Research Article

Compression Algorithm of Road Traffic Spatial Data Based on LZW Encoding

Dong-wei Xu,^{1,2} Yong-dong Wang,^{1,2} Li-min Jia,³ Gui-jun Zhang,^{1,2} and Hai-feng Guo^{1,2}

¹College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China

²United Key Laboratory of Embedded System of Zhejiang Province, Hangzhou 310023, China

³State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China

Correspondence should be addressed to Dong-wei Xu; dongweixu@zjut.edu.cn

Received 20 October 2016; Revised 19 December 2016; Accepted 10 January 2017; Published 16 February 2017

Academic Editor: Xiaolei Ma

Copyright © 2017 Dong-wei Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Wide-ranging applications of road traffic detection technology in road traffic state data acquisition have introduced new challenges for transportation and storage of road traffic big data. In this paper, a compression method for road traffic spatial data based on LZW encoding is proposed. First, the spatial correlation of road segments was analyzed by principal component analysis. Then, the road traffic spatial data compression based on LZW encoding is presented. The parameters determination is also discussed. Finally, six typical road segments in Beijing are adopted for case studies. The final results are listed and prove that the road traffic spatial data compression method based on LZW encoding is feasible, and the reconstructed data can achieve high accuracy.

1. Introduction

The advent of big data brings unprecedented opportunities as well as challenges, especially in the field of transportation and traffic engineering [1, 2]. With the rapid development of science and technology, the intelligent transportation system (ITS) has developed continuously, and its applications have become wide ranging. The ITS system can accomplish the tasks of road traffic data acquisition, processing, and transportation. Besides, it can complete the job of traffic state analysis, route guidance, and traffic control. As various road traffic detection systems are adopted in the road traffic field, the collected road traffic state data increase and become massive. This serious situation introduces a challenge for real-time transmission, storage, and guidance of massive road traffic data. Thus, it is necessary to find an efficient approach to compress real-time traffic states data which can save much storage space as well as providing some other applications [3]. And the compression method of road traffic states data has deeply promoted the managements for transportation administrators. Besides, useful compression method of road traffic data can also be applied to transportation research fields, and some inspirations may occur to the researchers.

The essence of road traffic state data compression is to represent the signal information with less data. Through effective compression and reconstruction, traffic data transmission and storage can be achieved [4–6].

In recent years, a great many of data compression methods have been explored in traffic and transportation fields. With the popularity of machine learning and data mining study among practitioners and researchers, some road traffic compression methods are presented. Due to the multidimensional and multigranularity characteristics of traffic and transportation big data, PCA method realizes the compression of road traffic states data through reducing the dimensions of original data [7]. As an emerging technology, compression sensing has also been used in data compression due to its superiority. Compression sensing breaks through traditional Nyquist sampling theorem restricts and can collect and compress data simultaneously. Making use of the redundancy characteristics of road traffic states, compression sensing technology achieves the estimation [8] and compression [9–11] of road traffic states data. Since the road traffic states data possess the spatial-temporal correlation and similar characteristics, Xiao et al. presented a spatial-temporal model based on road traffic data compression and

decompression technology of 2D discrete wavelet transformation, realizing the denoising compression of ITS system [12]. Ou et al. proposed road traffic volume data compression based on artificial neural network [13].

Some modified and improved methods also fill the compression gap. The embedded devices in motor vehicles also generate abundant data for researches to investigate the compression of road traffic states data [14]. Making use of the GPS positioning data produced by the mobile devices of travelers, Ma et al. presented a differential preprocessing method, and a dynamical Huffman algorithm was adopted to compress GPS positioning data [15]. Wang et al. put forward an encoding algorithm with self-adaptive switching mode according to specific format [16]. Hou presented a stop-wave mode based on the concept of the compression factor and its differential equation [17]. Song et al. proposed a hybrid spatial compression algorithm and error bounded temporal compression algorithm to compress the spatial and temporal information of trajectories, respectively [18]. However, many researches do not have a common baseline for their performance analysis and provide the infrastructure to operate on a publicly available dataset.

The existing road traffic data compression methods mainly focus on the compression of road traffic network data. However, in recent years, limited literature has been written on the road traffic spatial data compression methods of different road segments on similar time nodes. Some literatures on predictions are investigated temporally and spatially in recent years. The studies are not only in road traffic field, but also in the field of transportation.

The travel needs and travel routes of traffic participants exhibit certain regularity; thus, the road traffic spatial states of different road segments on similar time nodes represent strong relationships. That is, the changing curve of road traffic spatial state on different road segments on similar time nodes possesses some similarity. The correlation presents great probability for the compression of road traffic spatial data. Thus, based on the spatial correlation characteristics of the road traffic states, the road traffic spatial data on different road segments on similar time nodes are extracted for compression. LZW inherits the merits of LZ77 and LZ78 on compression efficiency and speed. Besides, the method easily achieves good performance. Thus, the LZW encoding is introduced in the study. Based on the spatial correlation of road traffic, a compression method of road traffic spatial data based on LZW encoding is proposed in this paper.

