IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received April 29, 2020, accepted June 4, 2020, date of publication June 12, 2020, date of current version June 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3001934

Vessel AIS Trajectory Online Compression Based on Scan-Pick-Move Algorithm Added Sliding Window

SHUANG SUN[1], YAN CHEN[1], (Member, IEEE), ZAIJI PIAO[2], AND JINSONG ZHANG[1]
[1]School of Maritime Economics and Management, Dalian Maritime University, Dalian 116026, China
[2]School of Marine Electrical Engineering, Dalian Maritime University, Dalian 116026, China

Corresponding author: Jinsong Zhang (jinsong_zhang@dlmu.edu.cn)

This work was supported in part by the China Scholarship Council under Grant 201806570020, in part by the National Natural Science Foundation of China under Grant 71271034, and in part by the National Natural Science Foundation of Liaoning Province under Grant 20180550963.

ABSTRACT The trajectory data of vessel AIS (automatic identification system) has important theoretical and application value for information supporting decisions. However, large sizes lead to difficulties in storing, querying, and processing. To solve the problems of high compression ratio and longtime consumption of the existing online trajectory compression algorithm, an SPM (scan-pick-move) trajectory data compression algorithm added sliding window is proposed. In order to better compress vessel trajectory data regarding compression efficiency, the sliding window is added to the classical SPM algorithm. In order to reduce trajectory data storage space, the maximum offset distance reference trajectory point is used as the criterion of whether the current trajectory point can be compressed. In this paper, the multi-dimensional space-time characteristics of trajectory data, such as distance error, compression ratio and compression time, are selected to evaluate the trajectory compression method from three levels: geometric characteristics, motion characteristics and compression efficiency. Compared with the existing SPM trajectory data compression algorithm, parallel experiments are conducted based on AIS data gathered over the duration of a month in the Japan Osaka Bay. The SPM trajectory compression algorithm added sliding window can significantly reduce the compression time and outperforms other existing trajectory compression algorithms in term of average compression error at high compression strengths. Also, the proposed method has high compression efficiency in the range of commonly used compression thresholds.

INDEX TERMS AIS data, compression ratio, scan-pick-move, sliding window, vessel trajectory compression.

I. INTRODUCTION

The automatic identification system (AIS) is becoming more and more convenient and efficient in ship navigation, monitoring and traffic flow research. According to SOLAS Convention, since 2002, more ships have been forced to install AIS equipment [1]. After nearly ten years of construction, the basic network framework of coastal AIS base stations have been formed. With the increase of the number of shipborne terminals, the frequent transmission of equipment information and the improvement of the collection base station, a large number of vessels AIS trajectory data have been generated. Such abundant AIS data can open up various research directions including visualization for detecting spatial distribution regularities in vessels, abnormal detection for maritime control, collision avoidance decision making and environmental protection [2]–[7]. A common task for all the AIS data studies corresponds to preprocessing massive historical records. An AIS message is transmitted by a ship at frequent intervals of approximately 2s–6min as shown in Table 1 [8]. However, the frequency of the clear change in speed and course is significantly lower than the recording rate. Consequently, most information in the raw AIS data is redundant in AIS trajectories that consist of massive similar trajectory points. To reduce the cost of storage and computing in data processing and to satisfy the response time requirement, compression of the vessel trajectory data is usually used before the detailed applications [9]–[11]. Trajectory data

The associate editor coordinating the review of this manuscript and approving it for publication was Yanli Xu.

TABLE 1. Upload time interval of AIS data.

Ship motion state	Time interval
Anchoring or berthing with moving speed no greater than 3kn	3min
Anchoring or berthing with moving speed greater than 3kn	10s
0~14kn	10s
0 ~ 14kn and change course	$3\frac{1}{3}$s
14~23kn	6s
14~23kn and change course	2s
>23kn	2s
>23kn and change course	2s

compression refers to the use of detection to eliminate redundant points in trajectory points. After eliminating redundant compression, the potential data mining speed of the same trajectory data is greatly improved.

One of the key bottlenecks of trajectory big data is the limitation of trajectory analysis, mining and application due to the massive data scale [12]–[15]. The purpose of trajectory data compression is to remove redundant location points on the premise of retaining the information contained in the data, so as to reduce the amount of data and the storage space occupied by the data. But at the same time of simplifying data points, a certain amount of information will be lost and some risks will be brought. Therefore, various methods of trajectory data compression are the focus of trajectory big data research. The classic algorithm in the field of trajectory compression is the Douglas-Peucker (DP) algorithm proposed by Douglas in 1973 [16]. DP algorithm is an off-line batch processing algorithm. It needs to collect the complete trajectory and then compress the data. In the field of vessel trajectory compression, most scholars improve the Douglas-Peucker algorithm. Etienne [17] *et al.* reduced the number of positions of a trajectory by adopting the DP algorithm filter while retaining only the virtual positions. Their purpose was to optimize the calculation time of traffic flow pattern recognition. They did experiments on 104201 records of 506 trajectories, with a compression ratio of 84.54% at a threshold of 10 meters. De Vries and Van Someren [18] used DP algorithm to retain the stop and move information on the ship's trajectory better. Zhang [19] *et al.* proposed a method to determine the maximum value to select a suitable threshold compression algorithm to ensure that the simplified point is within the threshold range. Zhao and Guoyou [20] proposed a path simplification method combining vessel trajectory and DP algorithm, which significantly reduced the compression time under high compression strength. Singh *et al.* [21] put forward a kind of SPM (scan-pick-move) algorithm in 2017. SPM algorithm uses the idea of DP algorithm to adjust the baseline selection method of error measurement and compares the error of trajectory points in turn. Han [22] *et al.* put forward to transform trajectory into space path and time series, and compress data in space and time at the same time. Most of the research on the compression algorithm which scientists and practitioners focus on are the road vehicle trajectory, offline compression, fixed graphics, terrain boundary line and graphics display [23]–[27], while the research on the online compression of vessel trajectory data is relatively less. Hershkovits and Ziv [28] add non-progressive coding to the sliding window algorithm to reduce the system resource consumption and compress the data.

With the improvement of AIS infrastructure and the development of AIS technology, AIS big data has become a development trend, and will eventually reach a level that computers can not directly store, which is exactly the biggest drawback of offline compression algorithm, only the data that has been collected completely can be compressed. Therefore, in this paper, according to the characteristics of vessel trajectory data, the online compression algorithm is used to improve the existing SPM compression algorithm. Our contributions can be summarized as follows:

(1) One compression technique is proposed, namely ISW-SPM (scan-pick-move Algorithm added Sliding Window), in order to maintain a flexible and condense representation of AIS trajectories.
(2) The maximum offset distance reference trajectory point is used as the criterion of whether the current trajectory point can be compressed.
(3) Compression results of single vessel trajectory in different compression thresholds are compared in the Sliding Window, SPM, ISW, and ISW-SPM algorithms.
(4) An extensive experimental study based on AIS data set in Osaka Bay provides concrete evidence that even after applying compression trajectories, high-quality feature points may still be attained in approximate evaluation by adding sliding window. In the data set, compression ratio, compression time and average error of multiple trajectories compression in different compression thresholds are also compared in the SPM, ISW, and ISW-SPM algorithms.

The remainder of this paper proceeds as follows. In section 2, the method for simplifying vessel trajectory based on the Sliding Window algorithm, SPM algorithm and SPM algorithm added Sliding Window are proposed. In section 3, basic information of AIS data is introduced and show the experimental results. Section 4 gives the discussion about the experimental results of three compression algorithms, and the conclusions are discussed in section 5.

II. MATERIALS AND METHODS

AIS data is complex and difficult to extract. In the analysis and research of the moving trajectory, only the information points that accurately describe the moving trajectory need to be stored, which are called the feature points. Other data points can be simplified, which is the basic idea of the trajectory data compression. AIS data preprocessing is the premise of trajectory data compression.

A. AIS DATA PREPROCESSING

AIS data includes ship MMSI, UTC time, longitude, latitude, speed, course, ship type, AIS berth type and other dynamic and static data. This paper takes ship longitude and latitude

TABLE 2. Conditions of segment data [29].

Condition name	Value	Condition name	Value
Upper bound of speed threshold	40kn	Lower bound of speed threshold	15kn
Upper bound of differential threshold	0	Lower bound of differential threshold	15
Global speed multiplier	5	Piecewise speed multiplier	7
Global difference multiplier	5	Piecewise difference multiplier	7
Minimum points of segment	7	Maximum points of segment	31
Minimum duration of segment	20s	Maximum duration of segment	5 min

data as the research object to carry out the research of vessel AIS trajectory compression algorithm.

1) TRAJECTORY SEPARATION

The specific process of ship trajectory separation is divided into two steps: separating the data of different ships; separating the data of different trajectories of the same ship. MMSI number is the only identification of a ship, which can be used as the basis for separating different ships; the ship's trajectory is not continuous in different time periods, and the same ship's different trajectory can be separated according to AIS data timestamp.

2) TIME INTERVAL STANDARDIZATION

There are standard specifications for the time interval of AIS data transmission; however, there are many cases that the time interval of collected AIS data is inconsistent with the standards. In addition, there are few ships equipped with high-cost gyrocompass in the inland river, and the change of AIS message transmission time caused by the change of ship course is rare according to the statistical results of AIS data in Osaka Bay and the time interval of AIS data transmission.

3) DATA CLEANING

The longitude and latitude data of vessel AIS are derived from GPS. In GPS relative positioning, the observation value may have errors. In the previous research [29], the abnormal points of trajectory have been eliminated according to conditions in Table 2. This paper will not repeat the cleaning algorithm. Only after cleaning, the data can be compressed to ensure the accuracy of compression.

After data preprocessing, a large number of trajectory segments can be got and Figure 1 is the display of randomly extracted trajectory of 100 trajectories in Osaka Bay.

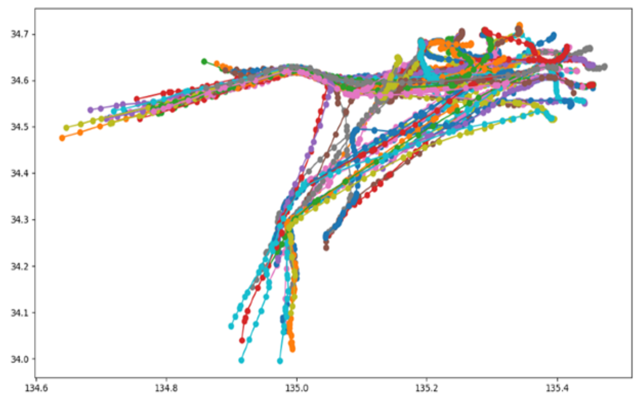


FIGURE 1. Trajectories in Osaka Bay.

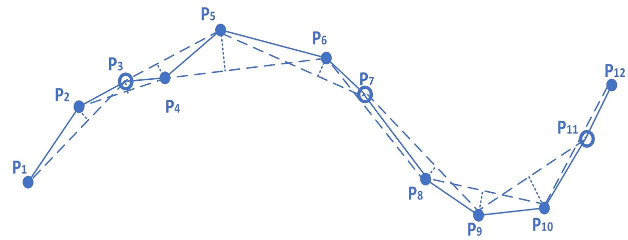


FIGURE 2. Classic Sliding Window compression algorithm.

B. CLASSIC SLIDING WINDOW COMPRESSION ALGORITHM

The core of the classic Sliding Window compression algorithm is to process only three points all the time, and use the idea of gradual compression to transmit the data in the form of a stream for online compression [30]. For example, the initial window is (P_1, P_2, P_3) and P_1 is the initial point, which is set as the key feature point to keep; calculate the Euclidean vertical distance between the point P_2 and the line between two endpoints; if it is greater than the specified threshold value, the P_2 is marked as the key feature point, the window slides and adds P_4, the current window is updated to (P_2, P_3, P_4), and P_4 becomes the new sliding starting point to gradually compress; calculate the Euclidean vertical distance between P_3 and P_2P_4 line again and if it is less than the preset threshold distance, discard the middle point P_3 and add P_5; update the current window to (P_2, P_4, P_5); and so on, keep P_4, P_5, P_6, delete P_7, and update the window to (P_6, P_8, P_9). Figure 2 shows the classic sliding window compression algorithm.

C. SPM COMPRESSION ALGORITHM

This algorithm is developed in the process of finding a faster off-line compression algorithm. SPM stands for Scan, Pick and Move. The SPM algorithm works in Figure 3 [21]: only on the trajectory data whose points are sorted in order of time, can it work like AIS trajectory; draw a straight line $P_1 - P_{12}$ between the first and last points of the trajectory; all the intermediate points are scanned in sequence, and only the point whose vertical distance d_2 from the line is greater than the threshold value δ are selected; the line moves to a new position by joining the newly picked point P_2 and the end point of the trajectory; and like this, $d_3 < \delta$, $d_5 > \delta$, all the intermediate points (d_3) between the start and the picked point are discarded; the line continues to move. Repeat processing until all points are scanned. The start point, all pick points, and end point define the compressed trajectory. The algorithm runs in linear time $O(n)$.

