Article

Ship Intention Prediction at Intersections Based on Vision and Bayesian Framework

Qianqian Chen 1, Changshi Xiao 1,2,3,*, Yuanqiao Wen 2,4, Mengwei Tao 1 and Wenqiang Zhan 3

1 School of Navigation, Wuhan University of Technology, Wuhan 430063, China; chenqq@whut.edu.cn (Q.C.); tao_mw@163.com (M.T.)
2 National Engineering Research Center for Water Transport Safety, Wuhan 430063, China; yqwene@whut.edu.cn
3 Institute of Ocean Information Technology, Shandong Jiaotong University, Weihai 264200, China; 210127@sdjtu.edu.cn
4 Intelligent Transportation Systems Research Center, Wuhan University of Technology, Wuhan 430063, China
* Correspondence: cs_xiao@hotmail.com

Abstract: Due to the high error frequency of the existing methods in identifying a ship’s navigational intention, accidents frequently occur at intersections. Therefore, it is urgent to improve the ability to perceive ship intention at intersections. In this paper, we propose an algorithm based on the fusion of image sequence and radar information to identify the navigation intention of ships at intersections. Some existing algorithms generally use the Automatic Identification System (AIS) to identify ship intentions but ignore the problems of AIS delay and data loss, resulting in unsatisfactory effectiveness and accuracy of intention recognition. Firstly, to obtain the relationship between radar and image, a cooperative target composed of a group of concentric circles and a central positioning radar angle reflector is designed. Secondly, the corresponding relationship of radar and image characteristic matrix is obtained after employing the RANSAC method to fit radar and image detection information; then, the homographic matrix is solved to realize radar and image data matching. Thirdly, the YOLOv5 detector is used to track the ship motion in the image sequence. The visual measurement model based on continuous object tracking is established to extract the ship motion parameters. Finally, the motion intention of the ship is predicted by integrating the extracted ship motion features with the position information of the shallow layer using a Bayesian framework. Many experiments on real data sets show that our proposed method is superior to the most advanced method for ship intention identification at intersections.

Keywords: ship intention identification; AIS; RANSAC; Bayesian framework; YOLO; intersection

1. Introduction

Due to the combination of information technology and ship technology, ship safety has experienced rapid developments. Although a variety of water situation awareness and intelligent collision avoidance technologies have been developed, ship collision accidents still occur frequently in ports and other complex waters, resulting in severe economic losses and environmental pollution. According to the Shanghai Maritime Safety Administration report, 47% of ship collisions occurred at waterway intersections (Shanghai Maritime Court, Shanghai, China, 2018). These accidents occur mainly because the seafarers are unable to correctly identify or predict the motion intentions of other ships in the intersections. Therefore, it is of significance to study the vessel intention perception methods at intersections and channels.

Although intention prediction has received increasing research in the field of road traffic in recent years, there has been little research on the intention prediction of maritime safety of ships. Moreover, the existing research on the prediction of ground vehicle driving intention cannot be directly applied to ships. For example, ships can travel over relatively wide areas without being restricted by roads. In addition, there is no fixed channel to
separate ships with different maneuvers, which increases the uncertainty of the ship’s trajectory, thus increasing the difficulty of ship navigation prediction. Beyond that, the movement of ships is susceptible to environmental conditions (such as wind, waves, and ocean currents) and surrounding ships in waterways. As a result, ships may exhibit different motion patterns even if they travel along the same route, further complicating the prediction of ship intentions. Most importantly, ships cannot perform maneuvering operations such as sudden stops, turns, or reversals unlike ground vehicle movements. Furthermore, it takes more time and space for a ship to transition from one state of motion to another. Therefore, the intention prediction method of ground vehicles cannot correctly describe the long-distance movement of ships.

In order to solve the problem of rapid identification of ship intentions, we propose a scheme to update and predict ship intentions based on channel video surveillance and radar data under the Bayesian framework. Firstly, a cooperative target composed of a group of concentric circles and a centrally positioned radar angle reflector was designed to obtain the relationship between radar and image data in this paper. Secondly, the RANSAC method was used to fit the radar and the image detection information, and then the homographic matrix from the radar coordinate system to the image coordinate system was obtained. Thirdly, a visual measurement model based on continuous object tracking was established to extract the motion parameters of the ship. Finally, the motion intention of the ship was predicted by integrating the extracted ship motion features with the position information of the shallow layer using the Bayesian framework. The main contributions of this work are threefold:

1. A ship motion model based on monocular vision is established, which can extract ship motion parameters real-time and is not affected by AIS delay;
2. The accuracy of ship intention prediction can be further improved by using the environmental information of the channel effectively;
3. A dynamic Bayesian model is established that can accurately identify the ship’s intention.

The remainder of this paper is organized as follows. Some related works are introduced in Section 2. In Section 3, we will discuss the methodology of our algorithm. Experimental and model prediction performances are reported in Section 4. Finally, the work is concluded in Section 5.