In this study, a compression method of road traffic spatial data based on LZW encoding is proposed to compress the road traffic spatial states data under the same time intervals, realizing efficient transmission and storage as well as display. The useful compression of road traffic states data can be efficiently used into feature extraction and traffic states prediction. Multivariate time series analysis is similar to the proposed method, which can take into consideration both spatial and temporal correlations. In our study, we used the spatial correlation characteristics of road traffic states to compress the states data. The aims of the two studies are different.

Some motivations are explained here. Although the proposed compression method of our study is tested on the road traffic states data, it is also very useful for transportation management as well as transportation prediction. Besides, the compression can be also used for feature extraction, which can be applied to evaluate the traffic running states.

Based on the characteristics of road traffic flow, the PCA method can be used to analyze the correlation of spatial road segments [19, 20]. Then, the spatial road segments are selected to extract the data for compression. The spatial road segments denote the different road segments; the data on these segments are extracted on the spatial road segments at the same time intervals.

The contributions of the proposed algorithm are three-fold:

- (1) The PCA method was introduced to the algorithm to select the road segments with spatial correlation.
- (2) A novel road traffic spatial data compression algorithm based on LZW encoding was proposed to construct the difference data on selected spatial road segments under the same mode.
- (3) The proposed algorithm could determine the optimal parameters in the training process based on spatial historical data and base data on road traffic states.

The rest of this paper is organized as follows. The modeling methodology of the proposed algorithm is discussed in Section 2. In Section 3, parameter determination of the road traffic spatial data compression study based on LZW encoding is presented. The experiment results are shown in Section 4. The conclusion and direction for future studies are discussed in Section 5.

2. Compression Algorithm of Road Traffic Spatial Data Based on LZW Encoding

2.1. Framework of the Algorithm. The process of compression and reconstruction of road traffic spatial data is shown in Figures 1 and 2, respectively. First, the PCA method was used to select the road segments with the characteristics of spatial correlation. The road traffic spatial data under the same mode on different road segments were acquired to construct the reference sequences of road traffic characteristics. Based on the analysis of spatial correlation, the base road segment was selected and the data on which were regarded as spatial base data. Second, the historical data on other spatial road segments under the same mode was extracted as training data. The optimal threshold of road traffic spatial difference data was determined based on road traffic spatial base data under the same mode. Third, real-time spatial data on other road segments under the same mode were acquired as experimental data and the road traffic spatial difference data were acquired on the basis of road traffic spatial base data under the same mode. Finally, the compression and reconstruction of road traffic spatial difference data were achieved through LZW encoding and decoding technology, respectively.

FIGURE 1: The process of road traffic spatial data compression based on LZW encoding.

TABLE 1: Road traffic characteristics reference sequence information chart.

Reference sequence ID	Mode	Road segment ID	Time	Road traffic state parameters
-----------------------	------	-----------------	------	-------------------------------

TABLE 2: Road traffic characteristics reference sequence description chart.

Reference sequence ID	Reference sequence name	Description
-----------------------	-------------------------	-------------

FIGURE 2: The process of road traffic spatial data reconstruction based on LZW decoding.

2.2. Acquisition of Road Traffic Spatial Base Data

2.2.1. Selection of Road Segments with Correlation Based on PCA Method. Road traffic flows possess the characteristics of periodicity, similarity, correlation, and so on. The road traffic flows of spatial road segments indicate a strong spatial correlation. Thus, the PCA method was used in this study to select the road segments with the characteristics of correlation.

PCA is a multivariate statistical method that eliminates the correlation among the variable indicators. n -dimensions of road traffic state data can be effectively reduced to two dimensions, which can be illustrated in a 2D figure. Taking advantage of these characteristics, the related road segments can be selected. The process has been described in previous studies [19, 20].

2.2.2. Division of Road Traffic Running Modes. The road traffic running modes can be divided into two levels: the road network level and road segments level. Assuming that the running modes division identification of road network level and road segments level can be divided into g and h

submodes, respectively, the road traffic running modes can be divided into $g \times h$ modes in total. The modes can be shown as $M_{ode} = \{M_{11}, M_{12}, \dots, M_{gh}\}$. g and h can be determined by the road traffic running modes division identification. The running modes division identification of road network mainly refers to the impact factors of road traffic running modes on different dates. The road traffic running modes division identification of road segments refers to the influence factors of the road traffic running modes of the specific condition of the road segments, which can be illustrated as in Figure 3.

2.2.3. Construction Design of Road Traffic Characteristics Reference Sequences. Assuming the collection period of road traffic state data was Δt , then time format of road traffic information template can be illustrated as in Figure 4. The table format of the road traffic characteristics reference sequence can be described as in Tables 1 and 2.

