

Article

Track Pairs Collision Detection with Applications to Ship Collision Risk Assessment

Jiahui Shi ^{1,2} and Zhengjiang Liu ^{1,2,*}

¹ Navigation College, Dalian Maritime University, Dalian 116026, China; dmunsgkl@163.com

² Key Laboratory of Navigation Safety Guarantee of Liaoning Province, Dalian 116026, China

* Correspondence: liuzhengjiang@dmlu.edu.cn

Abstract: The port waterway network plays an important role in the organization and management of port ship traffic. Due to limited ship operations, conflicts, congestion, and safety issues often arise in port waters. Conflicts between ships can be predicted by collision detection between ships. A novel collision detection algorithm for trajectory pairs is proposed by introducing variable time interval variables. In addition, to improve the overall accuracy of trajectory compression and reduce redundant calculation in collision detection, a multi-factor Douglas-Peucker algorithm adapted to ship trajectory compression is proposed with the consideration of speed and turn constraints. The maximum speed difference of the algorithm is increased by 1.5–2.5%, and the average speed difference increased by 2.0–4.5%. Based on the method mentioned above, the risk assessment framework of maritime collision is established and the risk situation of the waters near Ningbo Zhoushan Port is evaluated and analyzed by using ship historical track data.

Keywords: conflict detection; quantitative risk assessment; ship domain; trajectory compression



Citation: Shi, J.; Liu, Z. Track Pairs Collision Detection with Applications to Ship Collision Risk Assessment. *J. Mar. Sci. Eng.* **2022**, *10*, 216. <https://doi.org/10.3390/jmse10020216>

Academic Editor: Md Jahir Rizvi

Received: 7 January 2022

Accepted: 4 February 2022

Published: 6 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As the last nautical mile of ship transportation, the safety of navigation for vessels in the port waters directly affects the trade exchanges between countries [1]. To strengthen the safety of navigation in port waters, the International Maritime Organization (IMO) has established a traffic separation system and a ship reporting system. At the same time, countries have established a Vessel Traffic System (VTS) supervision system based on the characteristics of the water area. Figure 1 shows the management area of VTS near Ningbo-Zhoushan port and Yang-shan port, where there are several functional areas such as anchorage, fairway, alert area, and pilot embarking or disembarking area (point), et al. Ningbo-Zhoushan port and Yang-shan port both have become one of the biggest port all over the world, where the vessel traffic flow is very complex constrained by narrow waterways, islands, bridges, et al. Several managerial and navigational solutions such as the Traffic Separation Scheme (TSS) and VTS have been implemented in this area, however, the actual effectiveness of these solutions in generating navigational safety and the focus of subsequent navigational regulation still deserve further study.

Current ship collision risk evaluation theories and model studies can be classified according to macro and micro [2]. Macroscopic collision evaluation methods are mostly based on collision data, traffic flow statistics, and relevant hydro-meteorological data using statistical analysis to make a comprehensive evaluation of navigational safety in a certain range of waters. This study focuses on the evaluation of macroscopic collision in this water.

It can be subdivided into two categories, including the evaluation methods based on the number of collisions and collision rate and the evaluation methods based on the number of encounters and encounter rate. In [3], the authors had assessed the collision risk of the narrow waterway in Ningbo-Zhoushan with collision detection method and ship domain. This study first aligns and discretizes the trajectories, and then achieves collision

detection of trajectory pairs by detecting collisions between isolated trajectory points. It is easy to find that the efficiency will be low and the process is complicated.



Figure 1. The study area for collision risk assessment.

Regarding the study of conflict prediction methods within the channel, the authors of literature [4,5] classified conflicts into two categories, connecting segments and intersections, by introducing the concept of network. Literature [6] provides a literature review on the assessment of individual collision risk in ship navigation. Chai [7] proposed a quantitative risk assessment model with the consideration of human life loss and environmental pollution. Weng [8] made an estimation for the collision frequency of vessels in the Singapore Strait with the real-time vessel movement data. Goerlandt [9] developed a collision alert framework from the perspective of the concept of risk and the intended use of the model. Furthermore, the authors also introduced a method for measuring the ship collision risk, which was successfully applied to a case study. Debnath [10] made a study on the influential factors of collision risks and analyzed the associations between risks and geometric, traffic, and the regulation for controlling traffic in Singapore port fairways. Based on the ship domain safety areas derived by the VTS operator, Yoo [11] proposed a real-time collision risk assessment support system, whose validity was verified in the scenarios at Busan port.

Since macro collision risk evaluation is mostly obtained by statistical analysis based on historical big data, and the huge amount of data will bring some trouble to the research. In terms of compression of trajectory big data, the famous Douglas Peucker algorithm was proposed in [12]. Considering the importance of the velocity of a moving object in map-matching [13], mobility prediction, and moving pattern mining [14], Ying, et al. [15] proposed an algorithm for preserving velocity when simplifying trajectory.

The key of ship collision risk assessment is to calculate the number of conflicts between ships. At present, in the field of ship-based conflict detection research, most of the

continuous collision detection is carried out through the discrete collision detection method. Firstly, the continuous state of the trajectory pair is discretized, and then approximately continuous collision detection is performed by judging whether there is a collision between the finite states of the trajectory pair. Although its understanding and implementation are simple, its efficiency is very low. The detail for real-time collision detection can be found in the literature [16]. Tang [17] presented a novel continuous collision detection algorithm by introducing deforming non-penetration filters.

The purpose of this work is to evaluate the collision risk for vessels in specific water. The main contribution is divided into two parts. Firstly, the previous research work is improved and a new trajectory simplification method is designed. Secondly, a collision detection method between trajectories is proposed, which can effectively improve the efficiency of collision detection and reduce the difficulty of algorithm implementation. The framework of ship collision risk assessment is shown in Figure 2.

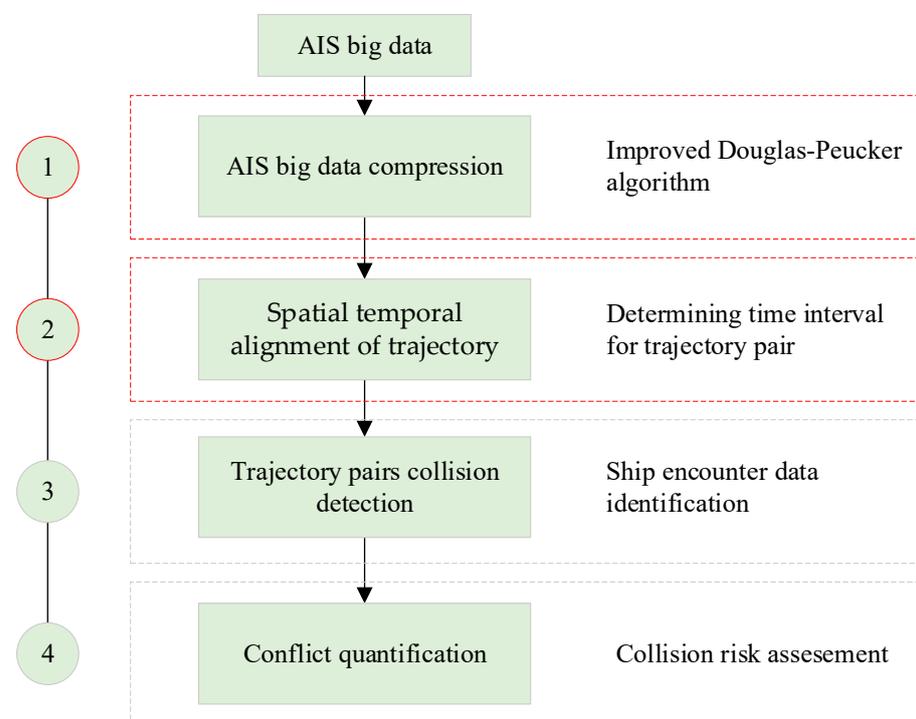


Figure 2. Framework of collision risk assessment.

2. Ship Collision Risk Assessment

2.1. AIS Compression Based on the Improved Douglas-Peucker Algorithm

Ship trajectory data is accumulated continuously through AIS equipment, and its broadcast interval can be as short as 5 s, so it will consume a lot of computing resources when analyzing ship collisions through massive AIS data. Especially for ships sailing in the waterway network, their navigation is limited by the direction of the waterway. According to statistics, in a ship trajectory, the key points to describe navigation behavior only account for about 8%. Therefore, we firstly compress the big AIS data. Therefore, our primary work is to compress AIS data.

Regarding the compression for AIS data, some works have been made in our previous work [18]. Considering that the historical trajectory of the ship contains information such as speed, position, heading, etc., this section improves the DP (Douglas-Peucker) algorithm by considering the influence of various factors on the compression algorithm. This section improves the DP algorithm considering the influence of ship speed based on Euclidean distance, realizes the compression and preprocessing of ship AIS data, reduces the number of AIS data, and improves the efficiency of ship collision detection.

2.1.1. Trajectory Preprocessing

We use the trajectory AIS data of Ningbo-Zhoushan, China in 2018, due to the original data is redundant, besides, the same ship may have many voyages into the same port, therefore, the AIS data should be preprocessed before grouping by voyage. The preprocessing steps are as follows (Algorithm 1).