D. SPM COMPRESSION ALGORITHM ADDED SLIDING WINDOW

The classic compression algorithms usually study the vehicle trajectory, display graphics and terrain boundary lines,

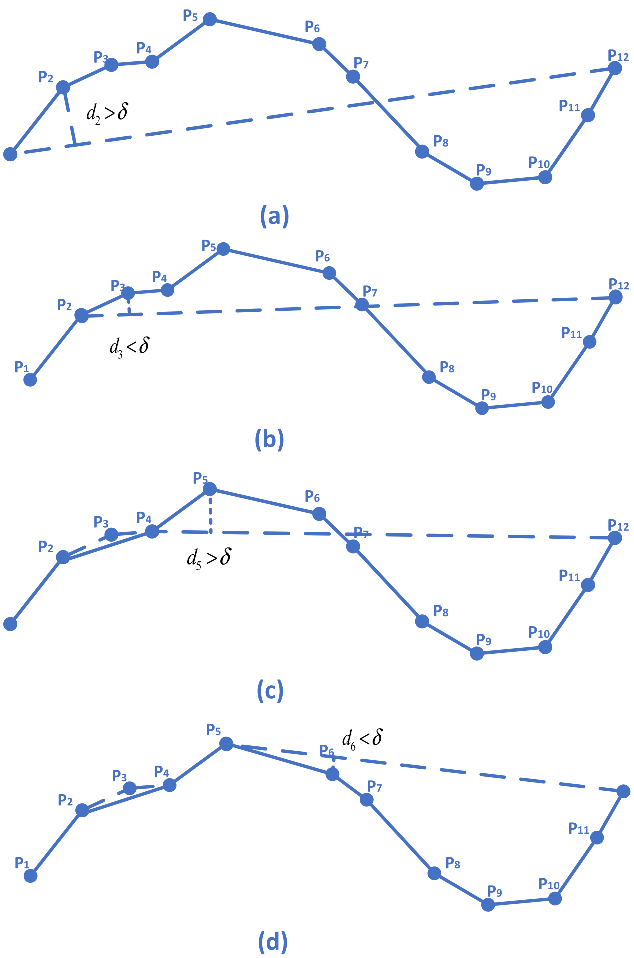


FIGURE 3. SPM compression algorithm.

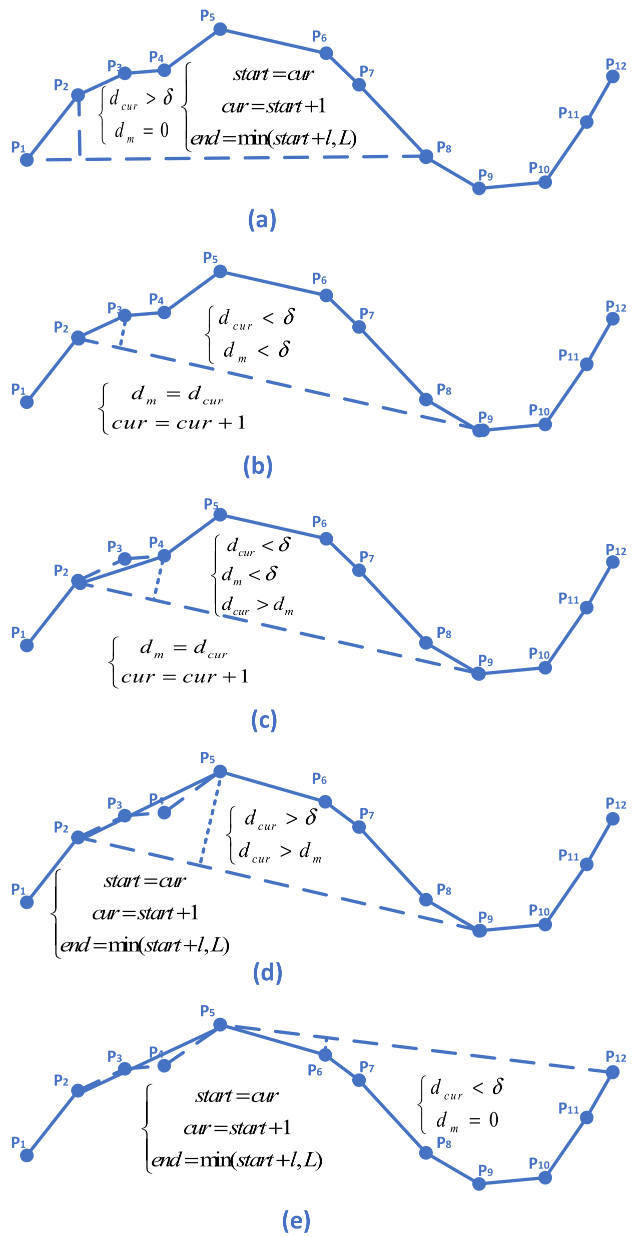


FIGURE 4. ISW-SPM compression algorithm.

only considering the compression of position information, while the ship itself is sailing in viscous liquid, with high uncertainty and poor maneuverability. In addition, there are also many characteristics such as turning limit, speed limit, AIS information loss. Whether it is the Douglas-Pecuker trajectory data compression algorithm, the classic sliding window trajectory data compression algorithm or the SPM compression algorithm, the vertical distance of the trajectory point offset from the trajectory direction is used as the selection standard of the trajectory feature points. However, when there is a change between the start trajectory point and the end trajectory point, it is necessary to calculate the vertical distance from all trajectory points to the starting and ending trajectory points, which increases the complexity and affects the efficiency of the algorithm [31]. To solve this problem, the maximum offset distance reference trajectory point in the sliding window is proposed as the criterion of whether the current trajectory point to be compressed can be compressed. In this way, when there is a change between the start trajectory point and the end trajectory point of the sliding window, only the vertical distance between the reference trajectory point of the maximum offset distance and the current trajectory point to be compressed and the straight line between the start trajectory point and the end trajectory point need to be calculated, so as to significantly reduce the complexity of the algorithm and the compression processing time. In the algorithm, the vertical distance from the point to the line need to be calculated. Take d_{cur} calculation of vertical distance from current point P_{cur} to line $P_{start} - P_{end}$ as an example. According to the principle of plane analytic geometry, the calculation formula of vertical distance d_{cur} is as follows:

$$d_{cur} = \sqrt{(x_{P_{cur}} - x_{P'_{cur}})^2 + (y_{P_{cur}} - y_{P'_{cur}})^2} \tag{1}$$

where P'_{cur} is the vertical coordinate from P_{cur} to the line $P_{start} - P_{end}$, and its position coordinate is given by (2) and (3) [32]:

$$x_{P'_{cur}} = \frac{x_{P_{cur}} + K \cdot (y_{P_{cur}} - y_{P_{start}}) + K^2 \cdot x_{P_{start}}}{1 + K^2} \tag{2}$$

$$y_{P'_{cur}} = x_{P_{cur}} + K \cdot (x_{P'_{cur}} - x_{P_{start}}) \tag{3}$$

where, $K = \frac{y_{P_{end}} - y_{P_{start}}}{x_{P_{end}} - x_{P_{start}}}$, similarly, maximum offset distance d_m can be calculated.