Let $p + 1$ denote the total number of selected road segments, which can be described as follows:

$$L = [L_1 \ L_2 \ \cdots \ L_{p+1}], \quad (1)$$

where $p + 1$ is the number of spatial road segments; L_i ($1 \leq i \leq p + 1$) denotes the i th road segments; L represents the set of selected road segments with correlation.

Based on the correlation of road traffic spatial data, the base road segment is acquired to extract the road traffic data as road traffic base data. The road traffic data on other p spatial road segments are extracted as historical data and real-time data.

2.3. Optimal Threshold Determination of Road Traffic Difference Data. The data on other spatial road segments are extracted as training data. Under M_{gh} mode, the road traffic spatial difference data under the same mode are acquired based on road traffic spatial base data to conduct the threshold processing. Through LZW encoding, the optimal

TABLE 3

Δt	The collection period of road traffic state data
$(m * \Delta t)$	The m th period collection of road traffic state data, $0 \leq m \leq N$
N	The number of daily collected road traffic data
i	The i th road segment under M_{gh} mode
$ST_i(m * \Delta t, M_{gh})$	The road traffic data on i road segment at $(m * \Delta t)$ moment under M_{gh} mode
$SB(m * \Delta t, M_{gh})$	The road traffic data on base road segment at $(m * \Delta t)$ moment under M_{gh} mode
$S_i(m * \Delta t, M_{gh})$	The road traffic difference data between the training data on i road segment and the base data on base road segment at $(m * \Delta t)$ moment under M_{gh} mode
$e_i(m, M_{gh})$	The road traffic difference data between the training data on i road segment and the base data on base road segment at Δt to $(m * \Delta t)$ time intervals under M_{gh} mode
$he_i(m, M_{gh})$	The road traffic difference data between the training data on i road segment and the base data on base road segment after threshold processing Δt to $(m * \Delta t)$ time intervals under M_{gh} mode
$E_i(m, M_{gh})$	The threshold of road traffic difference data at Δt to $(m * \Delta t)$ time intervals under M_{gh} mode
$pe_i(n, M_{gh})$	The data on the difference data between i road segment and base road segment after LZW encoding at Δt to $(m * \Delta t)$ time intervals under M_{gh} mode
$S'_i(n, M_{gh})$	The n th data on the difference data between i road segment and base road segment after LZW encoding at Δt to $(m * \Delta t)$ time intervals under M_{gh} mode
m	The number of difference data between i road segment and base road segment before LZW encoding at Δt to $(m * \Delta t)$ time intervals under M_{gh} mode
n	The number of difference data between i road segment and base road segment after LZW encoding at Δt to $(m * \Delta t)$ time intervals under M_{gh} mode

TABLE 4

j	The j th road segment, $1 \leq j \leq p + 1$
$SM_j(m * \Delta t, M_{gh})$	The real-time data on j road segment at $(m * \Delta t)$ moment under M_{gh} mode
$MS_j(m * \Delta t, M_{gh})$	The difference data on j road segment and base road segment at $(m * \Delta t)$ moment under M_{gh} mode
$err_j(m, M_{gh})$	The road traffic difference data between the real-time data on j road segment and the base data on base road segment at Δt to $(m * \Delta t)$ time intervals under M_{gh} mode

Based on LZW encoding, the road traffic spatial data compression can be achieved. The best threshold of the difference data between i road segment and base road segment can be introduced into the difference data on the j road

segment and the base road segment under the same M_{gh} mode. Combining the LZW encoding, the difference data compression of j road segment and base road segment can be realized. The main expressions can be described as follows:

$$\begin{aligned}
herr_j(m * \Delta t, M_{gh}) &= \begin{cases} 0, & err_j(m * \Delta t, M_{gh}) < E_{opt}(M_{gh}) \\ err_j(m * \Delta t, M_{gh}), & err_j(m * \Delta t, M_{gh}) > E_{opt}(M_{gh}) \end{cases}, \\
herrs_p(m * \Delta t, M_{gh}) &= [herr_1(m * \Delta t, M_{gh}) \quad herr_2(\Delta t, M_{gh}) \quad \cdots \quad herr_{p'}(m * \Delta t, M_{gh})], \\
perr_{p'}(m * \Delta t, M_{gh}) &= w(herrs_p(m * \Delta t, M_{gh})), \\
perr_{p'}(m * \Delta t, M_{gh}) &= [MS_1(m * \Delta t, M_{gh}) \quad MS_2(m * \Delta t, M_{gh}) \quad \cdots \quad MS_{p'}(m * \Delta t, M_{gh})].
\end{aligned} \tag{5}$$

The characteristics are explained in Table 5.

The compression ratio is p/p' .