Algorithm 1. Preprocessing AIS data

Require: AIS database *aisdata.db*, *threshold_time* = 30 min

```

1: Select distinct(MMSI) from aisdata.db squry_mmsi
2: For each sub_mmsi in squry_mmsi
3:   Select track from aisdata.db where MMSI = sub_mmsi
4:   If the number of track is less than 10 then continue
5:   Else For each point in track
6:     If the time difference time_diff between the next timestamp and the
       present timestamp is bigger than threshold_time then insert voyage into
       return_voyages, voyage set NULL
7:     Else insert point into voyage
8:     End If
9:     Insert voyage into return_voyages, voyage set NULL
10:    End for
11:  End If
12: End for
13: Return return_voyages

```

Due to the coordinate of trajectory are stored by longitude and latitude. Amis at the satisfaction of accuracy and efficiency of trajectory compression and clustering, there need to be transferred from geodetic coordinate system to Mercator. (φ, λ) refers to the geographical coordinates and Mercator coordinates is (x, y) , a refers to the long radius of the ellipse of the earth and e refers to the first eccentricity of earth ellipsoid that is the meridian parts at latitude, r_0 refers to the circle radius of base dimensions latitude. The formula for transferring coordinate from geographical to Mercator as Equation.

$$\begin{cases} r_0 = (a \times \cos \varphi) / \sqrt{1 - e^2 \sin^2 \varphi} \\ q = \ln \tan(\pi/4 + \varphi/2) - e/2 * \ln((1 + e \sin \varphi)/(1 - e \sin \varphi)) \\ x = r_0 \times \lambda \\ y = r_0 \times q \end{cases} \quad (1)$$

2.1.2. Multi-Factor Douglas-Peucker Algorithm

A trajectory is approximated by discrete samples of the AIS coordinate point. It can be shown a series of chronologically coordinate points, e.g., $T = \{p_1, p_2, \dots, p_n\}$, where each point consists of an ID, a coordinate set, and timestamp such as $P = \{MMSI, Longitude, Atitude, Course, Speed, Time\}$. DP algorithm [12] as one of the most accurate line simplification algorithms is proposed in 1973. The core concept of the DP compression method is to reduce the original trajectory point with compression threshold the perpendicular Euclidean distance of each remaining point. Before simplifying the trajectory data from the berthing data, there need to certify the compression threshold τ . DP algorithm recursively divides the curve composed of line segments. Initially, it is given all the points between the first and the last point. Firstly, the first and last points are kept, and then find the furthest from the line segment with the first and last. If the point is closer than τ to the line segment, then all points can be discarded, otherwise, that point must be kept. The algorithm recursively calls itself with the first and the furthest and then with the furthest and the last. The Ramer-Douglas-Peucker (RDP) algorithm, also known as the Douglas-Peucker algorithm and iterative endpoint fit algorithm, so in the following, RDP algorithm refers to DP algorithm.

Based on the DP algorithm and combined with speed and turning limitation, Multi-Factor Douglas-Peucker (MFDP) algorithm (Algorithm 2) is proposed. Before introducing the MFDP algorithm, we first describe the formal definitions for illustrating the parameters:

Definition 1. The perpendicular Euclidean distance $d_{\perp}(P_i)$ of each remaining point P_i is the perpendicular Euclidean distance $d_{\perp}(P_i, L_j)$ of a point P_i to line L . It can be calculated by the following Equation (2).

$$d_{\perp}(P_i) = d_{\perp}(P_i, L_j) \tag{2}$$

Definition 2. The azimuth $C_{i,j}$ of each remaining point P_i is the difference related to P_j . The azimuth difference $d_{\theta}(P_i)$ is the difference between the adjacent points. It is an important item to a unified coordinate system. Based on transformation by Equation (3) between geodetic coordinate system and Mercator, the formula of $d_{\theta}(P_i)$ is shown as Equation (4). In Equation (4), MP refers to meridian parts at latitude φ .

$$\begin{cases} MP = 7915.704471g \left[\tan \left(\frac{\pi}{4} + \frac{\varphi}{2} \right)^{e/2} \right] \\ C_{i,j} = \arctan \left(\frac{\lambda_j - \lambda_i}{MP(\varphi_j) - MP(\varphi_i)} \right) \end{cases} \tag{3}$$

$$d_{\theta}(P_i) = |C_{i+1,i} - C_{i,i-1}| \tag{4}$$

Definition 3. The speed difference $d_s(P_i)$ of each remaining point P_i is defined as follows. V_{i-1}, V_i, V_{i+1} refer to the speed of P_{i-1}, P_i, P_{i+1} respectively. t_{i-1}, t_i, t_{i+1} refer to the time of P_{i-1}, P_i, P_{i+1} , respectively.

$$\begin{cases} f_i(\theta) = \left(\left(\arctan \left(\frac{V_{i+1} - V_i}{t_{i+1} - t_i} \right) \right) - \left(\arctan \left(\frac{V_i - V_{i-1}}{t_i - t_{i-1}} \right) \right) \right) * 180 / \pi \\ d_s(P_i) = \begin{cases} f_i(\theta) + 360 & -360 < f_i(\theta) < -180 \\ |f_i(\theta)| & -180 < f_i(\theta) < 180 \\ 360 - f_i(\theta) & 180 < f_i(\theta) < 360 \end{cases} \end{cases} \tag{5}$$

Definition 4. MFDP algorithm aims at considerable and universal, so the above distance is part of structure distance $dist_i(P_i)$, which can be calculated by the following Equation (6), w_d, w_{θ} , and w_s refer to the weight of the perpendicular Euclidean distance, the angle distance, and the speed distance, respectively.

$$dist_i(P_i) = w_d d_{\perp}(P_i) + w_{\theta} d_{\theta}(P_i) + w_s d_s(P_i) \tag{6}$$

Due to the existence of weight of $dist_i(P_i)$ for simplification, therefore, the weight should be fixed when using. Under the certain necessities, the weight can be fixed differently as we need.

Algorithm 2. MFDP algorithm

Input: Track points list $D = [P_i, \dots, P_j]$, threshold $\epsilon = [\epsilon_d, \epsilon_{\theta}, \epsilon_s]$, weight $w = [w_d, w_{\theta}, w_s]$

Output:

- 1: **Function MFDP (D, i, j, e)**
 - 2: Find the point with maximum distance related to the baseline as d
 - 3: **If** d > dmax, split the track point list into two parts $D = [P_1, \dots, P_d], D = [P_d, \dots, P_j]$, recursive call
 - 4: **MFDP (D, i, d, e)**
 - 5: **MFDP (D, d, j, e)**
 - 6: **Else** output P_i, P_j
 - 7: **End if**
-

2.1.3. Error Analysis

The main measurements for the experimental evaluations are as follows, accuracy by turning angle SED_C and speed SED_V . The way of measuring the speed and turning angle difference between AIS trace and its compressed version is to measure the Synchronized Euclidean Distance (SED). SED measures the distance for any point in the track, this is the distance between its actual location and its synchronous position estimated via interpolation between its predecessor and successor points in the compressed track shown in Figure 3.

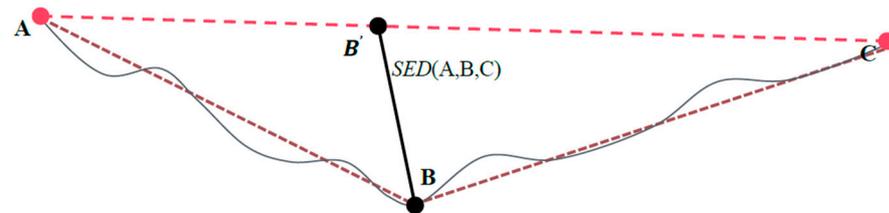


Figure 3. Synchronous Euclidean Distance. Where segment AC is the successive track points after compressing and point list [A,B,C] is the successive actual location of the track. Point B' is the generated synchronization point. $SED(A,B,C)$ is the Synchronous Euclidean distance between the point B' and B. The velocity of the Synchronized point B is $v'_B = v_A + v_{AC} \times (t_B - t_A) / (t_C - t_A)$, the time is $t'_B = t_B$ and the course is $c'_B = c_A + c_{AC} \times (t_B - t_A) / (t_C - t_A)$.

The Synchronous Euclidean Distance error of turning angle SED_C^i and speed SED_V^i for synchronization points are defined as follows.

$$SED_V^i = \sqrt{(v'_{ti} - v_{ti})^2 + (t'_i - t_i)^2} \quad (i = 1, 2, \dots, n) \tag{7}$$

$$SED_C^i = \sqrt{(c'_{ti} - c_{ti})^2 + (t'_i - t_i)^2} \quad (i = 1, 2, \dots, n) \tag{8}$$

The algorithm for generating synchronization points is presented in Algorithm 3.

Algorithm 3. Generating synchronization points

Subtrack $D = [P_1, \dots, P_j]$ is a points list of each trajectory in tracks; *marklabel* is the filtered point label list; *id* is each one filtered point label in *Subtrack*; *Outtrack* is the generated synchronization trajectory list; *L* is the number of points in *Outtrack*.

- 1: **For** each *submark* in *tracks* **do**
 - 2: **If** the number of *submark* is less than 3 then continue
 - 3: **Else**
 - 4: **For** each *id* in *marklabel* **do**
 - 5: **while** $L < id$ **do**
 - 6: $L++$
 - 7: Insert P_L into *Outtrack*
 - 8: Calculate the coordinate of synchronization point P_{id} under interpolation principle
 - 9: Insert P_{id} into *Outtrack*
 - 10: **End for**
 - 11: **End If**
 - 12: **End for**
 - 13: **Return** *Outtrack*
-

2.2. Collision Risk Assessment

At present, the research methods for ship collision accidents in a specific research water area within a specific period mostly refer to the results in the road and aviation fields. In the aviation field, a collision event is defined as the number of collisions within a specific period, which is also referred to as the frequency of collisions $f_{collision}$. It is

equal to the product of the number of collision incidents in a given time $N_{conflict}$ and the accident-causing coefficient $P_{causation}$.

$$f_{collision} = N_{conflict} \times P_{causation} \tag{9}$$

Note that the conflicts between ships can be divided into three categories based on the International Regulations for Preventing Collisions at Sea (COLREGS) [7], such as head-on, crossing, and overtaking conflict. All three types of conflict are distinguished through the course differences between ship pairs.

