Ocean Engineering 166 (2018) 37–46

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng

A method for simplifying ship trajectory based on improved Douglas–Peucker algorithm

Liangbin Zhao*, Guoyou Shi

Navigation College, Dalian Maritime University, 1, Linghai Road, Dalian, Liaoning, China

ARTICLE INFO

Keywords:
Water traffic
Ship trajectory
Automatic identification system (AIS)
Trajectory compression
Douglas–peucker algorithm

ABSTRACT

Automatic identification system (AIS) can provide massive ship trajectory data that is valuable for mining information in water traffic. However, large sizes lead to difficulties in storing, querying, and processing the aforementioned data. In the present study, to better compress ship trajectory data regarding compression time and efficiency, a method based on the improved Douglas–Peucker (DP) algorithm is presented. In the process of compression, the proposed method considers the shape of vessel trajectory derived from course information of track points. Parallel experiments are conducted based on AIS data gathered over the duration of a month in the Chinese Zhou Shan islands. The results indicate that this method can effectively compress ship trajectory information. Additionally, when compared with the traditional DP algorithm, this method can significantly reduce the compression time and exhibits better performance at high compression strengths. Also, the proposed method outperforms other existing trajectory compression algorithms in term of compression time.

1. Introduction

Recently, the value of ship trajectory data has increased. Thousands of ships sail around the world daily. Their mobility results in water traffic, which is a phenomenon that shows the behavioral patterns of ships. These patterns may support managers in the field of maritime supervision and management.

Automatic identification system (AIS) is an automatic tracking system for identifying and locating ships by exchanging data with other nearby ships and AIS base stations. Increasing research in the maritime field is focused on AIS data mining given the fast development of AIS terminal network, data storage, and data collection capacity. Several studies based on AIS data aim at various purposes including visualization for detecting spatial distribution regularities in ships, abnormal ship identification for maritime control, and decision-making for collision avoidance. Wu et al. (2016) created the maps of shipping density on a global scale with over 2.5 years of AIS data. They computed the average number of vessels crossing each grid per unit time, and 21,162,882,025 pieces of AIS data costs 16 h. Pallotta et al. (2013) presented an unsupervised and incremental learning approach for extracting vessel movement patterns. Their aim involved converting a large amount of AIS data into information supporting decisions. Two months of AIS traffic data in Strait of Gibraltar and three months of satellite AIS position data over the Indian Ocean are collected for pattern training and testing. Mou et al. (2010) used AIS data in studies on ship collision avoidance in busy waterways. In the study, a period corresponding to 62 days of data during June, July, and August 2007 in the North Sea was analyzed to calculate important parameters such as the closest point of approach (CPA) and time to the CPA (TCPA). Zhen et al. (2017) utilized AIS data and clustering method to obtain the clusters of encounter vessels and then constructed the semantic and mathematical relationship of vessel collision risk index for each cluster of encounter vessels with DCPA (Distance to Closest Point) and TCAP.

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 3–10 s. 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 track 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 application.

Qi and Zheng (2016) proposed a method for recognizing the vessel course alteration and extracting the representative point of a ship trajectory. Spiliopoulos et al. (2017a) described an approach, which is based on clustering technique and distributed processing technique, to transform billions of records of spatiotemporal (AIS) data into

* Corresponding author.
E-mail address: vszlb@126.com (L. Zhao).

https://doi.org/10.1016/j.oceaneng.2018.08.005
Received 15 September 2017; Received in revised form 23 July 2018; Accepted 6 August 2018

information for understanding the patterns of global trade. Patroumpas et al. (2017) presented a system to extract trajectory synopses from the incoming AIS positions retaining salient movement features only. Their method can instantly identify "critical point" along each trajectory, such as a stop, a turn, or slow motion. Therefore they may discard redundant locations along a "normal" course, and approximately reconstruct each vessel trajectory from such a synopsis consisting of critical points only. A ship trajectory in the small area is considered as line data in two-dimensional plane. The Douglas–Peucker (DP) algorithm (Douglas and Peucker, 1973; Ramer, 1972) was considered as one of the most accurate and effective methods to compress line data (Meratnia and Rolf, 2004; Muckell et al., 2010, 2014) and was widely adopted in compression of moving object trajectory.

In terms of DP-based compression of vessel trajectory, Etienne et al. (2012) used the DP algorithm filter to reduce the number of positions of a trajectory while retaining only the important positions. Their purpose involved optimizing the computation time of the traffic flow pattern identification. They conducted an experiment based on 104,201 records of 506 trajectories, and the compression rate was 84.54% for a threshold of 10 m (the precision of a GPS device). Li et al. (2010) proposed a method that considered the time attribute of ship trajectory. They considered a ship trajectory as line data in three-dimensional space. An experimental result based on 30,000 records indicated that the method can retain additional feature points that may preserve more information about the speed of trajectory. Zhu et al. (2014) conducted a similar study. De Vries and Van Someren (2012) adapted the DP algorithm to better retain stop and move information on the ship trajectory. Zhang et al. (2016) presented a method to select a suitable threshold for AIS trajectory based on determining the maximum value that can ensure that the track point of simplified trajectory is within the safety scope of the corresponding original track point. Thus, the calculated ship domain in the navigational situations is used as the threshold value. After the experiment based on 5,902,840 records from 962 ships in Chinese Qiong Zhou Strait, the result revealed that the compression rate corresponded to 98.25% with a threshold set as 0.8 times the ship length. Li et al. (2016) conducted several experiments to select an appropriate threshold that can guarantee a good balance between AIS trajectory simplification and visualization quality. The experimental AIS data included 29,015 coordinate points from 187 vessels.