2.5. Road Traffic Spatial Data Decompression Based on LZW Decoding. Based on LZW decoding technology, the data reconstruction of difference data between p road segments and base road segment can be realized. Combining the base

data, the decompression of p road segments real-time data can be achieved. The main expressions are as follows:

$$\begin{aligned}
dperr_p(m * \Delta t, M_{gh}) &= w'(perr_{p'}(m * \Delta t, M_{gh})), \\
CSM_p(m * \Delta t, M_{gh}) &= SB(m * \Delta t, M_{gh}) + dperr_p(m * \Delta t, M_{gh}),
\end{aligned} \tag{6}$$

TABLE 5

$E_{opt}(M_{gh})$	The optimal training threshold
$herr_j(m * \Delta t, M_{gh})$	The difference data between the real-time data on j road segment and the base data on base road segment after threshold processing at $(m * \Delta t)$ moment under M_{gh} mode
m	The number of difference data between j road segment and base road segment before LZW compression at Δt to $(m * \Delta t)$ time intervals under M_{gh} mode
$herrs_p(m * \Delta t, M_{gh})$	The road traffic difference data set of p road segments at $(m * \Delta t)$ moment under M_{gh} mode
$Perr_{p'}(m * \Delta t, M_{gh})$	The set of difference data on p road segment after LZW encoding at $(m * \Delta t)$ moment under M_{gh} mode
p'	The number after LZW encoding at $(m * \Delta t)$ moment
$MS'_j(m * \Delta t, M_{gh})$	The j' th data on difference data after LZW encoding at $(m * \Delta t)$ time moment under M_{gh} mode

where w' denotes the LZW decoding; $dperr_p(m * \Delta t, M_{gh})$ denotes the spatial difference data on p road segments after LZW decoding at $(m * \Delta t)$ moment under M_{gh} mode; and $CSM_p(m * \Delta t, M_{gh})$ denotes the reconstructed road traffic real-time data on p road segments at $(m * \Delta t)$ moment under M_{gh} mode.

3. Parameter Determination

In the process of road traffic spatial data compression based on LZW encoding, the following parameters were involved: $SB(m * \Delta t)$, $ST_j(m * \Delta t)$, $E_i(m * \Delta t)$, per , $err_i(m * \Delta t)$, n , where $E_i(m * \Delta t)$ can be acquired by $SB(m * \Delta t)$ and per , n , and $err_i(m * \Delta t)$ can be acquired by $SB(m * \Delta t)$, $ST_j(m * \Delta t)$, and $E_i(m * \Delta t)$. Parameter settings here are only concerned with the effect analysis of the road traffic spatial data compression based on LZW encoding. Separately analyzing the effect of each parameter on the accuracy of the algorithm cannot guarantee an optimal algorithm because these parameters influence the accuracy of the algorithm in different ways. All of the parameters in the road traffic spatial data compression results should be considered when conducting the algorithm analysis.

The compression ratios are introduced to measure the effect of parameters on the precision of the algorithm. The main expression can be described as follows:

$$CR_p(m * \Delta t, M_{gh}) = \frac{CM_a(m * \Delta t, M_{gh})}{CM_b(m * \Delta t, M_{gh})}, \quad (7)$$

where $CR_p(m * \Delta t, M_{gh})$ denotes the compression ratio of p road segment at $(m * \Delta t)$ moment under M_{gh} mode; $CM_a(m * \Delta t, M_{gh})$ denotes the number of road traffic data before compression at $(m * \Delta t)$ moment under M_{gh} mode; and $CM_b(m * \Delta t, M_{gh})$ denotes the number of road traffic data after compression at $(m * \Delta t)$ moment under M_{gh} mode.

Different $(SB(m * \Delta t, M_{gh}), ST_j(m * \Delta t, M_{gh}), per)$ corresponds to different $NMAE$. Thus, the following expression is reasonable:

$$CR_p(m * \Delta t, M_{gh}) = f(SB(m * \Delta t, M_{gh}), ST_j(m * \Delta t, M_{gh}), per). \quad (8)$$

That is, a certain distribution relationship f exists between $(SB(m * \Delta t, M_{gh}), ST_j(m * \Delta t, M_{gh}), per)$ and

TABLE 6: The road segments information.

Road segment ID	Road segment name
HI3009b	Xiao Jie Bridge East to Bei Xiao Jie Bridge
HI3008b	Bei Xiao Jie Bridge to Yong He Gong Bridge
HI7058b	Yong He Gong Bridge to Capital Library
HI7036b	Capital Library to An Ding Men Bridge
HI7057b	An Ding Men Bridge to Zhong Lou North Bridge
HI7056b	Zhong Lou North Bridge to Gu Lou Bridge

$CR_p(m * \Delta t, M_{gh})$. The process of finding the maximum $CR_p(m * \Delta t, M_{gh})$ that corresponds to $(SB(m * \Delta t, M_{gh}), ST_j(m * \Delta t, M_{gh}), per)$ is training optimal parameters. Thus, the following model can be obtained:

$$\min f(SB(m * \Delta t, M_{gh}), ST_j(m * \Delta t, M_{gh}), per)$$

$$CR_p(m * \Delta t, M_{gh}) = \frac{CM_a(m * \Delta t, M_{gh})}{CM_b(m * \Delta t, M_{gh})}. \quad (9)$$

Finally, the value of $(SB(m * \Delta t, M_{gh}), ST_j(m * \Delta t, M_{gh}), per)$ can be determined through statistical analysis of the reconstructed results of road traffic state.

4. Experiments

4.1. Data Acquisition

4.1.1. Road Segment Acquisition. The proposed compression algorithm is conducted with the road traffic spatial relevant data; thus, the selected data must exhibit the characteristics of spatial correlation. The road segments will be briefly explained here. The types of the road segments are express ways, the wide of which is similar. First, the volume data on six typical road segments in Beijing were adopted in the present study. The specific road segments were determined in Table 6.

Five days (June 11, 18, 19, 25, and 26 in 2011) of road traffic data were extracted to construct the reference sequences of road traffic characteristics. The road traffic state data collection interval is 2 min. As the correlation of road

TABLE 7: The cross correlation of the road segment.

Road segment	HI3009b	HI3008b	HI7058b	HI7036b	HI7057b	HI7056b
HI3009b	1	0.9218	0.9286	0.9289	0.9087	0.8349
HI3008b	0.9218	1	0.9645	0.9330	0.9011	0.8310
HI7058b	0.9286	0.9645	1	0.9285	0.8915	0.8294
HI7036b	0.9289	0.9330	0.9285	1	0.9107	0.8676
HI7057b	0.9087	0.9011	0.8915	0.9107	1	0.8199
HI7056b	0.8600	0.8616	0.8655	0.8676	0.8398	1

segments mentioned in the literatures [19, 20], the first two principal components can reflect most of the information of road traffic state. Based on PCA method, we can find that four road segments, HI3009b, HI3008b, HI7058b, HI7036b, exhibited strong correlation that can be determined by cross correlation.

The volume data on the six road segments from June 11, 2011, were extracted to determine the spatial correlation. The cross correlation is shown in Table 7. According to the table, the correlation of all road segments can be determined.

As shown in Table 7, the cross correlations between HI3008b and the other three road segments (HI3009b, HI7058, and HI7036b) were greater than 0.9. Thus, the HI3008b road segment served as the base road, and its collected data were considered as the base data. The volume data on the four road segments were selected for the case study to prove the performance of the proposed algorithm. This can be explained by the following reasons.

The change regularity of volume is mainly determined by the regularity of people's origin-destination (OD) travel. But for different date, people travel OD changes randomly. The travel on weekends has a comparative regularity. Thus, four days (June 18, 19, 25, and 26 on 2011) of road traffic data on spatial road segments were extracted to construct the reference sequences of road traffic characteristics.

4.1.2. Data Instruction. The collected road traffic data on the HI3009b, HI7058b, and HI7036b road segments from June 11, 2011, were considered as training data to conduct algorithm parameter settings. Under the same mode, the collected road traffic data on the HI3009b, HI7058b, and HI7036b road segments from four other days were regarded as real-time data to validate the proposed algorithm.

4.2. Results. The road traffic spatial volume data compression results based on LZW encoding on the HI3009b, HI7058b, and HI7036b road segments are illustrated in Figures 5–16.

The running time is provided here, which can indirectly reflect the calculation speed of the proposed method. Through several times testing, the average running time is approximate to 0.45. From the running time, we can see that the proposed method is simple and practicable.

The statistical reconstructed results of spatial volume data based on LZW encoding on HI3009b and HI7058b road segments from June 18, 19, 25, and 26, 2011, are illustrated in Tables 8 and 9, respectively. *CR*, *AE*, *marerr*, and σ denote the compression ratio, mean absolute error, absolute relative

FIGURE 5: HI3009b on June 18, 2011.

TABLE 8: Reconstructed results on HI3009b road segment.

Date	18	19	25	26	Average
<i>CR</i>	11.08	9.73	10.14	8.67	9.91
<i>AE</i>	11.24	11.28	14.14	11.92	12.15
<i>marerr</i>	12.42	14.82	14.09	13.81	13.79
σ	13.93	13.41	15.12	14.00	14.12

TABLE 9: Reconstructed results on HI7058b road segment.

Date	18	19	25	26	Average
<i>CR</i>	16.74	16.00	15.65	11.80	15.05
<i>AE</i>	6.65	6.93	6.83	7.41	6.96
<i>marerr</i>	6.87	7.67	7.07	8.51	7.53
σ	8.65	9.33	8.87	9.79	9.16

error percentage, and error standard deviation, respectively. Average denotes the mean value of the four indicators. *CR* is described in (7). *AE*, *marerr*, and σ can be described as follows:

$$AE = \frac{1}{p} \cdot \sum_p \left| CSM_p(m * \Delta t, M_{gh}) - SM_p(m * \Delta t, M_{gh}) \right|,$$

FIGURE 6: HI3009b on June 19, 2011.