Overtaking conflict means that ship pairs are sailing on almost parallel courses and the course difference should not exceed 10° . Crossing conflict means that the course difference of ship pair falls in the range $10\text{--}170^\circ$ or $190\text{--}350^\circ$. Head-on conflict means that the course difference of ship pair falls in the range $170\text{--}190^\circ$ [7]. Furthermore, the necessary condition for the occurrence of the above three types of conflicts is that one ship will intrude into the safety domain of the other when both ships sail with their current course and speed.

Considering the three types of conflicts mentioned above, the collision risk can be assessed by the sum of the number of collisions that all types will encounter.

$$f_{collision} = N_{conflict}^h \times P_{causation}^h + N_{conflict}^c \times P_{causation}^c + N_{conflict}^o \times P_{causation}^o \tag{10}$$

where $N_{conflict}^h$, $N_{conflict}^c$, and $N_{conflict}^o$ are respectively the number of conflicts occurring in each scenario of head-on, crossing, and overtaking between ship pairs in a specific area within a certain period, $P_{causation}^h$, $P_{causation}^c$, and $P_{causation}^o$ are respectively the causation probability for the head-on, crossing, and overtaking conflicts.

As mentioned, if we want to evaluate the number of collisions, we need firstly calculate the number of conflicts. The basis for whether there is a collision risk between two ships is whether there is an intrusion in the safety domain of the two ships under relative motion. Regarding the research on instantaneous collision detection, as mentioned in the first section, the most widely used currently is to evaluate the collision risk through the nearest encounter point or the ship domain.

2.2.1. Spatial-Temporal Alignment of Trajectory

In the past, collision detection was performed by judging the relative position between ship fields. Since the trajectory of a ship is composed of multiple position points, which can be regarded as a discrete object, therefore the collision detection between the trajectories needs to frequently detect the state of the two ships. To avoid unnecessary detection, it is necessary to filter out the state where there is no possibility of a collision at all. Traditional methods include hierarchical bounding boxes, space segmentation, and GJK, but most of the above methods are collision screening of isolated objects, and there is still a problem of low efficiency in trajectory collision detection. In this study, the spatiotemporal collision detection between trajectory pairs is simplified by time intervals.

Different from the previous study in [3,7,8], regarding the calculation for the number of vessel conflicts, we determined a self-adaptive time interval from the trajectory pairs as shown in Figure 4. In the previous study mentioned above, the trajectory was discrete by a fixed time interval, such as the value was set as three minutes in [7,8], and 10 s was taken in [3]. However, since the launch of an AIS device follows the principle of time division multiple access, the timestamps of each track are not consistent. Therefore, the collision detection based on fixed time intervals would need to interpolate the track to achieve the purpose of timestamp alignment. Furthermore, there is no guarantee that there is no waypoint when further collision detection is performed on the trajectory pair of each time interval through relative motion.

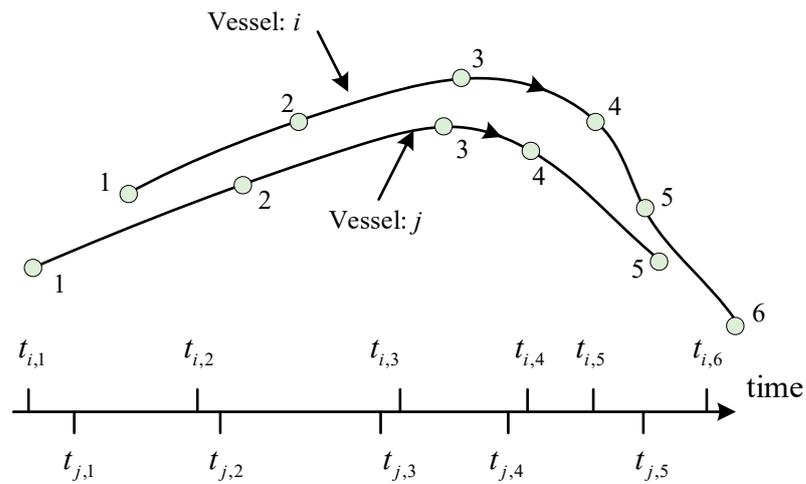


Figure 4. Determining the time interval from trajectory pairs.

The steps for determining the time interval for collision detection of trajectory pairs are as follows:

- (1) Determine the time intersection interval of the trajectory pair, if it exists, continue, otherwise exit.
- (2) Find the larger value of the trajectory to the start time.
- (3) Compare the magnitude between the two trajectories that are larger than the smallest value.
- (4) Use the smaller value in (2) and (3) as the collision detection time interval.
- (5) Set the smaller value in (3) as the larger value in (2), and repeat steps (2)–(5) until the end.

For example, in Figure 4, there are two trajectories labeled with a number from 1 to 5 and 1 to 6, respectively. Each trajectory has a series of waypoints and the arrival time at each waypoint was expressed as $t_{i,m}$, where m is the index of a waypoint. According to the mentioned process, due to the arrival time of vessel i is less than $t_{j,1}$ at waypoint 1, therefore, the left value is equal to $t_{j,1}$. Then, we can derive the right value $t_{i,2}$ by comparing the next adjacent arrival time $t_{i,2}$ and $t_{j,2}$. Finally, the time interval $[t_{j,1}, t_{i,2}]$ is obtained, subsequently, with the repetition of processes, the time interval set \mathcal{T} can be determined as $\mathcal{T} = \{[t_{j,1}, t_{i,2}], [t_{i,2}, t_{j,2}], \dots, [t_{i,5}, t_{j,5}]\}$. In this study, the radius of the safety domain R is various lengths of vessels, which is set as three times the length of vessel referred [8].

2.2.2. Trajectory Pairs Collision Detection

In the mentioned above, the time interval set has been determined, based on this foundation, we need to detect whether there is conflict at any time interval. Figure 5 is a conflict example for trajectory pair, where ship domain was utilized to detect collision between two vessels. If ship i enters the domain of another ship j during the time interval $[t_{i,4}, t_{i,5}]$, it is deemed as a conflict.

Whether an intrusion event occurs between ships i and j is determined here by the relative distance of movement and the size of the ship's domain as shown in Figure 6 where $D_{i,j}^t$ is the relative distance of ship pair at the time t , v_i^t is the speed of ship i at the time t , R_i is the safety domain of ship i , the position of the ship i at the time t is presented as $(\varphi_i^t, \lambda_i^t)$, $L_{i,j}^t$ is the distance of ship j can sail using the relative speed $v_{i,j}^t$ within the time interval Δt , $\theta_{i,j}^t$ is the angle between the relative speed and the connecting line of ship pairs. Some of the mentioned variables can be calculated as follows. Using Formula (1), we transferred the coordinate from geographical $(\varphi_i^t, \lambda_i^t)$ to Mercator (x_i^t, y_i^t) .

$$D_{i,j}^t = \sqrt{(x_i^t - x_j^t)^2 + (y_i^t - y_j^t)^2} \tag{11}$$

$$|\vec{v}_{ij}| = \sqrt{(|\vec{v}_i| \cos \alpha_i^t - |\vec{v}_j| \cos \alpha_j^t)^2 + (|\vec{v}_i| \sin \alpha_i^t - |\vec{v}_j| \sin \alpha_j^t)^2} \tag{12}$$

$$\theta_{i,j}^t = \arccos \left(1 - \frac{\left(\frac{|\vec{v}_i| \sin \alpha_i^t - |\vec{v}_j| \sin \alpha_j^t}{|\vec{v}_{ij}^t}| - \frac{x_j^t - x_i^t}{D_{i,j}^t} \right)^2 + \left(\frac{|\vec{v}_i| \cos \alpha_i^t - |\vec{v}_j| \cos \alpha_j^t}{|\vec{v}_{ij}^t}| - \frac{y_j^t - y_i^t}{D_{i,j}^t} \right)^2}{2} \right) \tag{13}$$

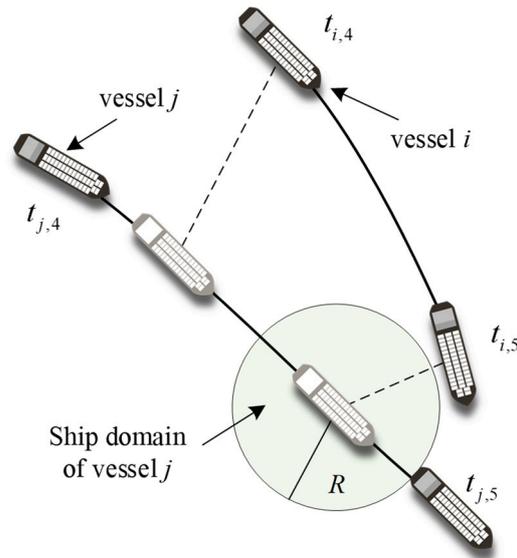


Figure 5. A conflict example for trajectories pairs.

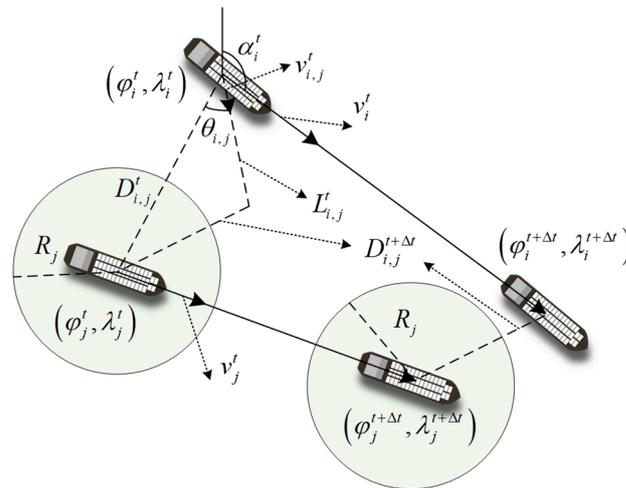


Figure 6. Relative motion of vessel pair at a time interval and ship conflicts.

Thus, the problem of ship collision risk assessment can be transformed into a statistical problem of the number of ship collisions in a specific area in a specific period. As shown in Figure 6, if $D_{i,j}^{t+\Delta t}$ is less than R_i , then the ship j will intrude the safety domain of the ship i in the time interval Δt .

$$D_{i,j}^{t+\Delta t} \leq R_j \rightarrow \sqrt{(x_i^{t+\Delta t} - x_j^{t+\Delta t})^2 + (y_i^{t+\Delta t} - y_j^{t+\Delta t})^2} \leq R_j \tag{14}$$

Assuming that two ships maintain their current course and speed during the time interval, Formula (14) can be transformed to

$$\begin{aligned} &\sqrt{\left(x_i^t - x_j^t + \left|\vec{v}_i\right| \sin \alpha_i^t \Delta t - \left|\vec{v}_j\right| \sin \alpha_j^t \Delta t\right)^2 + \left(y_i^t - y_j^t + \left|\vec{v}_i\right| \cos \alpha_i^t \Delta t - \left|\vec{v}_j\right| \cos \alpha_j^t \Delta t\right)^2} \leq R_j \\ \rightarrow &\sqrt{D_{i,j}^t + \left|\vec{v}_{i,j}^t\right|^2 + 2\Delta t \left[\left(x_i^t - x_j^t\right) \left(\left|\vec{v}_i\right| \sin \alpha_i^t - \left|\vec{v}_j\right| \sin \alpha_j^t\right) + \left(y_i^t - y_j^t\right) \left(\left|\vec{v}_i\right| \cos \alpha_i^t - \left|\vec{v}_j\right| \cos \alpha_j^t\right)\right]} \leq R_j \quad (15) \\ \rightarrow &L_{i,j}^t = D_{i,j}^t \cos \theta_{i,j}^t - \sqrt{R_j^2 - \left(D_{i,j}^t \sin \theta_{i,j}^t\right)^2} \leq \left|\vec{v}_{i,j}^t\right| \Delta t \end{aligned}$$

Therefore, whether there is a conflict between ship pairs during a time interval Δt or not can be determined by the Boolean variable $B(i, j, t)$.

$$B(i, j, t) \begin{cases} 1, & 0 < L_{i,j}^t \leq \left|\vec{v}_{i,j}^t\right| \Delta t \\ 0, & \text{otherwise} \end{cases} \quad (16)$$

Furthermore, due to the large span of the researched data in time and geographic scope, to avoid the extra calculation amount between trajectories that do not exist in conflict, it can be realized utilizing spatial retrieval. Currently, mature spatial retrieval technologies include the R-tree index, grid index, quad-tree index, and so on. In practical applications, we can combine one or more of the above technologies and merge their characteristics to form a new spatial indexing technology.

3. Case Study and Results

In this section, a series of experiments for the performance evaluation of the proposed MFDP and path generation algorithm by a real trajectory set obtained from Ningbo, CHINA was conducted. All the experiments are implemented in python 3.6 on an AMD Ryzen 7 5800H with Radeon Graphics 3.20 GHz machine with 16.0 GB of memory running Microsoft Windows 11. Our experimental results are as following discussions.

3.1. Experimental Data Source

In this study, AIS data were obtained from the Transportation and Maritime Safety Administration of China which is responsible for storing and maintaining various ship information along the coast of China.

The experimental data was all the Class-A AIS messages, that were collected in Zhou Shan Islands from selected obtained from 1 January 2018, to 10 January 2018. The data had been pre-processed utilizing the study which was described in the work [3]. The description of the measurement is shown in Table 1. The number of ships and the number of trajectories of each type are listed in Tables 2 and 3 gives the analysis of the ratio of ship length to width.

Table 1. Description of the measurement.

Measurement	Description	Unit
SED_{totv}	Total speed synchronization error of trajectories	Knots
SED_{avgv}	Average speed synchronization error of trajectories	Knots
SED_{totc}	Total course synchronization error of trajectories	Degree
SED_{avgc}	Average course synchronization error of trajectories	Degree
THR	Threshold (times the ratio of ship length to width)	None

Table 2. Quantity analysis for comparison between the existing method and proposed method.

THR	RDP		MCDP		Deleted Incorrectly
	Remained	Ratio (%)	Remained	Ratio (%)	
10	467,451	0.577976077	469,425	0.576193911	1974
20	343,302	0.690060227	347,703	0.686086918	4401
30	281,684	0.745690165	287,016	0.740876331	5332
40	242,889	0.780715051	248,949	0.775243964	6060
50	216,143	0.804861864	222,759	0.79888881	6616
60	196,856	0.822274546	203,506	0.816270795	6650
70	181,883	0.835792463	188,885	0.829470921	7002
80	170,002	0.846518863	177,063	0.840144054	7061
90	160,180	0.855386357	167,173	0.84907294	6993

Table 3. Speed loss of RDP and MCDP on the dataset with different thresholds.

THR	RDP		MCDP	
	SED _{totv}	SED _{avgv}	SED _{totv}	SED _{avgv}
10	93,320.9337	0.0506	90,102.9013	0.0496
20	146,309.7629	0.0824	139,549.0754	0.0803
30	188,422.1134	0.1085	178,265.4737	0.1052
40	224,651.3745	0.1312	210,409.2579	0.1266
50	256,911.2068	0.15082	239,676.8059	0.1453
60	285,895.6688	0.1681	265,157.2075	0.1616
70	311,667.0748	0.1840	288,211.9025	0.1763
80	335,843.8289	0.1992	309,908.0217	0.1907
90	359,218.8910	0.2129	329,731.6156	0.2036

3.2. Comparison of AIS Compression Method

3.2.1. Threshold Analysis

From the described in Section 3.1, we can find that Compression accuracy depends on the compression threshold. The larger the compression threshold, the lower the compression accuracy. So, the threshold selection will directly determine the quality of the simplified data. However, the optimal threshold value is difficult to determine which was affected by many factors. For example, Zhao et al., 2018 used the own ship’s length in different trajectories as the threshold value for satisfying the adaptability. Due to the data was collected from different types of ships, the characteristics of the trajectory are affected by a number of parameters, such as the length, the width, and the ratio of length to width. For example, the same length of ship sail on the same water, the track is different depending on the width of ships.

Considering the different maneuverability of ships, the bigger the ship, the more difficult it is to operate, therefore the track of the big one would have fewer points deviate the overall shape of ship trajectory than the small one. Therefore, if the threshold is global and suitable for the big ship, then the important track point for the small ship would be discarded. Conversely, if the threshold is suitable for the small ship, then there would retain many points of redundancy, and this leads to the low compression rate.

Consequently, the ratio of the length to width of ships was considered in the selection of threshold values. The threshold for each trajectory is different from each ratio of ship length to width. In the following experiments, which equal to 10 times or 100 times the ratio of ship length to width. Furthermore, when setting the threshold value, the sailing waters also can be considered, for example, the ratio in the harbor water can be set smaller than the ratio in the open water. Our practical experience shows that the appropriate ratio in the harbor water usually ranges from 30 times to 80 times. For different users, the threshold value varies to the purpose, thus whose variability should be considered in the application of the DP algorithm. Under certain conditions, the threshold value can be fixed differently as needed.

From the results shown in the left one of Figure 7, there was a strong correlation between the retention rate of characteristic points and the course threshold when the threshold of the Euclidean Distance and angle difference remained constant. From the results shown in the right one of Figure 7, there was a strong correlation between the retention rate of characteristic points and speed threshold when the threshold of the Euclidean Distance and the course difference remained constant. This showed that the MCDP method can retain the characteristic points that we wanted, and more features can be extracted.

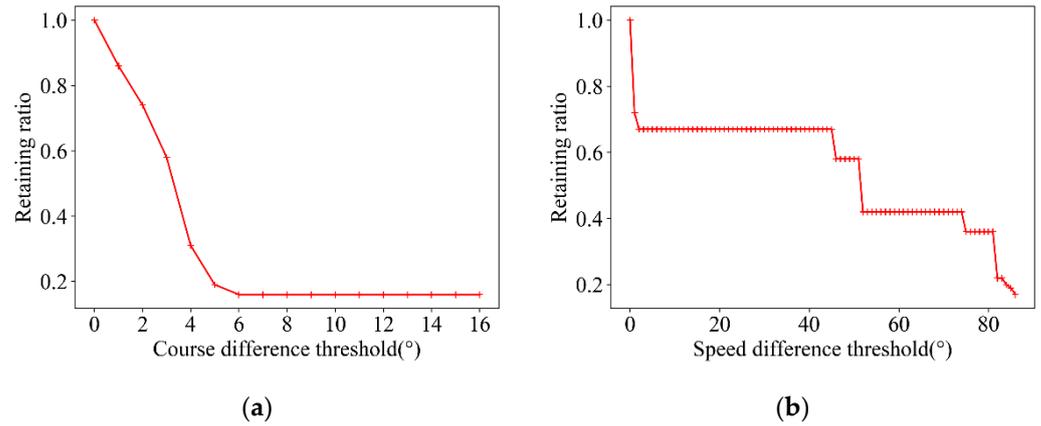


Figure 7. The relationship between retaining ratio and speed constraint or course constraint. (a) is the relationship between the retaining ratio and course constraint when the speed constraint is 87 and the threshold of the Euclidean distance is 100 times; (b) is the relationship between the retaining ratio and speed threshold when the course threshold is 27 and the threshold of the Euclidean distance is 100 times.

Figures 8a and 9a showed a strong correlation between the course difference, course threshold, and speed threshold when the threshold of the Euclidean Distance remained constant in standard error and mean error, respectively. The results in Figures 8b and 9b showed a strong correlation between the speed difference, speed threshold, and course difference when the threshold of the Euclidean Distance remained constant in standard error and mean error, respectively. Through analysis, it is also can be found that the MCDP method can retain the characteristic points that we wanted, and more features can be extracted.

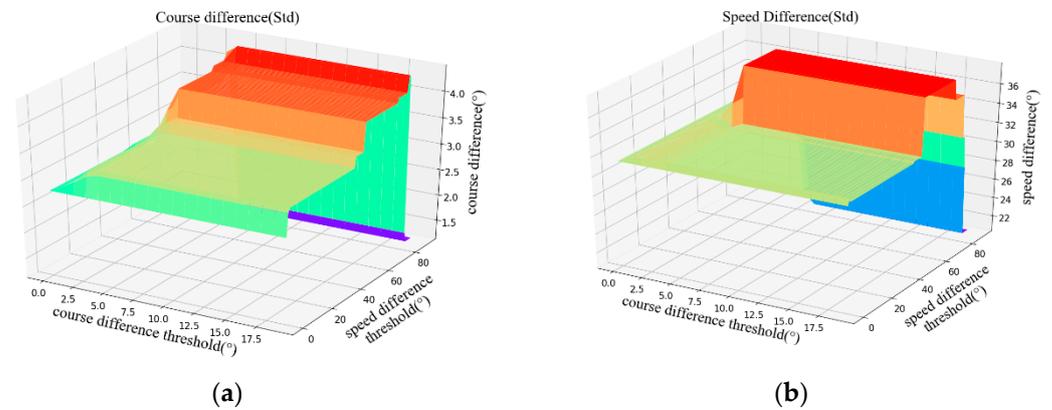


Figure 8. Correlation between the standard error of the course difference threshold, the speed difference threshold, and the course difference or speed mean difference. (a) shows the course difference; (b) shows the speed difference.

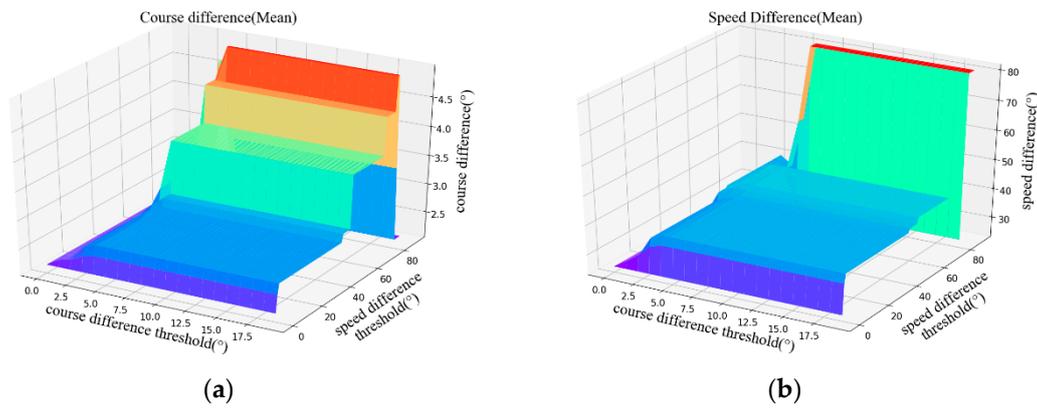


Figure 9. Correlation between the course difference threshold, speed difference threshold, and the mean error of course difference or speed difference. (a) shows the course difference; (b) shows the speed difference.

3.2.2. Experimental Comparison

To further verify quality improvement, the samples were tested with the provided method. The experimental results of the proposed method are shown in Tables 2–4. The results of the number of critical points deleted by RDP incorrectly are presented in Table 2, where RDP was tested with the threshold of Euclidean Distance, and MCDP was tested with the threshold of Euclidean Distance, angle constraint, and course constraint. From this, we can find that the provided method could retain more characteristic points than RDP. Tables 3 and 4 is the speed loss and course loss of RDP and MCDP on the dataset with different thresholds. In comparing the results, MCDP (rather than RDP) improved 1.5–2.5% in the average course synchronization error of trajectories and 2–4.5% in the average speed synchronization error of trajectories. From the comparison shown in Figure 10, it is apparent that the trajectory after simplification using MCDP was more similar to the original trajectory than when using RDP for the course difference and speed difference. Table 1 is the measurement, unit, and description of the measurement.

Table 4. Course loss of RDP and MCDP on the dataset with different thresholds.

THR	RDP		MCDP	
	SED _{totc}	SED _{avgc}	SED _{totc}	SED _{avgc}
10	1,711,254.8141	1.0810	1,663,394.0128	1.0656
20	2,713,959.2632	1.7590	2,620,121.4988	1.7256
30	3,503,489.6310	2.3169	3,373,388.1781	2.2699
40	4,174,016.6652	2.8159	3,990,199.7671	2.7510
50	4,740,253.5191	3.2101	4,524,679.0714	3.1326
60	5,245,404.4085	3.5704	5,001,570.1720	3.4821
70	5,716,247.3330	3.8996	5,441,852.0286	3.8040
80	6,134,831.4963	4.1966	5,837,147.1061	4.0917
90	6,540,768.7990	4.4750	6,211,981.4299	4.3621

Figure 11 shows the Synchronized Euclidean Distance of MCDP and RDP. In Figure 10, the various Synchronous Euclidean Distance error of speed and course of RDP is greater than MCDP, in other words, the trends both in speed and course simplified with MCDP algorithm were more accurate than that of RDP algorithm by SED_{totv} , SED_{avgv} , SED_{totc} , and SED_{avgc} . The reason is that the RDP algorithm just considered the spatial and temporal information and not the speed and course status. Our MCDP method fully and deeply tackled multiple constraints, e.g., velocity, course, and position, for improving the accuracy of the trajectory simplification. The Synchronous Euclidean Distance error of both methods increases with an increasing threshold value.

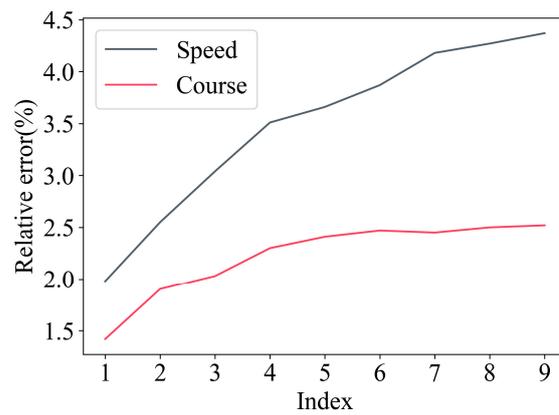


Figure 10. The relative error between Synchronous Euclidean Distance error of speed and course of RDP and MCDP.

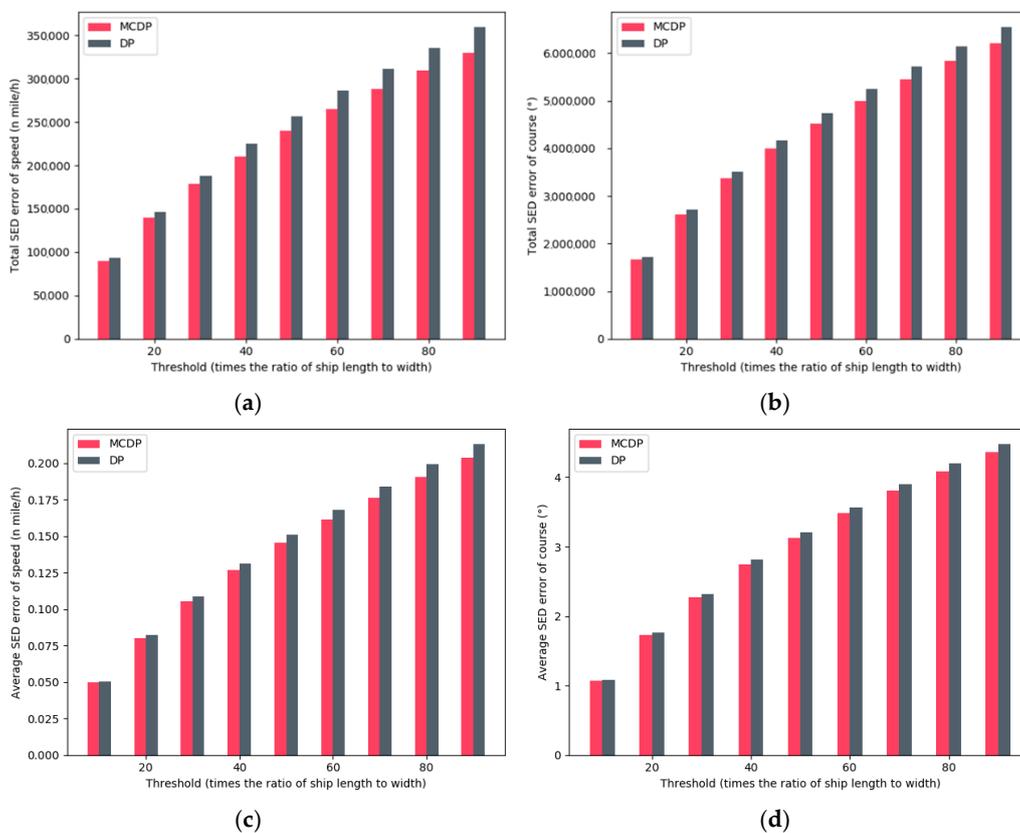


Figure 11. Comparison between Synchronous Euclidean Distance error of speed and course of RDP and MCDP. (a–d) show the total SED error of speed, total SED error of course, average SED error of speed, and average SED error of course, respectively.

Figure 12 shows the comparison in terms of the number of the remained critical points in different times of ship length to width. From this, we can find that the number of the remained critical points of both methods decreases with increasing threshold and MCDP can remain much more critical points than RDP.

To verify the advantages of MCDP, one trajectory was utilized to conduct the parallel experiments. In this experiment, the threshold value was set as 50 times the ship length to width. In Figure 13, the blue and red lines represent the original algorithm and the improved algorithm, respectively. The improved algorithm at each peak and valley can better preserve the original properties. The results reveal that the MCDP method can

preserve more attributes than RDP methods in terms of speed attribute and course attribute. Figure 14 shows the compression results of a vessel trajectory in the parallel experiment. Through analysis of this figure, it can be found that there will have more critical points been remained as denoted with an arrow in Figure 14. Therefore, the MCDP method is superior to the RDP method in retaining the attribute characteristics of trajectory including shape, speed, and course.

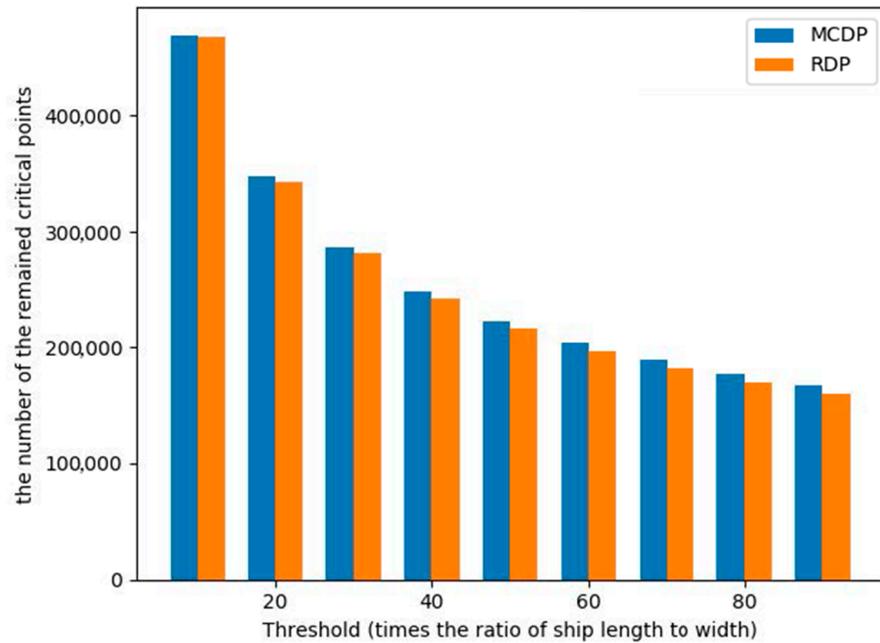


Figure 12. Comparison between MCDP and RDP.

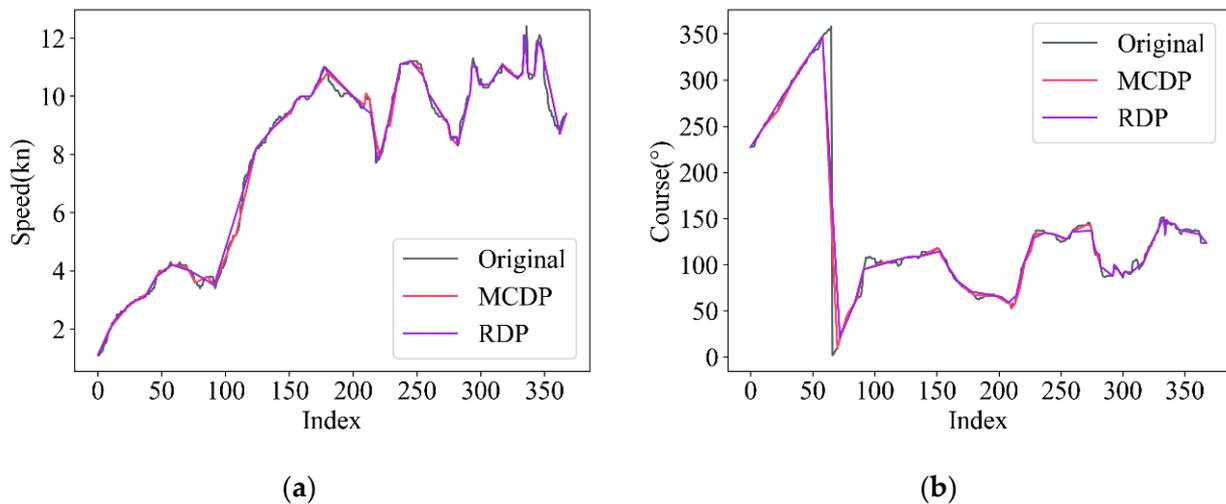


Figure 13. Charts comparing with the RDP algorithm and the MCDP algorithm (Our algorithm). (a) is the velocity simplification comparison with RDP, original, and MCDP; (b) is the course simplification comparison with RDP, original, and MCDP. The threshold was set as 50 times the ship length to width.

3.3. Collision Risk Analysis

In this section, collision risk was assessed with different meeting situations, such as head-on, crossing, overtaking situation. Therefore, we derived from existing studies [2,3,7], the causation probability values for head-on, crossing, and overtaking conflicts were set as $p_h = 4.90 \times 10^{-5}$, $p_c = 6.83 \times 10^{-5}$, and $p_o = 4.90 \times 10^{-5}$, respectively.

Table 5. Comparison of collision frequency in different duty periods.

Collision Type	Head On	Crossing	Overtaking	Overall
First officer	6.52×10^{-3}	5.18×10^{-2}	5.14×10^{-2}	1.097×10^{-1}
Second officer	2.21×10^{-3}	3.42×10^{-2}	2.75×10^{-2}	6.396×10^{-2}
Third officer	4.90×10^{-3}	3.61×10^{-2}	4.59×10^{-2}	8.689×10^{-2}

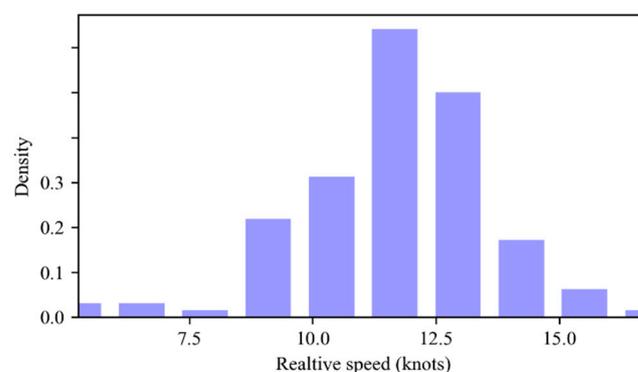
Table 6 shows the comparison of collision frequency in different encounter situations using analysis of variance (ANOVA) [3,8]. The *F* value and *p*-value in ANOVA are used to analyze the differences between and within groups. *F* value corresponds to a *p*-value, and the larger the *F* value, the smaller the *p*-value. If the *F* value is equal to 1, the difference between groups is consistent with the difference within groups. It can be found that *F* values of the three conflict types differ greatly and *p* values are small, therefore, the differences between groups are large, and it can be concluded that the collision frequency is affected by the driver’s experience or duty period.

Table 6. Comparison of collision frequency in different encounter situations.

Collision Type	Mean/Day	Std/Day	<i>F</i>	<i>p</i> -Value
Head-on				
First officer	7.24×10^{-4}	3.109×10^{-3}	5.92	0.008
Second officer	2.45×10^{-4}	6.49×10^{-4}		
Third officer	5.44×10^{-4}	2.45×10^{-4}		
Crossing				
First officer	5.75×10^{-3}	9.64×10^{-2}	2.68	0.089
Second officer	3.80×10^{-3}	2.99×10^{-2}		
Third officer	4.01×10^{-3}	4.25×10^{-2}		
Overtaking				
First officer	5.72×10^{-3}	1.36×10^{-1}	3.69	0.040
Second officer	3.06×10^{-3}	2.55×10^{-2}		
Third officer	5.10×10^{-3}	1.27×10^{-1}		

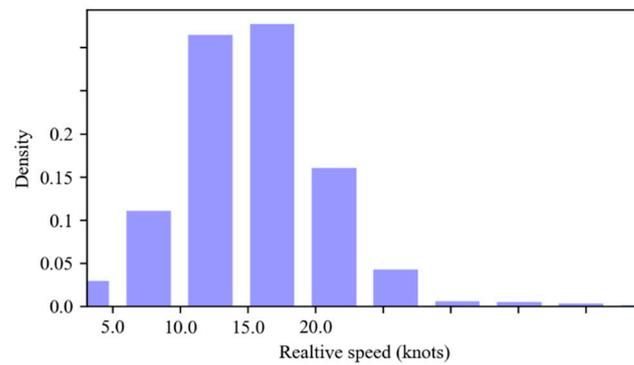
Figure 15 reveals that the difference between the relative speed distribution of ships involved in different conflict types. The major relative speed in head-on, crossing, and overtaking conflicts is range 11–12, 15–18, and 6–8 knots, respectively.

Figure 16 indicates that the length distribution for vessels involved in different types of conflict. From three types of distribution, it can be found that the majority of length in head-on and crossing conflicts is around 100 m. In overtaking conflict situation, the majority of length ranges from 100 to 150 m.

