

Application of Improved Hybrid Compression Algorithm in GPS Data Compression

She Xuebing^{1,a}, Zheng Xin^{2,b}

¹Department of College of information engineering, Jiangxi science and Technology Institute, Nanchang, 330098, China

²Department of College of information engineering, Jiangxi science and Technology Institute, Nanchang, 330098, China

^aemail:shexuebing1982@sohu.com, ^bemail:9714851@qq.com

Keywords: Hybrid compression algorithm; Huffman algorithm; Run-Length Encoding (RLE) algorithm; Vehicle monitor-ring system; GSP data

Abstract. In the paper we introduce a hybrid compression algorithm, which is the combination of Huffman algorithm and RLE algorithm, for compressing the GPS data. This algorithm acquires statistical characteristics of GPS data according to the NMEA0183 protocol, mixes Huffman algorithm and RLE algorithm to compress GPS data, to improve the coding efficiency and to restrain data expansion. Huffman algorithm has high compression rate on duplicated single-byte data while RLE algorithm has high compression rate on duplicated code segment. The flag bit is added in the process of encoding for encoding for the classification processing on GPS data in order to effectively identify the outputs of two kinds of algorithm when decoding and to ensure the complete decoding of compressed data. This improved hybrid compression algorithm is applied to local storage and 3G remote transmission of vehicle terminal GPS data, results show that the algorithm has clear improvement in compression performance of GPS data.

Introduction

In vehicle monitoring system, the vehicle terminal needs to upload the GPS signal (longitude, latitude, speed, direction, etc.) to the monitoring center in time. The monitoring center according to the communication protocol will receive the positioning information for local storage, easy for real-time monitoring and historical track playback. In the process of monitoring and management of the team, the GPS file is transmitted to the monitoring center. In the case of large amount of data, the data is compressed effectively, reduce information redundancy, reduce communication costs, reduce the occupancy of the transmission channel is the key problem to be solved in the process of data transmission. Because the GPS text data has a code segment and the height of the unit character.

In several lossless compression algorithms, the advantage of Huffman algorithm is based on the statistics of the frequency of individual characters in the information, and the output of the encoding is the same as the different word terminal, which ensures the unique translation of the code. Encoding process, the higher the data repetition, the higher the compression rate, but only for the characters one by one, so that the time to meet the code segment, but will reduce the compression efficiency of the. RLE algorithm has the advantage of the same value of continuous code segment with its value and string length (i.e., the number of repeat) to express, the same number of consecutive characters and the same number of times, the greater the compression ratio, the better, but because the RLE algorithm for fixed length characters, so it has limitations in the application process. In view of the characteristics of GPS data format, such as the kinds of algorithms are used to mix the two kinds of algorithms, such as the repeatability of the terminal, the 6 bit repetition of the first 7 bits and the degree of latitude. The improved Huffman and RLE hybrid compression algorithm in the process of encoding, the first of the GPS data to scan, extract the duplicate length code segment, will have a repeat code segment feature of the data to add a flag bit F1, the use of RLE algorithm for compression processing; the single byte of the data extracted from the F0, add a

logo Huffman, the use of compression algorithm for processing; other data directly copy, do not do any processing. The purpose of adding flag bits is to identify the Huffman encoding and RLE encoding in the decoding process of the hybrid algorithm, thus to achieve the full reduction of data after mixing the GPS data with encoding.

GPS data format

At present, the GPS satellite signal receiver adopts the NMEA0183 protocol communication standard which is developed by the National Institute of marine electronics. GGA (Positioning System Fix Data) DOP (Position and Active Satellites satellites Specific Minimum Data CNSS in Over Ground Speed Ground view DOP RMC GLL Global CNSS Geographic CNSS Recommended Course GSA GSV TVG and. Because the vehicle terminal needs to collect basic positioning information, so the terminal GPS system using the GPRMC statement. For example, a complete \$GPRM statement is received: \$GPRMC, 121530, A, 3413.5730, N, 10854.1613, E, 0.282269.026030512, *39. The GPRMC statement takes up 70 bytes (11 of them are used to separate the characters) and receives a data per second, which can be used to calculate the average local storage capacity of a vehicle is about 5.8M. According to the characteristics of GPS data format, the \$GPRMC pier is a repeating field, in the urban area of vehicle monitoring, longitude values of seven character, latitude value of 6 almost unchanged. Therefore, the repeating field compression high rate of RLE algorithm and the individual character repeat degree is high, pressure good shrinkage of the Huffman algorithm are mixed to the GPS data of compression.

Huffman algorithm

The performance index of the data compression is: encoding distortion, compression ratio, algorithm complexity, and anti-channel error rate. These factors in different applications, the focus is also different. In the process of GPS data compression, the main consideration of the compression ratio, because the terminal GPS data per second to receive a satellite signal, according to the communication protocol is the first time to upload a positioning information, data volume, encoding distortion is not very high, in the transmission process using 3G wireless communication, the channel is wider than GPRS, so the signal error rate is not high.