FIGURE 7: HI3009b on June 25, 2011.

$$marerr = \frac{1}{p}$$

$$\cdot \frac{\sum_p |CSM_p(m * \Delta t, M_{gh}) - SM_p(m * \Delta t, M_{gh})|}{SM_p(m * \Delta t, M_{gh})},$$

$$\sigma = \sqrt{\frac{\sum_p (y_p(m * \Delta t, M_{gh}) - \bar{e}_p(m * \Delta t, M_{gh}))^2}{p - 1}}, \quad (10)$$

where

$$y_p(m * \Delta t, M_{gh}) = CSM_p(m * \Delta t, M_{gh}) - SM_p(m * \Delta t, M_{gh}), \quad (11)$$

$$\bar{e}_p(m * \Delta t, M_{gh}) = \frac{1}{p} \sum_p y_p(m * \Delta t, M_{gh}),$$

where $y_p(m * \Delta t, M_{gh})$ denotes the error data between the original real-time data and the reconstructed real-time data on p road segments at $(m * \Delta t)$ moment under M_{gh} mode; $\bar{e}_p(m * \Delta t, M_{gh})$ denotes the mean error at $(m * \Delta t)$ moment under M_{gh} mode.

4.3. Sensitive Analysis. A sensitivity analysis is the study of how the uncertainty in the output of a mathematical model or system (numerical or otherwise) can be apportioned to different sources of uncertainty in its inputs [21]. In Section 4.2, four road segments are selected, and HI3008b is used for training and the others are used for testing. To test the effect of data size on the compression and reconstruction results, a sensitive analysis is urgently needed. Since the proposed algorithm is applicative for big data in road traffic transportation data, a sensitive analysis is also required to test the feasibility for little and medium-size data. The data size

FIGURE 8: HI3009b on June 26, 2011.

can be indicated by the collecting time. Thereby, a sensitive analysis is conducted through testing the compression and reconstruction results indicators under different collecting time, that is, CR , AE , $marerr$, and σ .

The process of parameters determination is performed in Section 3, but the optimal parameter is determined under fixed collecting time. For different collecting time, the optimal parameters will be different. Thus, collecting time is considered as a variable to test compression results. Besides, in this process, we also follow the rule in (9).

Here, a brief data declaration is provided. In Section 4.1, one-day collected data (720) are used for experiment. To test the feasibility of the proposed method, we calculate

FIGURE 9: HI7058b on June 18, 2011.

FIGURE 10: HI7058b on June 19, 2011.

the experimental index under different collecting time on HI7058b. This may be regarded as a test bed. The sensitive analysis can be seen in Tables 11–14.

From the sensitive analysis results shown in Tables 11–14, we can see that the compression ratio of big-size data is relatively greater than little and medium-size data. And AE , $marerr$, and σ are all less than 10. The results show that the proposed algorithm is feasible.

A comparison is also provided here. PCA method is a famous data compression method; thus, we compare the proposed method with PCA method. We compare the reconstruction indicators on on June 19, 2011. The specific results are shown in Table 15.

From Table 15, we can see that the CR of LZW encoding is dramatically greater than that of PCA. The AE , $marerr$, and σ of PCA and LZW are very similar. The comparison proves that the performance of the proposed method is comparatively better.

4.4. Analysis of Experiment Results. Based on the experiment results conducted in Section 4.2, the following analyses are presented:

- (1) From Tables 8–10, the following results can be obtained:

For the reconstructed volume data, the average compression ratios are 9.91, 15.05, and 5.94 for the HI3009b, HI7058b, and HI7036b road segments, respectively; the average mean absolute error rates are 12.15, 6.96, and 10.32 for the HI3009b, HI7058b, and HI7036b road segments, respectively; the average absolute relative error percentages are 13.79, 7.53, and 12.00 for the HI3009b, HI7058b, and HI7036b road segments, respectively; the average error standard deviations are 14.12, 9.16, and 13.37 for the HI3009b, HI7058b, and HI7036b road segments, respectively.

FIGURE 11: HI7058b on June 25, 2011.

TABLE 10: Reconstructed results on HI7036b road segment.

Date	18	19	25	26	Average
CR	6.32	5.71	6.43	5.29	5.94
AE	9.61	9.76	10.73	11.19	10.32
$marerr$	11.10	12.11	12.02	13.25	12.00
σ	12.33	12.65	13.85	14.64	13.37

As the statistical data show, we can find that the performance of the HI7058 road segment is better than that of the HI3009b and HI7036b road segments.

FIGURE 12: HI7058b on June 26, 2011.

FIGURE 14: HI7036b on June 19, 2011.

FIGURE 13: HI7036b on June 18, 2011.

FIGURE 15: HI7036b on June 25, 2011.