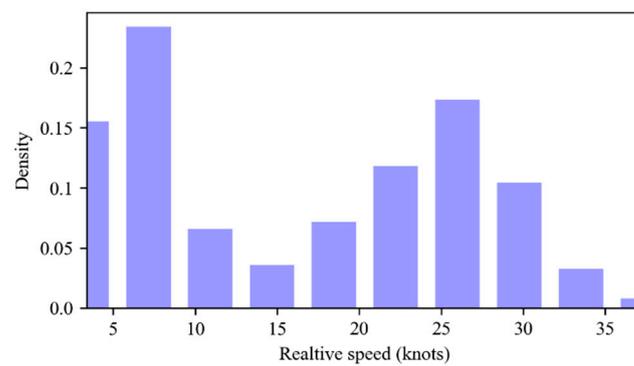


(a)

Figure 15. Cont.



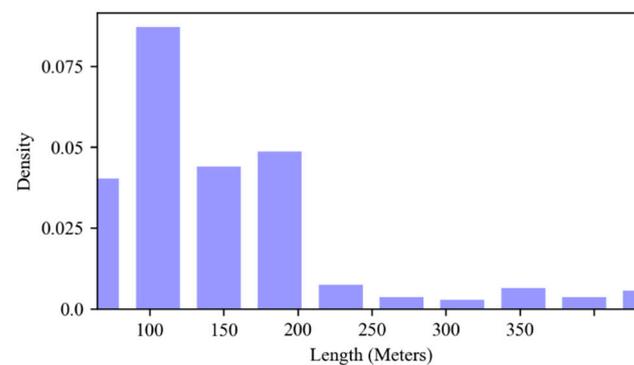
(b)



(c)

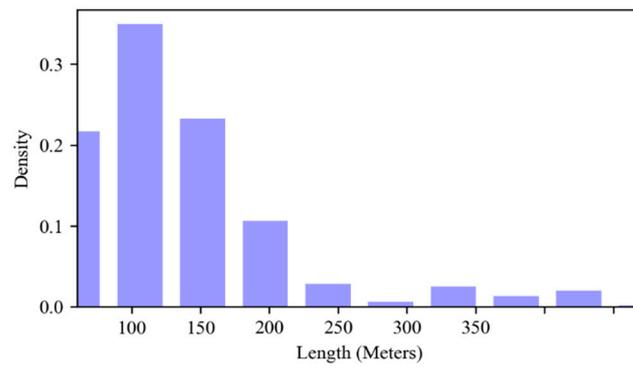
Figure 15. Relative speed distribution for vessels involving in different conflicts. (a) Head-on conflict; (b) Crossing conflict; (c) Overtaking conflict.

Figures 17–19 illustrate the spatio-temporal distribution under head-on, crossing, and overtaking conflicts for the specific time of the studied waters respectively. Where the red, yellow, and brown colors represent the conflicts at different watch stages, respectively. The red color represents the duty stage of the first officer from 00:00 to 04:00 and 12:00 to 16:00. The yellow color represents the duty stage of the second officer from 04:00 to 08:00 and 16:00 to 20:00. The brown color represents the duty stage of the third officer from 08:00 to 12:00 and 20:00 to 24:00.

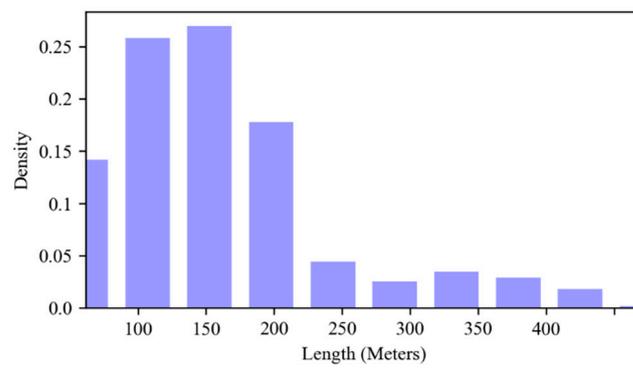


(a)

Figure 16. Cont.



(b)



(c)

Figure 16. Length distribution for vessels involved in different conflicts. (a) Head-on conflict; (b) Crossing conflict; (c) Overtaking conflict.

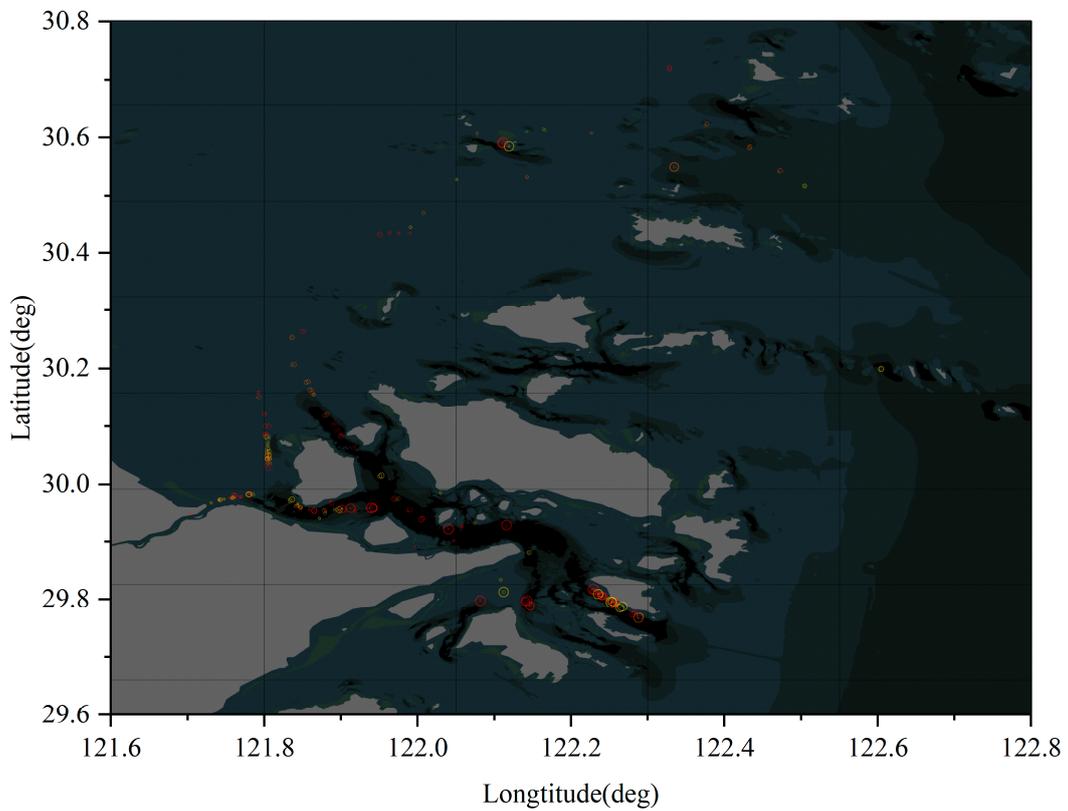


Figure 17. Spatio-temporal distribution of head-on conflicts.

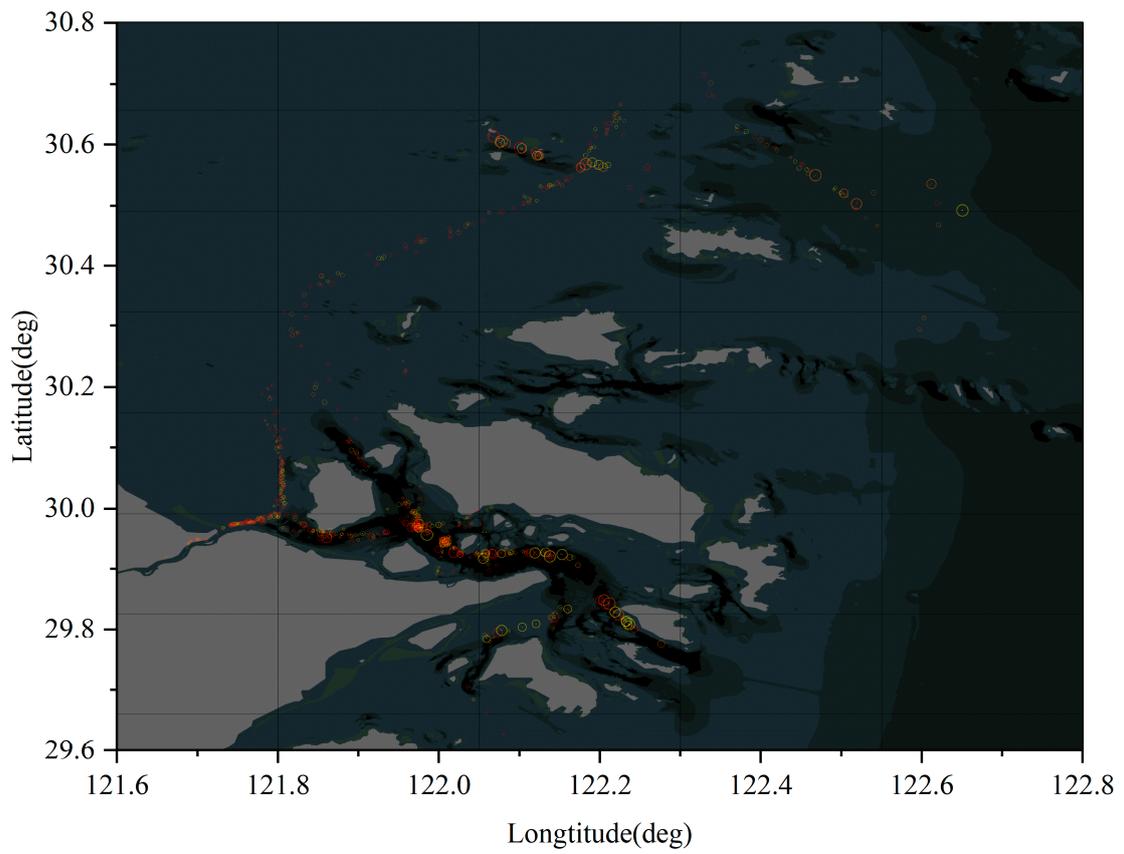


Figure 18. Spatio-temporal distribution of crossing conflicts.

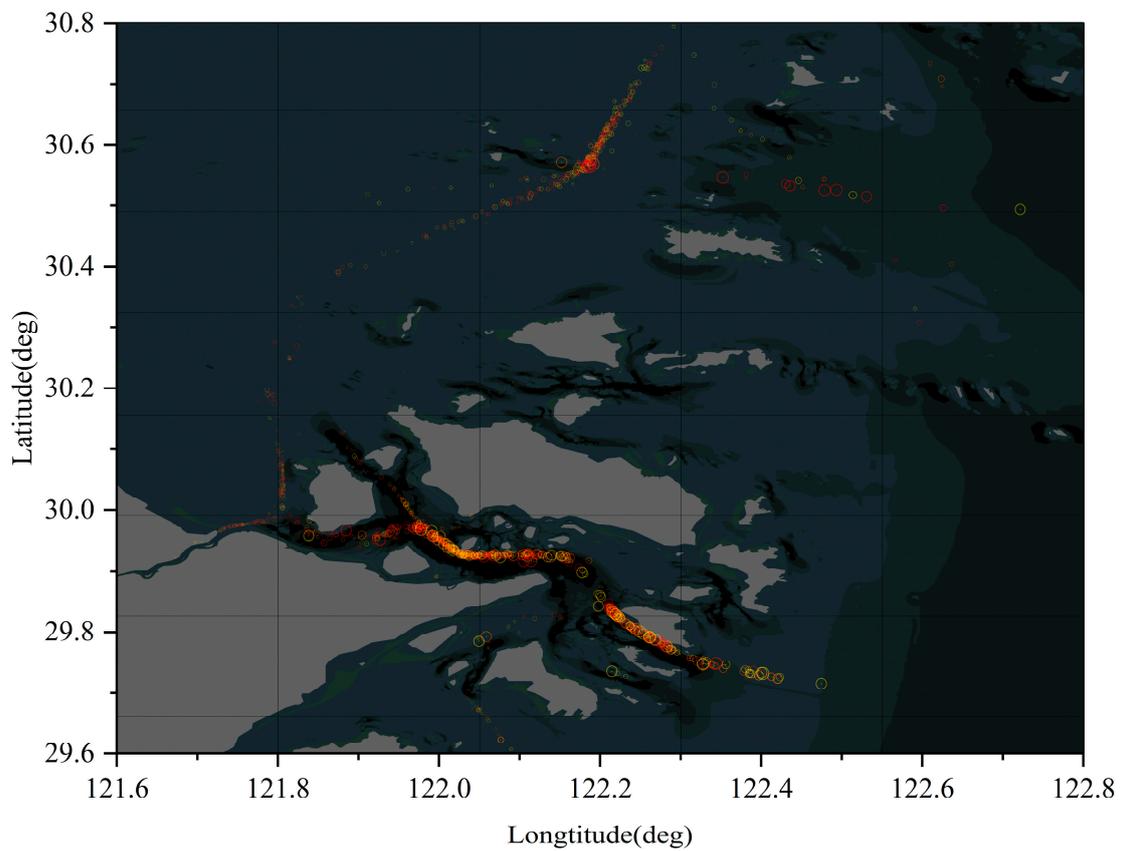


Figure 19. Spatio-temporal distribution of overtaking conflicts.

4. Discussion

Combining the water layout map in Figure 1 and the spatio-temporal trajectory distribution map, it can be found that the intersection of the Jin-Tang Bridge and the north side of the Xi Hou-men Bridge are both high incidence areas in the three types of conflicts. Furthermore, there are few conflicts in the Yang-shan harbor of Shang-hai port, the reason caused the conflict here should be limited by the narrow waters here so that the encounter distance between the ship pairs is less than the safety domain applied in this study. The width of the channel here is about 400 m, therefore, if there exists an encounter situation in the channel, it is very easy to form a conflict. In addition, during the maintenance and dredging of this channel, some engineering vessels are continuously operating in the channel, and other vessels may easily form conflicts with such vessels.

From the three spatio-temporal results, it can be found that conflicts are easy to occur in the traffic separation system (TSS) of Ningbo Zhoushan Port, especially at the intersection and the waterway in the south. The reasons for this phenomenon may be the following two aspects, on the one hand, the ship traffic flow is large, on the other hand, it may be due to there is a difference of the size of safety domain between the applied in this study and the actual in TSS.

Due to the simple pre-processing of the ship trajectory data in the early stage, the calculated collision frequency values will be smaller than the actual situation. The collision correlation between different ship types and ship sizes can be further analyzed in the later stage to further the risk evaluation of the waters.

5. Conclusions

The purpose of this study is to provide a framework for the design of a ship collision risk assessment system in port waters. In this paper, based on the estimated data of ships in port waters, we proposed a novel track pairs collision detection algorithm for evaluating the risk of ship collision in the port water. Based on considering the influence of ship speed on trajectory compression, the DP algorithm is improved, which not only ensures the compression ratio but also retains more key behavior points. The framework has great flexibility and can realize collision risk assessment in different waters, which is helpful to identify high-risk areas in port waters and improve ship management.

In the future, we can study the collision probability of this water by collecting collision accident data and determine a more suitable ship safety domain through comparative experiments.

Author Contributions: Conceptualization, J.S.; software, J.S.; supervision, Z.L.; writing—original draft, J.S.; writing—review and editing, J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number (51579025, 51709165) and the Fundamental Research Funds for the Central Universities (3132020134, 3132020139).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Some or all data, models, or code generated or used during the study are available from the corresponding author by request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. UNCTAD. *Review of Maritime Transport, 2021*. No. UNCTAD/RMT/2021; United Nations Publications: New York, NY, USA, 2021.
2. Xu, Q.; Wang, N. A survey on ship collision risk evaluation. *Promet-Traffic Transp.* **2014**, *26*, 475–486. [[CrossRef](#)]
3. Liu, D.; Shi, G. Ship collision risk assessment based on collision detection algorithm. *IEEE Access* **2020**, *8*, 161969–161980. [[CrossRef](#)]
4. Fan, H.S.L.; Cao, J.M. Sea space capacity and operation strategy analysis system. *Transp. Plan. Technol.* **2000**, *24*, 49–63. [[CrossRef](#)]

5. Li, Q.; Lam, J.S.L. Conflict resolution for enhancing shipping safety and improving navigational traffic within a seaport: Vessel arrival scheduling. *Transp. A Transp. Sci.* **2017**, *13*, 727–741. [[CrossRef](#)]
6. Ozturk, U.; Cicek, K. Individual collision risk assessment in ship navigation: A systematic literature review. *Ocean. Eng.* **2019**, *180*, 130–143. [[CrossRef](#)]
7. Chai, T.; Weng, J.; Xiong, D.-Q. Development of a quantitative risk assessment model for ship collisions in fairways. *Saf. Sci.* **2017**, *91*, 71–83. [[CrossRef](#)]
8. Weng, J.; Meng, Q.; Qu, X. Vessel collision frequency estimation in the Singapore Strait. *J. Navig.* **2012**, *65*, 207–221. [[CrossRef](#)]
9. Goerlandt, F.; Montewka, J.; Kuzmin, V.; Kujala, P. A risk-informed ship collision alert system: Framework and application. *Saf. Sci.* **2015**, *77*, 182–204. [[CrossRef](#)]
10. Debnath, A.K.; Chin, H.C.; Haque, M.M. Modelling port water collision risk using traffic conflicts. *J. Navig.* **2011**, *64*, 645–655. [[CrossRef](#)]
11. Yoo, Y.; Lee, J.-S. Collision Risk Assessment Support System for MASS RO and VTSO Support in Multi-Ship Environment of Vessel Traffic Service Area. *J. Mar. Sci. Eng.* **2021**, *9*, 1143. [[CrossRef](#)]
12. Douglas, D.H.; Peucker, T.K. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartogr. Int. J. Geogr. Inf. Geovisualization* **1973**, *10*, 112–122. [[CrossRef](#)]
13. Lou, Y.; Zhang, C.; Zheng, Y.; Xie, X.; Wang, W.; Huang, Y. Map-matching for low-sampling-rate GPS trajectories. In Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Seattle, WA, USA, 4–6 November 2009; pp. 352–361.
14. Xiang, L.; Wu, T.; Gong, J. A Geo-spatial Information Oriented Trajectory Model and Spatio-temporal Pattern Querying. *Acta Geod. Cartogr. Sin.* **2014**, *43*, 982–988.
15. Ying, J.J.; Su, J. On velocity-preserving trajectory simplification. In Proceedings of the Asian Conference on Intelligent Information and Database Systems, Da Nang, Vietnam, 14–16 March 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 241–250.
16. Ericson, C. *Real-Time Collision Detection*; CRC Press: Boca Raton, FL, USA, 2004.
17. Tang, M.; Manocha, D.; Tong, R. Fast continuous collision detection using deforming non-penetration filters. In Proceedings of the 2010 ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games, Washington, DC, USA, 19–21 February 2010.
18. Shi, J.; Liu, Z. Deep Learning in Unmanned Surface Vehicles Collision-Avoidance Pattern Based on AIS Big Data with Double GRU-RNN. *J. Mar. Sci. Eng.* **2020**, *8*, 682. [[CrossRef](#)]