Running time and compression rate are the two important indicators of compression result. Compression at high strength costs less time and retains less data, and compression at low strength costs more time and retains more data. In practical applications, the requirement varies to a large extent for different users. For example, the identification of ship dynamic parameters and statistical analysis of maritime traffic requires more details in the data, while the data in visualization analysis should be minimized to the maximum possible extent under the condition of maintaining data quality. Additionally, studies on mapping trajectory with different scales exhibit different requirements of data quality. Furthermore, while dealing with large datasets, the running time is an important consideration factor that should not be neglected by users.

However, previous studies do not sufficiently discuss the problem of running time and compressed data quality at high compression strengths in DP algorithm. To better apply DP algorithm to ship trajectory compression under the requirements in which the algorithm may utilize considerable time or a high compression strength, this study proposes a method that incorporates the course change in the ship trajectory and DP algorithm. When compared with the performances of the traditional DP algorithm, the proposed method can significantly reduce the running time at the same level of compression strength and additionally exhibits improved compressed data quality at high compression strength. Also, the proposed method outperforms other existing trajectory compression algorithms in term of compression time. Additionally, ship trajectory data, (which is based on the geographic coordinate system) cannot be used directly in the DP algorithm, (which is based on the Cartesian coordinates system). In the proposed method, the Mercator projection coordinate is applied to solve this problem.

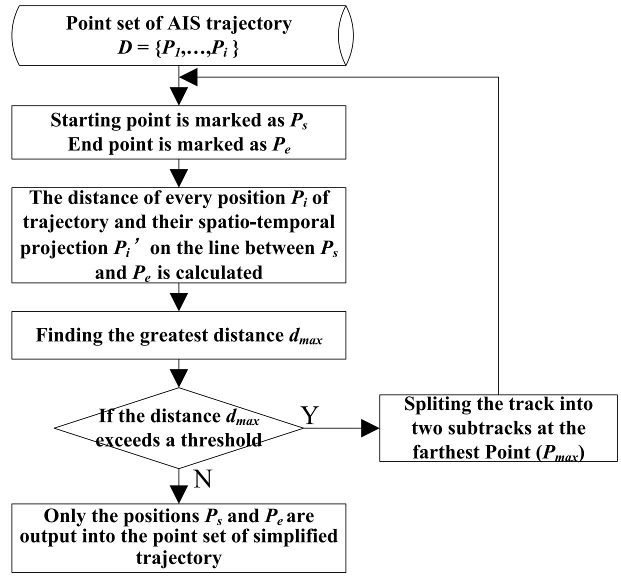

Fig. 1. DP algorithm theory.

The remainder of this paper is organized as follows. In section 2, the method for simplifying ship trajectory based on DP algorithm is introduced. In section 3, we introduce AIS data and the basic information on the data source in the study. Section 4 shows the experimental results, and the conclusions are discussed in section 5.

2. Simplification method for ship trajectory

2.1. Douglas–Peucker algorithm

2.1.1. Algorithm theory

The DP algorithm was presented by D. Douglas and T. Peucker in 1973. The algorithm splits the line data recursively and controls the compression quality by the threshold, and it is widely used in simplifying the trajectory of moving point objects due to its speed and accuracy.

The algorithm theory is illustrated in Fig. 1. An AIS trajectory is represented as a point set $D = \{P_1, \ldots, P_i\}$, as shown in Fig. 2(a). The maximum distance d_{max} between each point P_i of the trajectory and its projection P_i' on the line between the starting point P_s and end point P_e is calculated as shown in Fig. 2(b). The farthest point P_{max} is retained if the distance d_{max} exceeds a threshold. Subsequently, the trajectory is split at this position (P_{max}) as shown in Fig. 2(c). The algorithm is recursively applied to both trajectory subparts. As shown in Fig. 2(d) and (e), only points P_s and P_e of the trajectory subparts are retained if the distance d_{max} is lower than the threshold.

2.1.2. Application in ship trajectory

As described above, the data in the application of DP algorithm corresponds to plane Cartesian data. However, it is not possible to directly use the ship trajectory data (which is typically based on the geographic coordinate system) in the DP algorithm. The calculation of spherical distance is complicated. More importantly, it is difficult to calculate the distance between a track point and a line based on the geographic coordinate system. To calculate point-to-line, we convert the geographical coordinates to the coordinates in Mercator projection. (λ, φ) denote the geographical coordinates of a track point. The coordinates in Mercator projection of the track point (X, Y) are calculated as follows: