# Pure electric buses status information compression and transmission methods basis on optimized Huffman coding algorithms

**Jingfeng Yang[1, 2, 3], Yong Li[4], Nanfeng Zhang[1, 5], Jiarong He[6], Yueju Xue[1*]**

[1]*South China Agricultural University, Guangzhou, 510642, China*

[2]*Guangzhou Public Transport Data Management Centre, Guangzhou, 510620, China*

[3]*Guangzhou Transport Information Control Centre, Guangzhou, 510620, China*

[4]*Guangzhou Institute of Geography, Guangzhou, 510070, China*

[5]*Guangzhou Entry-exit Inspection and Quarantine. Guangzhou, 510623, China*

[6]*South China University of Technology, Guangzhou, 510641, China*

*Received 1 October 2014, www.cmnt.lv*

**Abstract**

To strengthen safety supervision, satellite positioning information is needed to be uploaded as well as battery status information of pure electric buses. For the unbalanced coverage of mobile communication base station, data loss and incomplete data are prone to emergence when a large amounts of real-time data transmission through 3G/GPRS network in the poor communication environment of subregion, which is difficult to achieve large-scale farmland information collection and transmission simultaneously. Therefore, optimized Huffman coding compression algorithm for text-oriented information is promoted in this paper. The whole process test results of data exchange and decompression show that, the data compression algorithms can realize data compression effectively. The data sampling period is relatively smaller and its adjacent data is closer, the higher compression rate can be obtained by corresponding compression algorithms, and it can the basically ensure the decompression data without distortion. This will improve the efficiency of information transmission of transportation, as well as to ensure the integrity of information, which is of great significance to realization of transportation energy conservation.

*Keywords:* vehicle information, data compression, pure electric buses, battery status information

## 1 Introduction

The key operational vehicle information safety supervision has been continuously strengthened along with the national, provincial governments at all levels. The standard terminal with satellite positioning function were forcibly mounted on passenger bus, tourist charter, dangerous goods vehicles, coaches, trucks, heavy trucks, buses, taxis and so on. Some cities also require passenger bus, tourist chartered buses, taxis and other installation equipment using a video camera in order to enhance information security regulatory standards. Regardless the using of satellite positioning terminal or video camera equipment, it must bear a lot of the cost of wireless data transmission. If using new energy vehicles, such as electric buses, hybrid buses, electric taxis, etc., due to the need to upload more types of data to the regulatory status of the platform, its wireless transmission costs will rise at the same time. How to meet the requirements of the competent authorities, and saves the cost of wireless data transmission has become an important problem to be solved at present industry. The compressed data transmission is one of the important ways to solve the problem, and it can ensure the real-time and data quality of data transmission in the city communication environment, many scholars have conducted extensive research in this areas. Vehicle information is now covered by the text, pictures, video and other types of information. In order to improve data transmission quality, many researchers have proposed a variety of compression and transmission methods, for example, Hashemian [1], Sharma [2], Shi [3], Meng [4] et al proposed for the Huffman coding technology for wireless transmission of GPS, including GPS coordinate, records and other text data compression and transmission problems. Hu [5], Jou [6], Wang [7], Barr [8] et al put forward the LZW compression algorithm, which can be optimized by establishing a method to quickly find a dictionary and it can greatly reduce the time of data compression to achieve optimal compression, and improve transmission efficiency. Wavelet compression [9] method for wireless transmission of compressed gradually becomes a hot spot in recent years. In terms of traffic information, based on GML data compressions, Zhang [10] achieved a traffic emergency data and realized a framework of urban traffic emergency system architecture; Through random matrices with restricted isometric conditions of the

***Corresponding author** e-mail address: xueyj@scau.edu.cn

original high-dimensional data onto a low dimensional space, Li [11] realized the data fast and efficient compression, and via a convex optimization algorithm to complete the data decompression in traffic information processing terminal after data transmission; According to the characteristics of traffic data, Zhao [12] using its research and comparative data compression methods based on principal component analysis (PCA) and independent component analysis (ICA) method, respectively. Scholars' research results could furnish some useful suggestions in the transmission terminal data to a certain extent. However, the problem can not be ignored is, because of the unbalanced nature of the distribution of urban mobile communication base stations, urban communication environment vary greatly, the data transmission capacity is in greater differences in different parts of the city, part of the region exists communication blind area. The method proposed by scholars rarely used in acquisition and transmission terminal data on one hand, also rarely take into consideration differrences the problem in the larger urban communication environment for real-time transmission of data, that is, how to realize text, pictures, video and other information integrated transmission while coverage of communication base station is uneven with large difference between the communication environment in urban areas. In addition, the large flow of data transmission needs to consume more energy; low-power transmission [13] has important significance to prolong the length of the various devices work. Therefore, based on priority principle approach for simple algorithms, short time-consuming, minimizing overhead of data collection terminals and the backend server, the pure electric buses status information compression and transmission methods basis on optimized Huffman coding algorithms is proposed in this paper. Data collection vehicle information compression and transmission method can be achieving the optimized by encryption, compression, transmission, reception, decryption process, which is important to improve the transmission efficiency of vehicle information to ensure the integrity of information and transportation industry to achieve energy saving.

## 2 Optimized Huffman coding algorithms

According to national standards, Ministry of Transportation and the provincial technical standards requirements, the current standard of vehicle information including position, speed, altitude, license plates, start/flameout, lights, driving status, drivers and other types of text that collected through the data collection terminals. In addition, the new energy automotive vehicle information needed to collect additional battery status such as temperature, and overall cell voltage, current and other battery management information. The data with different format, size, and type will be uploaded via the data exchange and the wireless communication module of the vehicle terminal.

Vehicle status terminal with satellite positioning usually set 15 seconds for the data collection period when collecting for pure electric bus vehicle information. The collecting cycle for battery status information of vehicle battery management system is a relatively short period. Part of collecting cycle of vehicle battery management system is in less than five seconds, whose data generated in terms of quantity and frequency are more than the data collected by vehicle status terminal with satellite positioning.

The whole process from the point of view of data transmission, namely, data collection, transmission, reception, storage throughout the process point of view, when compared to a large number of simultaneous transmission vehicle terminal especially, the main pressure is focused on how the backend server receives a large number of short, transmission, integration of data transmitted by the terminal.

The way to solve the problem of large amounts of data transmission and access mainly includes three. First, data optimization in vehicle terminal, extract the key significance with application data, the remaining data is temporarily stored in the terminal memory device. Second, data are identified in the data receiving platform to extract of critical data, but it must guarantee the reliability of data transmission. The last, data are compressed by vehicle terminal firstly, and decompressed on the platform.

The compression-decompression method can ensure the integrity of data transmission, but also can reduce the data communication flow, which can save communication costs, but its disadvantages are also very obvious. Optimization transmission of terminal is with less disadvantages, the main disadvantage is it requires the performance processing each group of data compression, which would lead to transmission time delay. Real-time upload of vehicle information is demanding on safety supervision. Moreover, with higher hardware configuration of vehicle terminal at present, the capable of handling large amounts of data vehicle terminal hardware compression technology is promoted. Data collecting, data compressing and transmitting could be performed at the same time, which could not be simultaneously affectted.

### 2.1 DATA COMPRESSION ALGORITHMS

Due to the limitations of the vehicle terminal hardware conditions, the applications for data transmission compression by vehicle terminal were less in the past for a long period of time in the practical. Images, video and other data that takes up a lot of storage space are usually collected in offline mode, namely, images, video and other data in not considering the real-time requirement when used by off-line way.

Although the vehicle terminal hardware technology continues to improve, but data collection, compression and transmission operate at the same time will occupy the terminal necessarily overhead occupy the terminal. Therefore, the selection of data compression algorithms in this paper compared with traditional data compression algorithms. The main principles include that, it must reach the compression effect and the algorithm should not be too complicated in limited bandwidth resources, which would avoid the influence of compression and transmission efficiency, and not occupy too much memory. Taking up too much memory will affect data collection, as well as other applications occupy a larger memory and use a common transmission protocol.

Based on the operating environment consideration of pure electric bus, the selection of compression algorithm for

operating state of the commercial vehicle in this paper must be followed principles of less calculation, fast compression, simple algorithm and easy to implement. The algorithms must to be reached an optimum compression effect, which cannot be too complicated to avoid affecting the compression efficiency of the transmission under the limited bandwidth resources to meet the requirements of the terminal hardware and communications environment. Text data is the main data for the commercial vehicle terminal collecting (including GPS/Beidou positioning information). Optimized Huffman coding technique [1, 15] for effectively lossless compression is proposed for the GPS/Beidou position data and the other text of commercial vehicles operating state to be transmitted in this paper.

Huffman coding principle is: Set consisting of the text is represented as a set of characters with $I=\{I_1, I_2, I_3 \cdots I_n\}$, where $I_x$ represents different characters of text. Assuming the frequency of a character $I_x$ is $F_x$, the code length is $L_x$. To make the total length of code source text file of the shortest, you need to determine the coding mode. Making the minimum value of $\sum_{i=0}^{n} F_x L_x$, this encoding is called Huffman coding [3]. Huffman code based on Huffman tree, Huffman tree construction steps are shown as follows [3, 4].

Step 1. Set the given $n$ weights $\{w_1, w_2, w_3, \cdots, w_n\}$ according to $n$ binary trees, where every tree $T_x$ has only one weight $w_x$ is the root of 1, the left and right sub-tree are empty.

Step 2. Select the root node of the binary trees in $F$ with the minimum weights of sub-tree as to construct a new binary tree, and the root of the new lien binary is the value of its left and right sub-tree root node the sum of the weights.

Step 3. Delete the two trees in $F$, while the new binary was added to $F$.

Step 4. Repeat Step 2 and Step 3 until $F$ set up with only one tree, this tree is the Huffman tree.

Typically, a large number of repeated characters are existed, such as GPS data. GPS and the other repeated characters text data included by vehicle information can be regarded as redundant information to be removed. On the basis of Huffman compression coding table for rapid processing after compression by contrasting data, then stored in the storage buffer for post-processing. Among them, Huffman compression coding tables are the number of times pre-generated by the background server for text data in the statistical characters, and pre-stored in the Flash of vehicle terminal. The Realization of the process is shown in Figure 1.

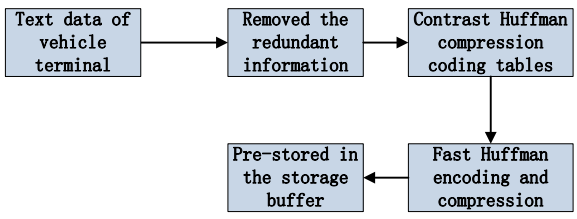


FIGURE 1 the compression flow chart of vehicle terminal text information

Huffman encoding is completely according to the characters of probability to construct coding, there is no error protection function. The algorithm needs to calculate the source character of probability statistics and obtains the probability distribution of the source symbols. This process requires a large cache using the hardware implementation of the algorithm. Generally, it believes that the decoding is complexity, and it is not conducive to the realization of hardware implementation [15, 16]. Huffman coding and arithmetic coding is a typical probability model, many scholars proposed the dictionary model for optimizing the original model, the typical optimizing algorithms such as LZW algorithm [4, 15], etc.

For compression and decompression speed, LZW algorithm is fast, and requires only one scan of the compressed data to be compressed. The compression ratio of the algorithm is very good for the continuous input stream to be compressed with high rate repetition. But the main drawback of this algorithm is the adaptability is rather poor, and the compression rate is less than Huffman algorithm. Moreover, the LZW algorithm usually needs to combine with other algorithms to reach the expected aim for some file with very low complexity. For vehicle operation status data, the data type apparently does not apply to LZW algorithm.

Huffman coding method has certain limitation, such as it need to scan flow of input symbols for two times, be stored or transmitted by Huffman tree before the storage or transmission of Huffman coding results. However, the algorithm has high compression rate, simple and practical, and all the coding has uniqueness while decoding, very suitable for vehicle information data that has high identification requirements. In order to overcome the practical problem of cache, high complexity caused by the Huffman coding, we improved the structure of Huffman tree based on the original. The main steps are as follows:

Step 1. Initialize the established binary Huffman tree, one of the root node has a weight of 0.

Step 2. To encode without the new characters of newly generated two nodes, the connection of the parent node weights add 1, and another node, which is defined as node weights of 0 nodes to define new weight of 0.

Step 3. The new characters have been encrypted through search was coded character location, in the case of the characters of the weight add 1 with coding and compared with the original weight of the same node.

Step 4. Repeat Step 2 and Step 3 until all characters are encoded.

The improved way of Huffman coding by weight set to 0, the original requires two symbols flow is simplified to a scan. The algorithm only needs to scan a probability of 1 characters, and it can only exchange the node between the binary tree needs to node numbering, which reduces the occupation of the cache, and the complexity of the algorithm is decreased. Its deficiencies are mainly embodied in the complete lost of coding error protection function, and the data requirements are relatively high.

### 2.2 DATA DECOMPRESSION AND PARSING

To distinguish the uploaded data from different vehicle terminal and to ensure uniqueness of uploaded data, uploaded data must be parsed and recognized. Default the parsing code for different vehicle terminal before uploading the data. Read and Match the parsing code by the reception server to achieve data uniqueness verifying. The specific processing is shown as follows.

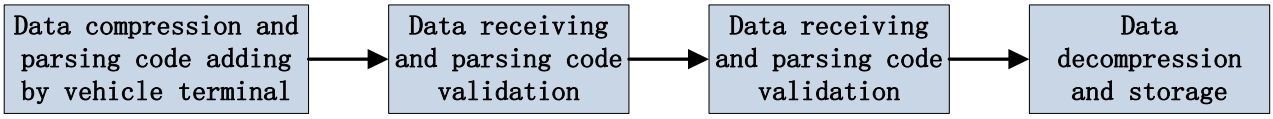


FIGURE 2 The Flow Chart of Data Parsing

Pictured above, it needs to setting parsing code corresponding to vehicle information first. Data compression and parsing code adding are processed by vehicle terminal. Backend server receives the data compression package, and matches the SIM card number, parsing code and other information corresponding to set parsing codes correspondding table. The perfectly matched data will be decompressed and written to the database.

## 3 Experiments and results

To meet the requirements of real-time data transmission, this paper uses 3G network as a data transmission channel for data transmission. In 3G network, the prerequisite to achieve the theoretical rate is the premise of carrier frequency for data communication, completely and exclusive to a user only. Due to the self-interference system affected by the communication environment, the use of the number of actual formation, the actual data transmission rate and theory of data transmission rate is often of large difference in the vehicle information collection environment.

### 3.1 DATA TRANSMISSION TEST

Data transmission test mainly completes two test including the network delay test and TCP transmission rate test in the actual environment in city village. Unicom's 3G networks were selected for communication channel for testing in city village environment. Data of vehicle status terminal with satellite positioning and vehicle battery management system were collected and uploaded, then received by server.

Round trip time (RTT) is usually used as a measure of 3G network time delay. Low delay time is of extreme importance in actual application. Lower delay is advantageous for the uplink direction to improve capacity and data throughput, and increase the coverage of high bit rate, obtain higher transmission bit rate. Typically, data transmission testing uses PING method, namely by obtainning response time of sending and receiving data packets to get connected server response time. When sending upload packets in requesting, all the serial number sequence will be added one firstly before sending, the replying information will also be marked the corresponding serial number. Then observe the Ping packet response messages to detect the link for data transmission delay calculation, such as packet loss, grouping repetition and fault sequence etc.

The following table shows the data uploaded test results in a village environment of a city. To be clear, several uploaded tests were started at 10 in the morning, ICMP packets for a total of 120 were send every time, the 3G bandwidth rate of 70-120kbps (city 3G bandwidth rate is about 700-900kbps), TCP transmission rate is close to 3G bandwidth rate.

TABLE 1 Testing results of raw data uploading

| Serial Number/Test Items | Maximum Value of Ping Testing | Minimum Value of Ping Testing | Average Value of Ping Testing | Received Volume | Packet Loss Volume | Packet Loss Rate |
|---|---|---|---|---|---|---|
| 1 (Packets:120) | 7043ms | 1190ms | 2478ms | 115 | 5 | 4.17% |
| 2 (Packets:120) | 6478ms | 948ms | 2136ms | 117 | 3 | 2.50% |
| 3 (Packets:120) | 5735ms | 1087ms | 2332ms | 115 | 6 | 4.17% |
| 4 (Packets:120) | 5845ms | 1101ms | 2447ms | 113 | 7 | 5.83% |
| 5 (Packets:120) | 6089ms | 1017ms | 2311ms | 116 | 4 | 3.33% |
| Mean | 6238.00ms | 1068.60 ms | 2340.80ms | 115.20 | 5.00 | 4.00% |

TABLE 2 Testing results of compressed data uploading

| Serial Number/Test Items | Maximum Value of Ping Testing | Minimum Value of Ping Testing | Average Value of Ping Testing | Received Volume | Packet Loss Volume | Packet Loss Rate |
|---|---|---|---|---|---|---|
| 1 (Packets:120) | 4546ms | 1078ms | 2213ms | 233 | 22 | 8.63% |
| 2 (Packets:120) | 4997ms | 1210ms | 2319ms | 230 | 25 | 9.80% |
| 3 (Packets:120) | 4912ms | 1109ms | 2190ms | 238 | 17 | 6.67% |
| 4 (Packets:120) | 5221ms | 1223ms | 2342ms | 239 | 16 | 6.27% |
| 5 (Packets:120) | 4893ms | 1005ms | 2201ms | 234 | 21 | 8.24% |
| Mean | 4913.80ms | 1125.00ms | 2253.00ms | 234.80 | 20.20 | 7.92% |

The result of data packets uploaded test shows that, the maximum delay time is 7043ms, the average delay time of 2340ms, which meets the basic requirements in real time data uploaded. 3G network takes the average packet loss rate 4.17% of data uploading while the network bandwidth is 70-120kbps. Relative to the GPRS data transmission mode, 3G network has been able to meet the requirement of actual application in the setting of upload strategy and selecting transport protocol, but also laid the foundation for the data compression test environment. In addition, it appears from the data uploading

test results, village area of the communication environment is relatively poor, data compression and transmission will still be of importance as well.

Due to environmental influences such as shaking hands error occurred etc., a large amount of data in the case of serious delays caused by transmission network, also had a serious packet loss in the original data uploading test. Direct transmission of original data easily lead to incomplete data, the delay is difficult to reflect the real-time data and so on in 3G networks.

Table 2 is uploading test result by APN for compressed data without identification and screening. The uploading data test method is similar to the original test data uploading method, each transmission is with 255 data packets improved by optimized Huffman coding. The difference between the two methods is to establish the corresponding APN network to upload data using 3G in the same use area.

The results of RTT test in APN network show data compression can effectively reduce the occurrence of packet loss in the area of villages of poor communication environment, while the uploading data is compressed by optimized Huffman coding before transmission.

## 3.2 DATA COMPRESSION AND DECOMPRESSION TEST

RTT test results showed that the time delay is less, the less the number of packet loss when transmitting a small amount of data. Transmitting a small amount of data of pure electric buses can effectively improve the efficiency of data transmission, but also to ensure the real-time data in the area of poor communication environment.

The data compression testing is shown in Table 3 for data transmission of original, after Huffman Coding, LZW algorithms and optimized Huffman Coding.

The main evaluation on the effect of data compression is mainly based on the compression rate and transmission time. Compression rate is one of the most important indexes to measure the data lossless compression algorithm. Compression rate is a direct measure of data reduction, the running time reflects the complexity of the algorithm, and the combination of the two indexes can be further estimated energy efficiency of communication. Transmission error rate is the rate between data length of accurately praised data had (compressed data containing the decompression process) not received by backend server and the data length before compressing. Packet loss rate is the ratio between the packet numbers of server has not been received and data packets had been transmitted by vehicle terminal.

It should be noted that, according to current technical standards for data collection and transmission requirements, this paper adopts 15s, 10s, 5s three data collection cycle, and the collection time is 300 minutes of pure electric buses collected operating status data as the test data source. This paper also shows the result according to the comparison of the original data, Huffman encoding, LZW algorithm, optimized Huffman coding after data transmission.

TABLE 3 Testing result of data compression

| Serial | Data Type | Collect Cycle (s) | Compressed Data Length (kb) | Compression Rate (%) | Transmission Time (s) | Error Rate of Transmission (%) | Packet Loss Rate of Transmission (%) |
|---|---|---|---|---|---|---|---|
| 1 | Original | 15 | 335.98 | 0.00% | 49.71 | 3.71% | 6.13% |
| 2 | Original | 10 | 479.04 | 0.00% | 75.66 | 3.99% | 6.98% |
| 3 | Original | 5 | 943.46 | 0.00% | 169.45 | 4.65% | 8.69% |
| 4 | Huffman | 15 | 180.26 | 46.35% | 25.22 | 2.78% | 4.69% |
| 5 | Huffman | 10 | 262.68 | 45.17% | 38.41 | 2.94% | 4.99% |
| 6 | Huffman | 5 | 511.43 | 45.79% | 91.33 | 3.05% | 5.21% |
| 7 | LZW | 15 | 202.32 | 39.78% | 24.33 | 2.38% | 3.89% |
| 8 | LZW | 10 | 287.52 | 39.98% | 42.37 | 2.75% | 4.24% |
| 9 | LZW | 5 | 569.67 | 39.62% | 104.97 | 2.98% | 4.57% |
| 10 | Optimized Huffman | 15 | 142.12 | 57.70% | 19.46 | 2.24% | 3.79% |
| 11 | Optimized Huffman | 10 | 204.37 | 57.34% | 27.76 | 2.39% | 3.88% |
| 12 | Optimized Huffman | 5 | 408.62 | 56.69% | 68.65 | 2.49% | 4.09% |

TABLE 4 Testing result of filtered data compression

| Serial | Data Type | Collect Cycle(s) | Compressed Filtered Data Length (kb) | Compression Rate (%) | Transmission Time (s) | Error Rate of Transmission (%) | Packet Loss Rate of Transmission (%) |
|---|---|---|---|---|---|---|---|
| 1 | Original | 15 | 115.39 | 0.00% | 17.97 | 2.56% | 1.98% |
| 2 | Original | 10 | 180.12 | 0.00% | 28.37 | 2.62% | 2.01% |
| 3 | Original | 5 | 367.95 | 0.00% | 57.79 | 2.89% | 2.15% |
| 4 | Huffman | 15 | 66.34 | 42.51% | 9.34 | 2.27% | 1.31% |
| 5 | Huffman | 10 | 103.43 | 42.58% | 14.03 | 2.31% | 1.44% |
| 6 | Huffman | 5 | 207.12 | 43.71% | 29.42 | 2.35% | 1.49% |
| 7 | LZW | 15 | 76.53 | 33.68% | 10.64 | 2.02% | 0.89% |
| 8 | LZW | 10 | 117.67 | 34.67% | 17.08 | 2.17% | 0.98% |
| 9 | LZW | 5 | 236.14 | 35.82% | 31.65 | 2.29% | 1.12% |
| 10 | Optimized Huffman | 15 | 53.39 | 53.73% | 7.65 | 1.83% | 0.73% |
| 11 | Optimized Huffman | 10 | 82.76 | 54.05% | 9.89 | 1.97% | 0.78% |
| 12 | Optimized Huffman | 5 | 169.14 | 54.03% | 21.87 | 2.08% | 0.87% |

Experiment result shows that, compressed transmission effect was more notable. The transmission amount of compressed data is reduced greatly, accounting for about between the original amount of data of 1/4-1/3, and the corresponding transmission time is shortened into 1/4-1/3. In addition, it can be seen from the cycle of data collection, data sampling period is relatively smaller and its adjacent data is closer, all kinds of algorithm could get a higher compression rate. In particular, the correlation data has obtained the very good compression rate, which greatly enhances the communication efficiency of system, but also save the cost of communication.

For best results of actual test when using the optimized Huffman coding techniques, the data collection period is 15s. The test results further verify that a lower transmission error rate and transmission loss error can be obtained in the case of the transmission condition of data with less quantity and short transmission time.

The optimized Huffman coding technology has advantages of less calculation amount, high compression rate, and it is suitable for data transmission in the operation area of pure electric bus. For the LZW algorithm, the algorithm has a better compression efficiency for the correlation data in global or local, but because data field correlation of the pure electric bus operation status is small, resulting in limitation of application applied aspects of the object, and thus its compression efficiency is relatively poor in terms of optimized Huffman coding algorithm. To enhance the efficiency of real-time applications of pure electric buses operating status data, and avoid unwanted real-time transmission data occupying a lot of memory and data transmission channel, the paper also carries out data compression and transmission test after the data is filtered and identified, which are shown in Table 4.

The amount of data to be transmitted will be greatly reduced when the filtered and compressed data transmission mode is selected. Real-time data needed to transmitting can be greatly compressed in the transmission mode, which ensure data transmission for real-time application while the error rate of data transmission, data transmission packet loss rate fell to a lower level. Further, it can be seen from the data collection cycle that, the data sampling period is relatively smaller the closer adjacent data is; the data compression rate of various algorithms can be obtained. In particular, an excellent compression rate can be obtained in the case of the data with great correlation, which can greatly enhance the efficiency of communication systems, but also save a lot of communication costs.

In order to further verify the results of data compression algorithm, this paper compares the original data with the parsed decompressed data. Through the data exchanging, vehicle terminal could obtain the speed sensor data. Exchange data through data compression method based on the technology of Huffman coding for data compression, and then the original data file and compressed data files are transmitted to the backend server through 3G network. After transmission, read the received data file for decompression and inverse quantization, and compare the difference between the original data and data contrast reduction before and after data compression.

To be clear, the speed testing data was selected for a passenger bus driving data from 8:00 am to 12:00. An average of 5 seconds data collection cycle was set in the pure electric bus vehicle terminal, and it generated for a total of 2772 (theoretical value should be 2880, because of various reasons in the process of data collection, there are 108 errors, lost or not collected data samples) speed data during the period. Before the data compression, the original speed changing curve is shown in Figure 3. In Figure 3, the horizontal represents the speed of data collection time; the ordinate represents speed value, where are the same as Figure 4. After the data compression and transmission, the parsed speed changing curve is shown in Figure 4.