As shown in Figure 3, the sliding window of current trajectory data compression is $W = (P_{start}, P_{cur}, P_{end}, P_m)$, where P_{end} depends on the given window size l and the number of trajectory points L, that is $end = \min(start + l, L)$. Calculate the vertical distance d_{cur} and d_m from points P_{cur} and P_m to line $P_{start} - P_{end}$ respectively. If d_{cur} or d_m is greater than the thresholdδ, it means that the trajectory point is far away from the trajectory direction, and the trajectory cannot be approximately fitted, then add P_{cur} to the trajectory set Q. And then make $P_{start} = P_{cur}$, meanwhile set a new sliding window. Otherwise, P_{start} does not change, P_{cur} of the sliding window is moved back one point along the trajectory sequence, and whether P_m needs to be updated is determined according to the size of d_{cur} and d_m. For a new sliding window, repeat the above process until the compression process is finished. In this way, in the processing of each sliding window, it is not necessary to calculate the vertical distance from other trajectory points between P_{start} and P_{end} to the straight line $P_{start} - P_{end}$. Although there will be some distance errors as shown in formula (6), it greatly improves the efficiency of the algorithm and reduces the complexity of the algorithm.

By introducing the maximum offset distance point P_m as the reference trajectory point of the SPM algorithm, this paper provides the discrimination condition of feature point selection, and sets the corresponding threshold δ of trajectory compression distance according to the actual situation. If the current point P_{cur} to be compressed meets any of the following conditions, this point will be taken as the feature point:

Condition 1: the vertical distance from P_m to line $P_{start} - P_{end}$ is greater than the given threshold δ

Condition 2: the vertical distance from P_{cur} to line $P_{start} - P_{end}$ is greater than the given threshold δ

The pseudo code of SPM algorithm based on adding sliding window is summarized in Algorithm 1:

Time complexity analysis: the time complexity of the algorithm is $O(n)$, and n is the total number of trajectory points to be compressed.

Algorithm 1 *Compress*()

```
Input: trajectory set to be compressed P = {P_i}, threshold
of trajectory compression distance δ, window size l,the
number of trajectory points L
Output: List of feature points Q
{
  Q = {P_1}
  Sliding Window W = (P_start, P_2, P_l, P_m) ,
  start = 1, end = 2, P_m = 0
Do
    d_cur = vertical distance (P_cur, line P_start − P_end)
    d_m = vertical distance     (P_m, line P_start − P_end)
    if (d_cur > L or d_m > L) // judge feature points
    then
      Q = Q ∪ {P_cur} // add P_cur in feature point list
    W = {P_cur, P_start+1, P_start+l, 0} // set a new sliding
window
 end if
 if (d_cur ≤ Landd_m ≤ L)
   if (d_cur > d_m) then
     P_m = P_cur // update P_m based on distance
   end if
end if
}while (end > n)
return Q
```

TABLE 3. Transcoded AIS data.

DATE	TIME	messageID	MMSI	SOG/kn	LON/deg	LAT/deg
2018/5/1	0:00:02	1	431006467	10.9	135.2502633	34.69016167
2018/5/1	0:00:02	1	636091407	11	135.1633017	34.59303167
2018/5/1	0:00:03	1	565694000	11.8	134.9719217	34.10307333
2018/5/1	0:00:04	1	431000613	11.4	135.31627	34.62358
2018/5/1	0:00:08	1	431003155	12.5	135.097225	34.20200167
2018/5/1	0:00:09	1	441568000	13.3	135.0946333	34.59061667

III. RESULTS

A. EXPERIMENTAL DATA SET AND EXPERIMENTAL ENVIRONMENT

In order to verify the performance and effectiveness of the SPM compression algorithm added Sliding Window, more than 300 ships of Japan Osaka Bay are used to collect data sets over a period of time. The single vessel trajectory is the AIS data randomly selected from a ship in the data set between 2018-01-01,00:00:00 and 2018-02-01,00:00:00, and the ship's MMSI (Maritime Mobile Service Identify) is 431000331. The ship position, ship speed and other information in AIS data will be greatly changed due to equipment, signal drift and other reasons. In order to ensure the availability of data, it is necessary to preprocess the data. After transcoding and preprocess, the data format needed in this paper is shown in Table 3 [36]. For multiple vessel trajectories, 300 ships with more frequent activities are extracted from the data set to obtain the motion trajectory data for one month, with a total of 1402015 trajectory points. The experiment is carried out in Windows 10 system, the algorithm is written in Python language, version 3.7, developed in PyCharm IDE.

B. EXPERIMENTAL EVALUATION CRITERIA

In order to evaluate the compression effect, the most common performance evaluation methods are used: compression time T, compression ratio R, compression error E [33]. They are defined as follows:

1) COMPRESSION TIME T

Time before compression T_1 and time after compression T_2.

$$T = T_2 - T_1 \tag{4}$$

2) COMPRESSION RATIO R

The number of trajectory points before compression M and after compression N.

$$R = (1 - \frac{N}{M}) \times 100\% \tag{5}$$

3) DISTANCE ERROR E

The average distance between the non-feature point and its two adjacent feature points, l_k is the distance between the non-feature point k and the adjacent feature points.

