei. Journal of Electronic Measurement and Instrumentation [J]. Journal of Electronic Measurement and Instrumentation, 2016,30 (04): 638-644.
[5] Liu Xiong’en, Huang Xiao Yang. A Lossless Compression Method for Discrete-tone Images [J]. Journal of Xiamen University (Natural Science), 2014,53 (01): 31-35.
[6] Fang Jianping, Hao Yue, Liu Hongxia, Li Kang. A Hybrid Run-Length Coding for Soc Test Data Compression [J]. Acta Electronica Sinica, 2005 (11): 55-59.