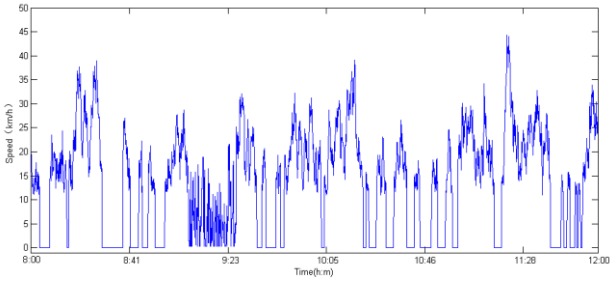


FIGURE 3 The velocity curve before data compression

Figure 4 shows the decompressed and parsed data, which have been compressed, transmitted, parsed, decompressed and warehoused to the backend server.

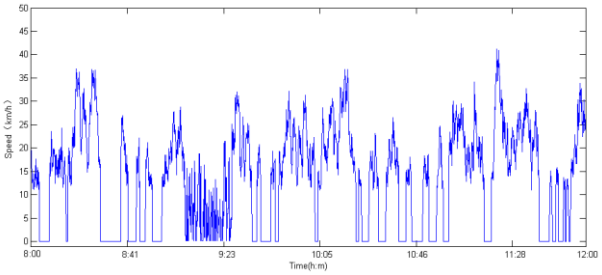


FIGURE 4 The velocity curve after data decompression

From the speed curve of the data before and after compression, the speed error value curve can be seen, the speed data consistent with original data basically before compressed through the process of compression, transmission, parsing, decompression, and data storage. Coding, compression and other data processing technology have met the basic requirements in actual application. To accurately describe the data differences between situation before and after the data compression and transmission, this paper calculated the mean variance to discriminate compression deviation after transmission of the data. After calculation, the mean square difference before and after data compression and transmission is of only 0.2789, which is within a reasonable range. The main reason for the deviation of the data exists in data loss, data failing to parse and the other reason during data transmission.

A total of 2702 data is to accurately have been parsed during 2880 speed data transmitted by vehicle terminal, the proportion of accurate transmission and parsing data is up to 96.25%, which basically meet the requirements of actual application. 108 errors, loss, and no sample data contained

the error data of 25, accounting for 0.87% of the total, loss of data 49, accounting for 1.70% of the total loss rate, no sample data of 34 due to various reasons did not collect, accounting for 1.18% of data should be collected of the total. The availability testing experiment results of compression and transmission algorithm show that, the algorithm has high efficiency and high transmission integrity.

## 4 Conclusions

Compression algorithm method is researched in this paper to the vehicle information including text information collection types (including position information), image and video to compress the data by using Huffman coding technology. On this basis, all kinds of vehicle information is transmitted based on NMEA-0183 various kinds of vehicle GPS monitoring transmission protocol. Through the process of vehicle information compression, transmission, decompression, parsing and storage whose data from vehicle status terminal with satellite positioning and vehicle battery management system, the compression algorithms are verified by speed data as an example. The algorithm achieved great test results in the experiments, the proportion of accurate data transmission and parse is up to 94.69%, basically meet the requirements of data uploading actual applications in different city region for the urban public transport, passenger transport, taxi and other industries. Meanwhile, the algorithm play an important role for enhancing the efficiency of wireless transmission applications and reducing the cost of data transmission as well as the time, which has the practical significance for the deployment of this base station, the selection of communication mode and the other uses. The data compression algorithms mentioned in this paper provides methods with transmission reliability and minimal distortion. It can provide reliable data transmission method for different mobile base stations coverage of the urban environment, which is important to improve vehicle transmission efficiency of information, ensure the integrity of information and the realization of transport energy conservation.

## Acknowledgments

The research is supported by Science Program of General Administration of Quality Supervision, Inspection and Quarantine the People's Republic of China (2014IK183) and Guangdong Research Projects Special Funds of Industry-Academy-Research Cooperation (2012B091100345) and Guangzhou Science and Technology Project (2014Y2-00044).

## References

[1] Hashemian R 2004 *IEEE Transactions on Communications* **52**(1) 6-8
[2] Sharma M 2010 Compression Using Huffman Coding International *Journal of Computer Science and Network Security* **10**(5) 133-41
[3] Shi P, Li M, Yu T 2013 Design and Application of an XML Date Compression Algorithm Based on Huffman Coding *Journal of Beijing University of Chemical Technology (Natural Science Edition)* **40**(4) 120-4 *(in Chinese)*
[4] Meng F Y 2010 *Researeh and Application of Huffman Coding in Environmental Protection Real Monitoring System* Thesis of Degree of Master of Ocean University of China 11-13 *(in Chinese)*
[5] Hu P, Zhang J Z 2012 Study and Implementation for Terminal Data Compression Technologies in Remote Fault Diagnosis System *Computer Engineering and Applications* **48**(34) 130-5
[6] Jou J M, Chen P Y 2006 *IEEE Transactions on Communication* **47**(9) 1278-83
[7] Wang Q, Qi C, Luo X 2005 Modified LZW algorithm and its parameters optimization *Journal of Chongqing University of Posts and Telecommunications* **17**(3) 351-5 *(in Chinese)*
[8] Barr K C, Asanovic K 2006 Energy-aware lossless data compression *ACM Transactions on Computer Systems* **24**(3) 250–91
[9] Yang S H, Jia Y, Zhou S W 2013 Virtual nodes-based wavelet compression algorithm for wireless sensor networks Ruanjian Xuebao *Journal of Software* **24**(3) 557-63 *(in Chinese)*
[10] Zhang H Y 2011 *Based on ArcGIS Server Urumuqi Traffic Emergency System Data Compression Technique Research* Thesis of Degree of Master of Xinjiang University *(in Chinese)*
[11] Li Q Q, Zhou Y, Yue Y, et al 2012 Compression method of traffic flow data based on compressed sensing *Journal of Traffic and Transportation Engineering* **12**(3) 113-20 *(in Chinese)*
[12] Zhao Z Q, Zhang Y, Hu J M, Li L 2008 Comparative Study of PCA and ICA Based Traffic Flow Compression *Journal of Highway and Transportation Research and Development* **25**(11) 109-14 *(in Chinese)*
[13] Xu H L, Zhang H, Shen Y, et al 2013 Environment monitoring system for flowers in greenhouse using low-power transmission *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)* **29**(4) 237-44 *(in Chinese)*
[14] Liu X J, Qiu X L, Sun C F, et al 2010 Design and application of knowledge model and PDA-based precision farming system *Transactions of the CSAE* **26**(1) 210-5 *(in Chinese)*
[15] Wang Q, Qi C, Luo X M et al 2009 Modified LZW algorithm and its parameters optimization *Journal of Chongqing University of Posts and Telecommunications* **3**(1) 85-91 *(in Chinese)*
[16] He X M 2011 *Lossless data compression and decompression software and hardware realization* Thesis of Degree of Master of University of Electronic Science and Technology of China 9-11 *(in Chinese)*

**Authors**

**Jingfeng Yang**

**Current position, grades:** post-doctoral student at the South China Agricultural University. Technology expert for transportation information system at Guangzhou Public Transport Data Management Centre and at Guangzhou Transport Information Control Centre.
**University studies:** MS degree at South China Agricultural University in 2008.
**Scientific interests:** telematics, GIS application and intelligent traffic, intelligent transportation network and transportation - geographic information system.

**Yong Li**

**Current position, grades:** associate professor at Guangzhou Institute of Geography, Guangdong Provincial Academy of Sciences.
**Scientific interests:** Study on the public traffic system using Geography Information System (GIS).
**Experience:** modeling and forecasting of the public transit passenger flow, public traffic system of Guangzhou City.

**Nanfeng Zhang**

**Current position, grades:** Dr., Senior Engineer, Chief Engineer at Guangzhou Entry Exit Inspection and Quarantine Bureau laboratory.
**University studies:** PhD degree at South China Agricultural University in 2014.
**Scientific interests:** mechanical and electrical integration, vehicle detection and fault diagnosis.
**Experience:** Member of the Society of automotive engineering of China, science and technology expert of Guangdong Provincial Department of Science and Technology, expert of Guangzhou City Auto Repair Association.

**Jiarong He**

**Current position, grades:** deputy chief of integrated four divisions at Foshan municipal people's government office, post-graduate student at South China University of Technology.
**Scientific interests:** software engineering technology.
**Experience:** participated in big software engineering projects, e.g. at the Department's OA automation construction and the Digital Urban Management projects.

**Yueju Xue**

**Current position, grades:** Dr., professor-researcher at the College of Information, South China Agricultural University.
**Scientific interests:** data mining, big data processing technology.
**Scientific interests:** data processing technologies, modeling of IOT and intelligent computing.