$$E = \frac{1}{M-N} \sum_{k=1}^{k=M-N} l_k \tag{6}$$

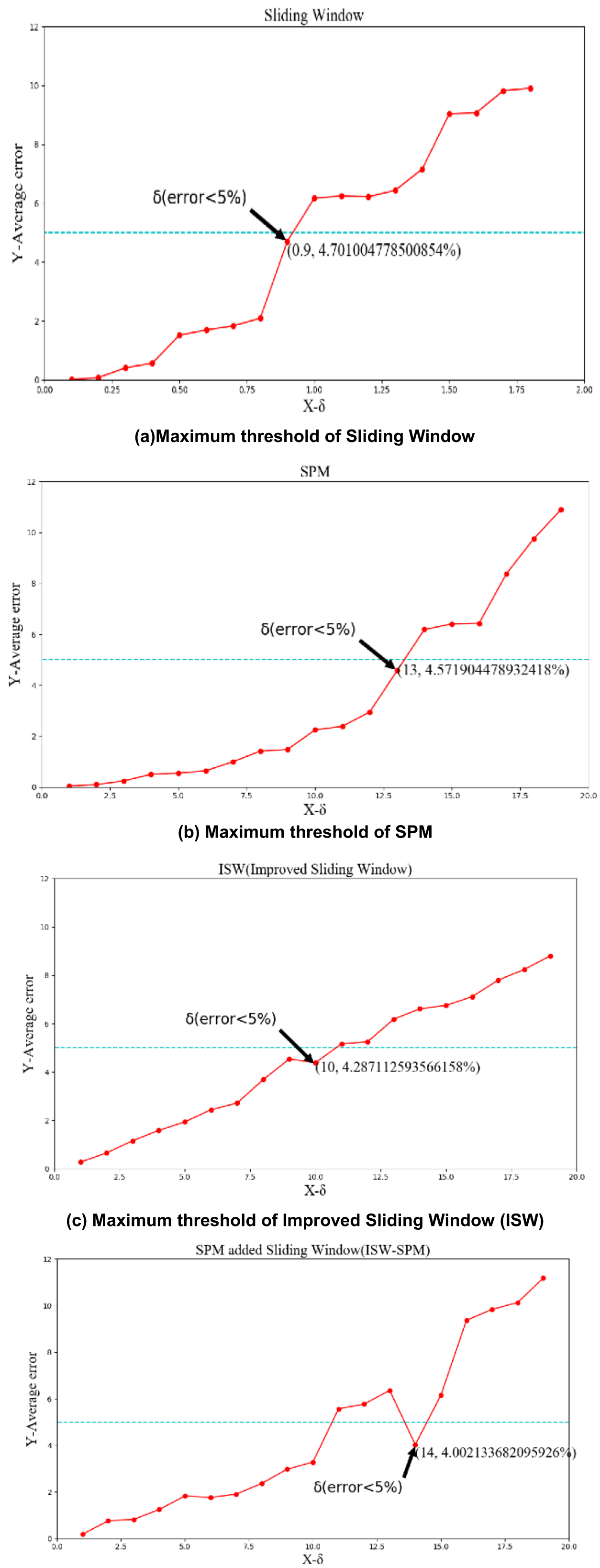
C. COMPRESSION EFFECT

1) COMPRESSION RESULTS OF SINGLE VESSEL TRAJECTORIES

According to experience, the maximum threshold which satisfies the error ratio of less than 5% is selected, then the four thresholds of ship 431000331's single trajectory are (0.9,4.70%), (13,4.57%), (10,4.29%), (14,4.00%). The results show that the threshold value and the average error are linear on the whole, but not strictly linear growth. The selections of threshold values are shown in the Figure 5 and 6.

Next, take the maximum thresholds of four algorithms to compress the trajectory of ship 431000331. The compression results of single vessel trajectory are shown in the figure 6.

It can be seen from figure 6 that after the compression of single vessel trajectory, the SPM algorithm added sliding window can obtain better compression accuracy. The number of original trajectory points and various compression algorithms are as follows: the number of original trajectory points is 597, the number of trajectory points compressed by the Sliding Window algorithm is 192, points compressed by the SPM algorithm is 397, points compressed by the improved Sliding Window algorithm is 93, and points compressed by the SPM algorithm adding sliding window is 154. In addition, the compression time of the four algorithms are as follows: 174ms (sliding window algorithm), 9.3ms (SPM algorithm), 269ms (improved sliding window algorithm), 18ms (SPM algorithm with sliding window added). The number of compression points of the SPM algorithm is larger than that of Sliding Window algorithm, although the algorithms have the same time complexity, the compression time of the SPM