The dominant idea of Huffman compression algorithm is to scan the file, according to the probability of the occurrence of scanned data symbols for encoding, for the large probability of the input in accordance with the use of a shorter code word. In lossless compression encoding, the average bit rate of Huffman encoding can be close to the source of information entropy. According to the principle of Huffman coding, data compression after the shortest length can be calculated by

$$H = - \sum_{i=1}^n P_i \log_2 \frac{P_i}{2}, \text{ which } P_i \text{ said } i \text{ of the kind of operation code in the program probability.}$$

Gpsdata.txt compression algorithm based on Huffman compression, the compression results are shown in Table 1 (part of the data), which only has the GPS text file \$GPEMC.

Table 1 Huffman compression algorithm for GPS data of encoding

Encoding character	The frequency of character/ time	Encoding results
G	10	1010
P	10	1010
R	10	1010
M	10	1010
C	10	1010
1	42	00
2	21	010
3	10	1010

RLE algorithm

In the vehicle monitoring system, the terminal GPS data format and the front of the latitude and longitude information of the terminal GPRMC and the information of the latitude and longitude are repeat code section, so the RLE algorithm is used to compress this part of the code segment. The principle of the RLE algorithm scans the code segments that are duplicated in the data stream which is composed of characters, and replaces the duplicated code segments with the value of their value and the number of repeats and the flag bits.

In the compression process, non-repetitive bytes can have arbitrary length without being control byte interrupt, unless specified tag byte appear in the non-repetitive section (at most two bytes to encode) is scarce. According to this principle can statistics appear to repeat the code . We use Microsoft Studio Visual 2008 environment in accordance with the principle of RLE compression algorithm, using C++ to write RLE encoding program, GPS data format in the "0512" (monthly), "GPRMC" (10854.1), "3413.5", "" (latitude), and other heavy and complex fields were placed in the case of different code segments (repeat length of different code) of the impact of the compression rate, the results are shown in table 2.

Table 2 data compression algorithm for GPS data

Repeat code segment number	The length of the repeat code/byte	Compression ratio/%
6	4(0512)	45
6	6(\$GPRMC)	67
6	7(10854.1)	78
6	6(3413.5)	67

As can be seen from table 2, GPS data, in the case of the number of repeat code is equaled, the length of the repeat code segment appears longer, the use of RLE algorithm for the better compression effect.

Implementation of improved hybrid compression algorithm

Huffman compression algorithm and the RLE algorithm with high compression ratio of Huffman data are scanned to extract characters. The F0 algorithm is used to extract the data. When the data is extracted, the Huffman algorithm is used to perform data compression. The RLE algorithm is used to carry out data compression. The F1 algorithm is used to extract the data. This encoding's solution is mixed in two different ways, but due to the addition of a sign bit Huffman compressed encoding data and RLE compressed encoding data to distinguish, so, in the process of decoding, if you encounter a sign bit F0, that is the use of Huffman decoding algorithm; if you read the sign bit F1, you know that the back is a code segment of data, that is, the use of RLE decoding algorithm. As a result of the encoding process in order to decode the program to distinguish between two kinds of encoding, in the repeat character or repeat the code segment in front of a bit of information to add a logo, so in the hybrid compression process requires additional huf_sig F1 (using Huffman encoding number of characters in 1) F0 and rle_sig (the number of characters in the RLE series is more than 1 of the code segment), the effect of GPS data compression further discussion.

According to the principle of the improved hybrid algorithm, the paper shows that the length of the file is composed of two parts, which is compressed by Huffman algorithm, and the other part is RLE compression algorithm. So the length of the mixed compression encoding after the same as the two part of the compression of the encoding data length and. In order to compare the Huffman algorithm and the RLE algorithm, we add the data monitoring and control code in the process of encoding. The main purpose is to count the effects of single byte repetition and the repetition rate and length of the code segment on the compression ratio of the hybrid compression algorithm.

Conclusion

The improved hybrid compression algorithm makes full use of the characteristics of Huffman algorithm for single byte data compression and RLE algorithm to compress the GPS data in the vehicle monitoring system, and proposes a method to identify Huffman encoding and RLE encoding in the process of decoding. The compression results of the same GPS text file with Huffman algorithm and RLE algorithm are improved by the improved hybrid compression algorithm. However, because of the distinction between the two types of encoding in front of each of the output code will also have a certain impact on the performance of the new scheme, the problem needs to be further explored.

Reference

- [1] Hisashi Kurasawa , Atsuhiko Takasu , Jun Adachi. Load Balancing Scheme on the Basis of Huffman Coding for P2P Information Retrieval[J].IEICE transactions on information and systems, 2013,E92-D(10).
- [2] Li Changzhen.GPS global positioning system [J]. modern physics knowledge, 2010,12(C00): 90-92.
- [3] Wang Xiaoming. The study of the adaptive maximum stroke length encoding compression algorithm [J]. Journal of Liaoning University of Technology. 2013.2(1):31-33.
- [4] Qian Dejun, Zhang Zhe, Hu Chen.NMEA0183 protocol analysis of [J]. electronic devices. 2011.4(30):698-701.