- (2) The road traffic spatial volume compression ratio for the HI7058b road segment is higher than that for the HI3009b and HI7036b road segments.

The main reason is that the cross correlation of road traffic spatial volume for HI7058b is higher than that for HI3009b and HI7036b. From Table 7, a similar conclusion can be reached. Consequently, the volume compression ratio for the HI7058b road segment is higher than that for the other two road segments.

- (3) The precision of the reconstructed results of volume for the HI7058b road segment is higher than that for the other two road segments.

From Figures 5–16, we can get that the precision and stability of the reconstructed volume data for the HI7058b road segment is higher than those for the HI3009b and HI7036b road segments based on LZW encoding. The phenomenon is mainly caused by the cross correlation between the base volume data and real-time volume data on base road segment and other spatial road segments, respectively. From Table 7, similar conclusions can be reached.

- (4) Some peak points are missing and the phenomenon can be described by the following reason.

TABLE 11: The sensitive analysis results on June 18 on HI7058b.

Data size	180 (0:00–6:00)	360 (0:00–12:00)	540 (0:00–18:00)	720 (0:00–24:00)
CR	5.45	10.29	14.21	16.74
AE	4.09	5.92	6.35	6.65
marerr (%)	9.92	8.13	7.24	6.87
σ	5.43	7.90	8.40	8.65

TABLE 12: The sensitive analysis results on June 19 on HI7058b.

Data size	180 (0:00–6:00)	360 (6:00–12:00)	540 (12:00–18:00)	720 (0:00–24:00)
CR	3.00	10.59	15.00	16.00
AE	3.84	6.16	6.45	6.93
marerr (%)	10.98	9.14	7.86	7.67
σ	4.79	8.51	8.74	9.33

FIGURE 16: HI7036b on June 26, 2011.

The road traffic data at the peak points has a sudden change compared to the data on the base road segment. In the feature extraction process, the features are extracted based on the threshold processing of the difference data. If the features needed to be retained, we can shift down the threshold. In the reconstruction process, the peak points can be retained at the cost of compression ratio. From the reconstructed results shown above, the peak points are sustainable.

- (5) Several errors occur in the road traffic state reconstruction of this algorithm.

The errors are mainly caused by the following two reasons:

- (1) Obtaining the corresponding road traffic spatial states with a perfect match based on LZW encoding is difficult because of the limitations of the road traffic running characteristics.

- (2) The parameters exhibit a certain deviation. Determining the optimal parameters is irregular because they vary for different road traffic state datasets. The selected optimal parameters are determined based on the historical road traffic state data. Therefore, the current optimal parameters are approximately different from the historical optimal parameters.

5. Conclusions

An effective road traffic data compression algorithm can boost the data transportation and storage effectiveness of a road traffic system. The PCA method can be used to select the road traffic segments with strong correlation. Based on the spatial correlation of the road traffic spatial data, this study proposes a road traffic spatial data compression algorithm that uses LZW encoding. The contributions of this study can be effectively used for the road traffic spatial data compression of different road segments. Besides, the high spatial correlation roads are selected by PCA, which can also be used in transportation research. Further, the compression method can motivate some interesting ideas in transportation research field as well.

For the road segments with high spatial correlation, the proposed algorithm performs effectively. According to the reconstructed results of the HI3009b and HI7036b road segments, the algorithm is sensitive for correlation. The stronger the correlation is, the better the performance of the algorithm is. Thus, to ensure improved performance, the cross correlation should be greater than 0.95. Then, the expected compression ratio can be obtained.

Considering the remarkable performance of the proposed algorithm, we will explore the traffic state compression based on the spatial-temporal correlations in our next study.

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

TABLE 13: The sensitive analysis results on June 25 on HI7058b.

Data size	180 (0:00–6:00)	360 (0:00–12:00)	540 (0:00–18:00)	720 (0:00–24:00)
CR	3.27	10.59	12.86	15.65
AE	4.19	5.92	6.43	6.83
marerr (%)	10.00	8.16	7.22	7.07
σ	5.26	7.76	8.39	8.87

TABLE 14: The sensitive analysis results on June 26 on HI7058b.

Data size	180 (0:00–6:00)	360 (0:00–12:00)	540 (0:00–18:00)	720 (0:00–24:00)
CR	2.09	4.86	15	11.80
AE	4.29	6.51	7.32	7.41
marerr (%)	11.70	9.60	9.03	8.51
σ	5.60	9.14	9.84	7.79

TABLE 15: The indicators results of PCA and LZW encoding on June 19, 2011.

Road segment	CR		AE		marerr		σ	
	PCA	LZW	PCA	LZW	PCA	LZW	PCA	LZW
HI3009b	2.01	9.73	7.43	11.28	8.10	14.82	9.94	13.43
HI7036b	2.01	5.71	9.43	9.76	13.31	12.11	11.82	12.65
HI7058b	2.01	16.00	7.15	6.93	7.30	7.67	9.23	9.33

Acknowledgments

This work was supported by the Zhejiang Provincial Natural Science Foundation (Grant no. LQ16E080012), the National Natural Science Foundation of China (Grant no. 6157331), and Open Fund for a Key-Key Discipline of Zhejiang Province (2015001).

References

- [1] E. I. Vlahogianni, B. B. Park, and J. W. C. Van Lint, "Big data in transportation and traffic engineering," *Transportation Research Part C: Emerging Technologies*, vol. 58, p. 161, 2015.
- [2] Q. Shi and M. Abdel-Aty, "Big Data applications in real-time traffic operation and safety monitoring and improvement on urban expressways," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 380–394, 2015.
- [3] Z. Zhang, Q. He, H. Tong, J. Gou, and X. Li, "Spatial-temporal traffic flow pattern identification and anomaly detection with dictionary-based compression theory in a large-scale urban network," *Transportation Research Part C: Emerging Technologies*, vol. 71, pp. 284–302, 2016.
- [4] K. Sayood, *Introduction to Data Compression*, Elsevier Press, Newnes, 4nd edition, 2012.
- [5] G.-H. Ahn, Y.-K. Ki, and E.-J. Kim, "Real-time estimation of travel speed using urban traffic information system and filtering algorithm," *IET Intelligent Transport Systems*, vol. 8, no. 2, pp. 145–154, 2014.
- [6] X.-H. Yao, F. B. Zhan, Y.-M. Lu, and M.-H. Yang, "Effects of real-time traffic information systems on traffic performance under different network structures," *Journal of Central South University of Technology (English Edition)*, vol. 19, no. 2, pp. 586–592, 2012.
- [7] Z. Q. Zhao, Y. Zhang, J. M. Hu et al., "Comparative study of PCA and ICA based traffic flow compression," *Journal of Highway and Transportation Research and Development*, vol. 25, no. 11, pp. 109–113, 2008.
- [8] D. W. Xu, H. H. Dong, H. J. Li, L. M. Jia, and Y. J. Feng, "The estimation of road traffic states based on compressive sensing," *Transportmetrica B*, vol. 3, no. 2, pp. 131–152, 2015.
- [9] Q.-Q. Li, Y. Zhou, Y. Yue, and A. G.-O. Yeh, "Compression method of traffic flow data based on compressed sensing," *Journal of Traffic and Transportation Engineering*, vol. 12, no. 3, pp. 113–126, 2012.
- [10] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: from theory to applications," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4053–4085, 2011.
- [11] B. Li, J. Z. Xie, and B. L. Wang, "Signal reconstruction based on compressed sensing," *Computer Technology and Development*, vol. 19, no. 5, pp. 23–25, 2009.
- [12] Y. Xiao, L.-Y. Lu, S. Gao, Y.-M. Xie, and D.-Y. Xu, "Traffic data denoising compression for intelligent traffic systems based on 2-D discrete wavelet transformation," *Beifang Jiaotong Daxue Xuebao/Journal of Northern Jiaotong University*, vol. 28, no. 5, pp. 1–5, 2004.
- [13] X. L. Ou, J. T. Ren, and Y. Zhang, "A neural network mode for urban volumes compression," *Cybernetics and Informations*, vol. 1, no. 4, 2003.
- [14] G. J. Xu and H. Wang, "Implementation of locating data real-time compression of embedded GPS system with car," *Information Technology*, vol. 4, p. 14, 2006.
- [15] Q.-L. Ma, W.-N. Liu, and D.-H. Sun, "A high-speed compression scheme for vast quantities of GPS data," *Journal of Sichuan University*, vol. 43, no. 1, pp. 123–128, 2011.
- [16] Q.-Z. Wang, K. Wang, and Z.-S. Yang, "Coding algorithm of traffic flow in intelligence guidance system based on adaptive

- switching mode,” *China Journal of Highway and Transport*, vol. 6, article no. 14, 2009.
- [17] M. Hou, *QoS Management with Differentiated Services IP over the Internet*, Kingston, Ontario, Canada, 1999.
- [18] R. Song, W. Sun, B. Zheng et al., “RESS: a novel framework of trajectory compression in road networks,” *Proceedings of the VLDB Endowment*, vol. 7, no. 9, pp. 661–672, 2014.
- [19] J. Dong, Y. Zhang, Z. Zhang, and X.-X. Kuang, “Principal component analysis of dependency of urban intersections,” *Journal of Southwest Jiaotong University*, vol. 38, no. 6, pp. 619–622, 2004.
- [20] Y. Pan and J. Zhang, “Traffic-flow prediction based on dependency analysis of urban intersections,” *Transportation and computer*, vol. 23, no. 1, pp. 31–34, 2005.
- [21] A. Saltelli, “Sensitivity analysis for importance assessment,” *Risk Analysis*, vol. 22, no. 3, pp. 579–590, 2002.