(a)Maximum threshold of Sliding Window

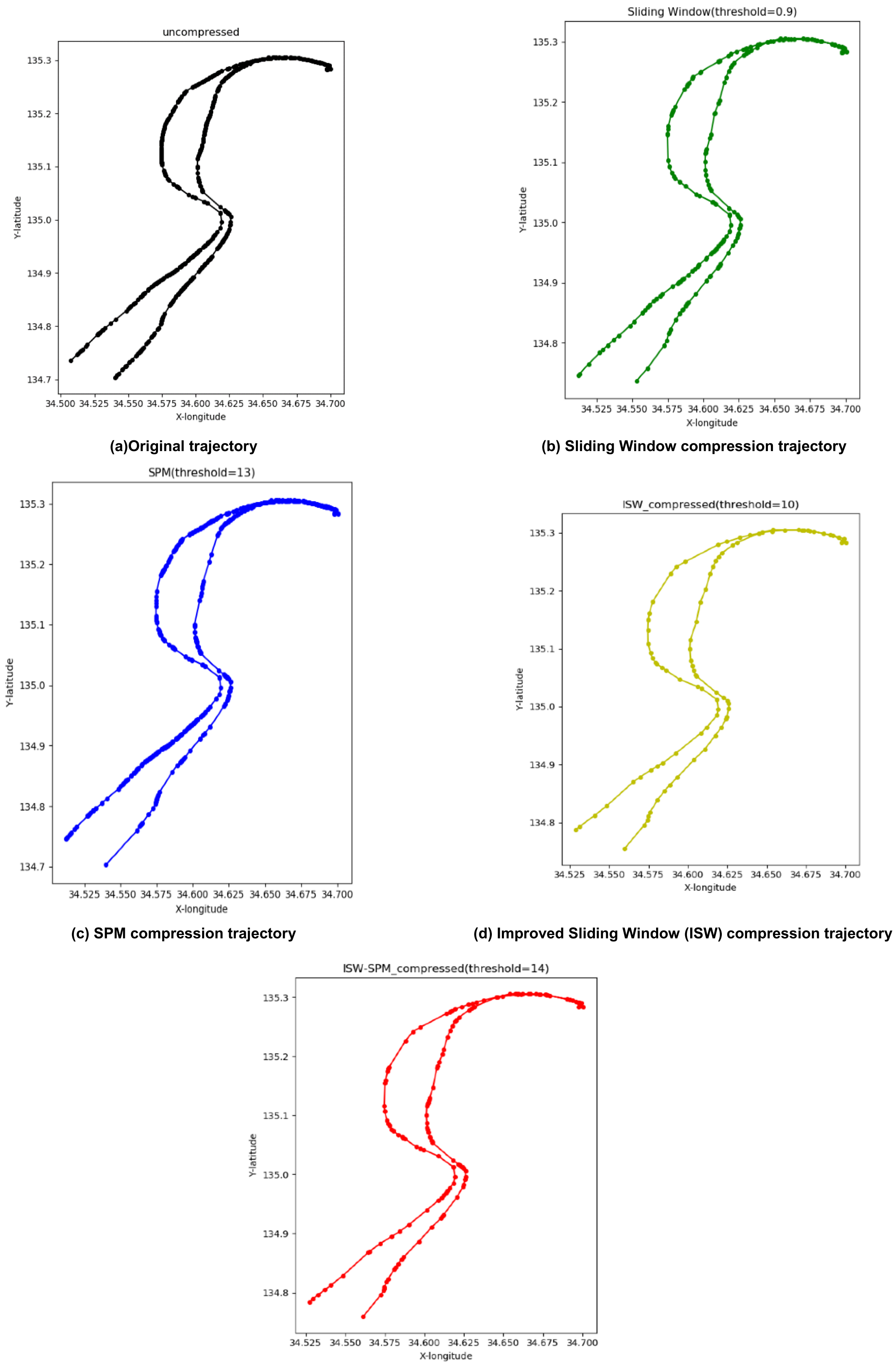
(b) Maximum threshold of SPM

(c) Maximum threshold of Improved Sliding Window (ISW)

(d) Maximum threshold of SPM added Sliding Window (ISW-SPM)

FIGURE 5. Selection of maximum threshold value(X/m, Y/m).

algorithm is much shorter than that of the Sliding Window algorithm. The comparison results are shown in Table 4.



(a)Original trajectory

(b) Sliding Window compression trajectory

(c) SPM compression trajectory

(d) Improved Sliding Window (ISW) compression trajectory

(e) SPM added Sliding Window (ISW-SPM) compression trajectory

FIGURE 6. Compression results of single vessel trajectory(X/deg, Y/deg).

The improved sliding window algorithm has the highest accuracy and better reflects the original trajectory characteristics, but the compression time is longer than the other three algorithms. The proposed algorithm in this paper

TABLE 4. Experimental results of single vessel trajectory.

Algorithm	Original points	Points after compression	Compression time
Sliding Window	597	192	174ms
SPM	597	397	9.3ms
ISW	597	93	269ms
ISW-SPM	597	154	18ms

TABLE 5. Experimental results under different thresholds.

Algorithm	Threshold / m	Points after compression	Compression ratio /%	Compression time/ms	Average error / m	Maximum error/m
SPM	5	1000039	28.7	36956	1.0	37.9
SPM	10	833781	40.5	30819	2.6	98.4
SPM	20	809333	42.3	30418	7.0	192.0
SPM	30	733875	47.7	30348	12.3	344.1
SPM	40	702541	49.9	27257	18.0	500.1
SPM	50	691166	50.7	27309	24.6	770.0
ISW	5	309160	77.9	39773	1.2	19.9
ISW	10	210841	85	36654	2.6	26.7
ISW	20	144700	89.7	30028	5.5	135.6
ISW	30	116207	91.7	28541	8.3	168.7
ISW	40	98877	92.9	26343	11.1	188.2
ISW	50	86939	93.8	25261	13.6	386.1
ISW-SPM	5	449600	67.9	40235	1.3	64.7
ISW-SPM	10	317514	77.4	34811	2.7	95.8
ISW-SPM	20	215680	84.6	33298	5.2	140.4
ISW-SPM	30	169503	87.9	28572	7.5	195.0
ISW-SPM	40	141216	89.9	27286	9.5	224.1
ISW-SPM	50	122492	91.3	25441	11.3	249.4

reduces part of the compression time but obtains better compression accuracy and good compression.

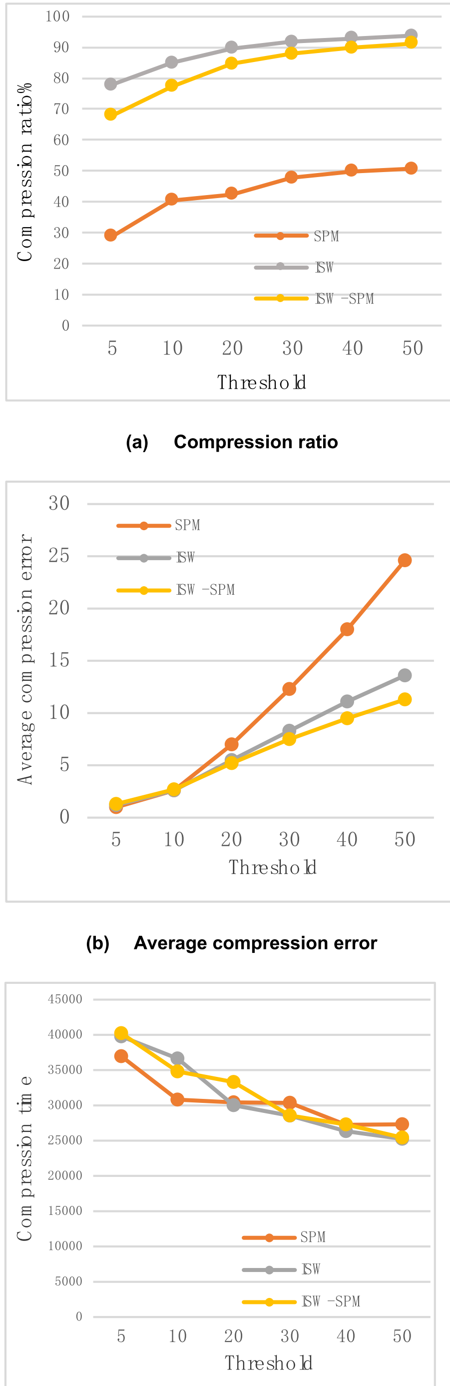
2) COMPRESSION RESULTS OF MULTIPLE VESSEL TRAJECTORIES

In order to comprehensively analyze the consistency of error loss of trajectory compression algorithm on multiple compression scales, multiple scales of 5m, 10m, 20m, 30m, 40m and 50m are selected as the compression threshold parameters. The applicable algorithms are SPM algorithm, improved Sliding Window algorithm and SPM algorithm added sliding window. Table 5 shows the results of different algorithms under different compression thresholds. The compression time T, compression ratio R, compression error E are calculated according to (4-6):

It can be seen from Table 5 that after the compression of multiple vessel trajectories, different thresholds have great influence on the three algorithms.

IV. DISCUSSION

Figure 7 is the three algorithms comparison diagram of multiple vessel trajectories. The compression ratios are expected that the less the number of points is after the compression, the higher the compression ratio would be. The compression ratios of the three algorithms go up with the increase of threshold value, whereas the compression ratio of the ISW algorithm and ISW-SPM algorithm is relatively high and stable. The compression ratios of SPM algorithm are almost a half that of ISW and ISW-SPM. From the results of single vessel compression, we also can find that the ISW-SPM algorithm can still maintain a good trajectory shape.



(a) Compression ratio

(b) Average compression error

(c) Compression time

FIGURE 7. Comparison of three algorithms in multiple vessel trajectories.

The error of the SPM algorithm increases rapidly with the increase of threshold value, especially from 10m. The errors of the ISW algorithm and ISW-SPM algorithm are relatively low. The algorithm proposed in this paper has the smallest and most stable error change rate with the increase of threshold value. Moreover, from the results of single ship compression, the ISW-SPM algorithm can still maintain a good trajectory shape.

In terms of compression time, when the threshold value is 5m, the compression time of the SPM algorithm is significantly lower than that of the other two algorithms. With the increase of threshold, the compression time of the three algorithms is reduced. Among them, the ISW-SPM algorithm has a slow descent rate, but the descent rate does not change much. When the threshold value is 40m, the compression time of the three algorithms is similar. With the increase of the threshold value, the compression time of the ISW algorithm and ISW-SPM algorithm is significantly lower than that of the SPM algorithm. Compared with the high compression ratio of other algorithms, the proposed algorithm not only obtains better compression accuracy, but also has greater advantages in compression time compared with the other algorithms. The compression time of the algorithm is about 6.7% lower than that of the SPM algorithm on 50m threshold, because the algorithm constantly compares points in the dynamic threshold stage. For the maximum error, it can be found that the error increases faster with the increase of threshold value and the maximum errors of the SPM and ISW algorithm do not accord with the reality when the threshold value are 40m and 50m. By comparing the compression ratio, average compression error and compression time of the three algorithms, it can be found that the algorithm inherits the good compression time of the SPM algorithm, and the online compression of trajectory needs to analyze and process the data online, so the lower compression time makes the algorithm have better effect, and the reduction of compression ratio makes the data storage become the feature point that can accurately reflect the trajectory style.

V. CONCLUSIONS

(1) Aiming at the online compression of vessel trajectory, this paper improves the classic SPM (scan-pick-move) compression algorithm, which combines the sliding window, and applies it to the vessel AIS trajectory data compression, improves the operation efficiency of the computer, and provides a theoretical basis for the future processing of the vessel AIS trajectory big data, ship motion pattern recognition, ship motion feature extraction, single vessel behavior research.

(2) For the improved compression algorithm of SPM, window length and distance threshold are recommended. Users can choose different compression thresholds according to their different compression requirements. The recommended threshold can be used as the initial recommended threshold for compression, which needs to be further adjusted. A related opportunity for future work is to devise methods for optimal threshold selection to achieve the best trade-off between accuracy and compression ratio for a specific ship.

(3) The experimental results show that the algorithm can get a better trajectory pattern, improve the compression efficiency and reduce the compression time. Compared with the classic Sliding Window algorithm, classic SPM algorithm, the improved Sliding Window algorithm, although the proposed algorithm's compression ratio is generally higher than that of the improved sliding window algorithm, the compression processing efficiency of the proposed algorithm is better. In the aspect of big data processing, AIS trajectory data online processing, real-time compression and calculation efficiency, SPM algorithm is not as good as the proposed algorithm, and the distortion rate of ship trajectory is higher, AIS trajectory number is higher according to the behavior characteristics of the ship, the retention capacity is insufficient.

(4) As a good research object of big data, AIS data contains rich ship navigation laws. With the continuous improvement of AIS data collection technology, AIS data magnitude will increase explosively. Mining AIS big data reasonably can fill the blank of previous expert experience. In the future, when dealing with AIS data and extracting the behavior features, the multi-dimensional characteristics of ship navigations behavior can be used to make the algorithm more targeted to extract the key feature points of vessel trajectory, and improve the compression ratio and efficiency while reducing the compression risk.

REFERENCES

[1] W. Bing, "AIS application and prospects," *China Water Transp.*, vol. 12, no. 7, pp. 70–92, 2012.
[2] R. J. Bye and A. L. Aalberg, "Maritime navigation accidents and risk indicators: An exploratory statistical analysis using AIS data and accident reports," *Rel. Eng. Syst. Saf.*, vol. 176, pp. 174–186, Aug. 2018.
[3] L. Zhu, S. Haker, and A. Tannenbaum, "Flattening maps for the visualization of multibranched vessels," *IEEE Trans. Med. Imag.*, vol. 24, no. 2, pp. 191–198, Feb. 2005.
[4] D. Chen, X. Wang, Y. Li, J. Lang, Y. Zhou, X. Guo, and Y. Zhao, "High-spatiotemporal-resolution ship emission inventory of China based on AIS data in 2014," *Sci. Total Environ.*, vol. 609, pp. 776–787, Dec. 2017.
[5] L.-Z. Sang, A. Wall, Z. Mao, X.-P. Yan, and J. Wang, "A novel method for restoring the trajectory of the inland waterway ship by using AIS data," *Ocean Eng.*, vol. 110, pp. 183–194, Dec. 2015.
[6] L. Goldsworthy and B. Goldsworthy, "Modelling of ship engine exhaust emissions in ports and extensive coastal waters based on terrestrial AIS data—An Australian case study," *Environ. Model. Softw.*, vol. 63, pp. 45–60, Jan. 2015.
[7] G. Pallotta, M. Vespe, and K. Bryan, "Vessel pattern knowledge discovery from AIS data: A framework for anomaly detection and route prediction," *Entropy*, vol. 15, no. 12, pp. 2218–2245, Jun. 2013.
[8] F. Xiao, H. Ligteringen, C. van Gulijk, and B. Ale, "Comparison study on AIS data of ship traffic behavior," *Ocean Eng.*, vol. 95, pp. 84–93, Feb. 2015.
[9] F. X. Zhu, L. M. Miao, and W. Liu, "Research on vessel trajectory multi-dimensional compression algorithm based on Douglas–Peucker theory," *Appl. Mech. Mater.*, vol. 694, pp. 59–62, Nov. 2014.
[10] K. Xu, J.-Y. Qiu, and Y. Li, "Offline efficient compression algorithm for AIS data retains time elapsing dimension," *Comput. Sci.*, vol. 44, no. 11A, pp. 508–512, 2017.
[11] L. P. Perera and B. Mo, "Ship performance and navigation data compression and communication under autoencoder system architecture," *J. Ocean Eng. Sci.*, vol. 3, no. 2, pp. 133–143, Jun. 2018.

[12] W. Luo, H. Tan, L. Chen, and L. M. Ni, "Finding time period-based most frequent path in big trajectory data," in *Proc. Int. Conf. Manage. Data-SIGMOD*, New York, NY, USA, 2013, pp. 713–724.

[13] M. Li, X. Li, and J. Yin, "TORD problem and its solution based on big trajectories data," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 6, pp. 1666–1677, Jun. 2016.

[14] S. Fei, X. Zhang, L. Tang, Z. Liu, X. Yang, and K. Dong, "Temporal and spatial distribution of high efficiency passengers based on GPS trajectory big data," *J. Geo-Inf. Sci.*, vol. 17, no. 3, pp. 329–335, 2015.

[15] A. Belhassena and H. Wang, "Trajectory big data processing based on frequent activity," *Tsinghua Sci. Technol.*, vol. 24, no. 3, pp. 317–332, Jun. 2019.

[16] D. H. Douglas and T. K. Peucker, "Algorithms for the reduction of the number of points required to represent a digitized line or its caricature," *Can. Cartographer*, vol. 10, no. 2, pp. 112–122, 1973.

[17] L. Etienne, T. Devogele, and A. Bouju, "Spatio-temporal trajectory analysis of mobile objects following the same itinerary," *Adv. Geo-Spatial Inf. Sci.*, vol. 10, pp. 47–57, Jul. 2012.

[18] G. K. D. de Vries and M. van Someren, "Machine learning for vessel trajectories using compression, alignments and domain knowledge," *Expert Syst. Appl.*, vol. 39, no. 18, pp. 13426–13439, Dec. 2012.

[19] S.-K. Zhang, Z.-J. Liu, Y. Cai, Z.-L. Wu, and G.-Y. Shi, "AIS trajectories simplification and threshold determination," *J. Navigat.*, vol. 69, no. 4, pp. 729–744, Jul. 2016.

[20] L. Zhao and G. Shi, "A method for simplifying ship trajectory based on improved Douglas–Peucker algorithm," *Ocean Eng.*, vol. 166, pp. 37–46, Oct. 2018.

[21] A. K. Singh, V. Aggarwal, P. Saxena, and O. Prakash, "Performance analysis of trajectory compression algorithms on Marine Surveillance data," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Mysore, India, 2017, pp. 1074–1079.

[22] Y. Han, W. Sun, and B. Zheng, "COMPRESS: A comprehensive framework of trajectory compression in road networks," *ACM Trans. Database Syst.*, vol. 42, no. 2, pp. 1–49, Jun. 2017.

[23] A. Dobrkovic, M.-E. Iacob, and J. van Hillegersberg, "Maritime pattern extraction and route reconstruction from incomplete AIS data," *Int. J. Data Sci. Anal.*, vol. 5, nos. 2–3, pp. 111–136, Mar. 2018.

[24] J. M. Mou, C. van der Tak, and H. Ligteringen, "Study on collision avoidance in busy waterways by using AIS data," *Ocean Eng.*, vol. 37, nos. 5–6, pp. 483–490, Apr. 2010.

[25] L. Wu, Y. Xu, Q. Wang, F. Wang, and Z. Xu, "Mapping global shipping density from AIS data," *J. Navigat.*, vol. 70, no. 1, pp. 67–81, Jan. 2017.

[26] K. Georgios, N. Pelekis, and Y. Theodoridis, "Trajectory compression under network constraints," *Adv. Spatial Temporal Databases*, vol. 5644, pp. 392–398, Jul. 2009.

[27] J.-Y. Zhang, W. Xu, W. Zhang, X. Meng, and Y. Zhang, "A novel compression algorithm for infrared thermal image sequence based on K-means method," *Infr. Phys. Technol.*, vol. 64, pp. 18–25, May 2014.

[28] Y. Hershkovits and J. Ziv, "On sliding-window universal data compression with limited memory," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 66–78, Jan. 1998.

[29] S. Shuang, C. Yan, and Z. Jinsong, "Trajectory outlier detection algorithm for ship AIS data based on dynamic differential threshold," *J. Phys., Conf. Ser.*, vol. 1437, no. 1, pp. 12–13, 2020.

[30] G. Miao, S. Guo-You, and L. Wei-Feng, "Online compression algorithm of AIS trajectory data based on improved sliding window," *J. Traffic Transp. Eng.*, vol. 18, no. 3, pp. 222–231, 2018.

[31] L. I. Xiaoming, G. U. Yupei, and Z. Juntao, "GPS trajectory map matching algorithm based on sliding window," *J. Xi'an Technol. Univ.*, vol. 6, pp. 459–462, 2017.

[32] J.-G. Wu, L. Min, W. Guang, and L. L.-F., "An improved trajectory data compression algorithm of sliding window," *Comput. Technol. Develop.*, vol. 25, no. 12, pp. 53–57, 2015.

[33] X. Qu, Q. Meng, and L. Suyi, "Ship collision risk assessment for the Singapore strait," *Accident Anal. Prevention*, vol. 43, no. 6, pp. 2030–2036, Nov. 2011.

SHUANG SUN was born in Zhalantun, China, in 1991. She received the B.S. degree in information management and information system and the M.S. degree in management science and engineering from Dalian Maritime University, Dalian, China, in 2013 and 2016, respectively, where she is currently pursuing the Ph.D. degree. Her research interests include big data processing and analysis, data mining, clustering analysis, and maritime safety.

YAN CHEN (Member, IEEE) was born in 1952. She graduated from the Dalian Maritime College majoring in ship electrical engineering, in 1976. She is currently a Professor and the Ph.D. Supervisor with the School of Maritime Economics and Management, Dalian Maritime University. She mainly engaged in modern transportation management, decision support systems, knowledge management, data warehouse, and data mining.

ZAIJI PIAO was born in Zhalantun, China, in 1990. He received the B.S. degree in automation from Dalian Ocean University, Dalian, China, in 2013, and the M.S. degree in control engineering from Dalian Maritime University, Dalian, in 2016, where he is currently pursuing the Ph.D. degree. His research interests include ship motion control, advanced control theory, and maritime science and safety.

JINSONG ZHANG was born in Tianjin, China, in 1985. He received the Ph.D. degree in management science and engineering from Dalian Maritime University, Dalian, China, in 2013. He is currently a Lecturer with the School of Maritime Economics and Management, Dalian Maritime University. His research interests include text mining, big data processing and analysis, and knowledge management.