2. Related Work

There are several relevant studies in the literature on the navigational intention prediction of ships. For example, Tang Huang et al. [1] from Dalian Maritime University found that there were field errors and obvious noise in the original AIS track data set of ships and that these data were often irregular timing data. Therefore, they proposed an improved trajectory similarity measure with a directional resolution to improve the accuracy of trajectory clustering. On this basis, they used the Long Short-Term Memory (LSTM) neural network to accurately predict ship behavior. Steidel et al. [2] believe that traditional maritime abnormal behavior detection and the prediction of a ship’s navigational intention mainly focus on the development of methods to extract typical ship motion patterns from historical traffic data without considering contextual information. Therefore, he proposed a method to predict a ship’s intention by combining historical ship traffic data with information about shipping routes. Later, Pietrzykowski used AIS online data to analyze if the ship’s operation and identification behavior were potentially risky [3]. Based on this, Zhang Hong used a data mining method to analyze the AIS data of tuna purse-seine fishing ships to identify the operation state of tuna purse-seine fishing ships [4]. In addition, Gao et al., for example, constructed a Bi-LSTM-RNN model, which can be used for AIS date and time series feature extraction and online parameter adjustment to realize online real-time prediction of ship behavior. This algorithm enhanced the correlation between historical data and future data, thus improving the prediction accuracy [5]. Later, Kawamura used a GPU SPH simulator to predict the motion of a 6-dOF ship in harsh water transport conditions [6]. The study of Murray et al. deconstructed ship behaviors
into clusters according to specific regions, and each cluster contained similar trajectory behavior characteristics. A deep learning framework was used to predict ship behavior for each specific cluster. This algorithm believes that the future navigation trajectory of a ship can be predicted by inputting the past trajectory of a ship in a specific area. However, the influence of environmental factors was not considered in this method, which is somewhat different from the actual situation of ship navigation [7].

Ma et al. conducted two valuable studies [8,9]. First, they proposed to extract the mutual behavior between ships from AIS tracking data to capture the spatial dependence between ships that meet and then predicted the ship’s intention and collision risk based on a LSTM network. Later, they devised a data-driven approach that linked movement behavior to future and early risks, and predicted a ship’s collision risk by classifying behavior into appropriate risk levels. In reference [10], it was assumed that all ships in the scene could not share the same motive and decision of motion, and the observation-inference-prediction-decision (OIPD) model was proposed to avoid collisions by repeatedly iterating the difference between observation and prediction information. Tang [11] used a grid-based method to discretize the historical AIS data into track segments and established a probabilistic directed graph model. Through this model, the state characteristics of each node ship can be counted, and the navigation state of the ship can be detected by the probability graph obtained.

Some recent studies include, for example, [12] a proposed a new spatial-temporal geographic method to solve the risk behavior of multi-ship collision based on ship movement. The direction-constrained space-time prism was used to characterize the possibility of the ship’s interaction, which enabled the assessment of the ship’s potential collision risk. Suo proposed a modeling method based on a cellular automata simulation to analyze and evaluate maritime traffic risks in a port environment in real-time [13]. Then, Alvarellos et al. established a deep neural network to predict the damage to ships anchored in ports caused by environmental effects within 72 h by considering the influence of ship size, sea state, and weather conditions on ship motion [14]. Based on the above work, Xue proposed a knowledge learning model under multi-environment constraints to analyze ship risks in port waters and improve the decision-making basis for autonomous navigation of intelligent ships [15].

Some other works are also of great research value in predicting ship intentions. Alizadeh proposed a point-based and track-based model, considering the constant distance between the target and the sample trajectory. The LSTM method was then used to measure the dynamic distance between the target and the sample trajectory, and predict the short-term and long-term trajectory of the ship [16]. Subsequently, Praczyk et al. extracted spatial direction (Euler Angle) from the inertial navigation system and used an improved neural network to predict ship behavior [17]. Zissis used an artificial neural network to predict the future position, speed, and heading behaviors of ships on a large scale based on historical AIS data [18].

Mining-related behavior characteristics from AIS data is a general method to study ship behavior. However, due to the large ship flow in inland river confluence waters, AIS data bandwidth is insufficient, and uploading is not timely, which leads to the low real-time prediction of ship navigation intention. It is easy to miss the best decision time and increase the risk of ship collision. In this paper, the motion intention of ships is predicted by combining visual and radar data based on the Bayesian framework, and the problem of accuracy and real-time intention prediction is successfully solved. Firstly, to obtain the relationship between radar and image, a cooperative target composed of a group of concentric circles and a central positioning radar angle reflector is designed. Secondly, the corresponding relationship of radar and image characteristic matrix was obtained after employing the RANSAC method to fit the radar and the image detection information. Then, the homography matrix was solved to realize the radar and the image data matching. Thirdly, the extended YOLO detector was used to track the ship motion in the image sequence, and the visual measurement model based on continuous object tracking was
established to extract the ship motion parameters. Finally, the motion intention of the ship was predicted by integrating the ship motion features extracted with the position information of the shallow layer using the Bayesian framework. At the same time, in order to verify the feasibility of the proposed method, experimental scenarios were designed according to the scene characteristics of the intersection area of Wuhan Yangtze River and Han River, and the ship intention recognition algorithm was verified successfully.

3. Methodology

The proposed method takes a dynamic Bayesian algorithm as the main framework and introduces a new image measurement method to extract the motion characteristics of the ship. To better introduce the idea of the proposed algorithm, this section is divided into three sub-sections. In Section 3.1, we will describe how to process sequential images to measure ship speed and angular velocity, in which the measurement model will also be introduced in detail. In Section 3.2, we will discuss how to map an object in the image sequence to an electronic chart. In Section 3.3, how to use a dynamic Bayesian algorithm to estimate the navigation intention will be introduced in detail. The overall description of our algorithm is shown in Figure 1.

Figure 1. Framework based on vision and Bayesian framework for intent prediction of vessels.

3.1. Radar and Vision Fusion Calibration

Monocular vision and millimeter-wave radar signal have their own characteristics; for example, monocular vision has the advantages of simple structure and strong robustness, while millimeter-wave radar has the advantages of accurate positioning, etc. [19]. Compared with the measurement results of a single sensor, radar and vision fusion measurements can obtain more ship attitude and motion information [20]. However, this method usually requires calibrating radar and visual measurement results first. Therefore, it is necessary to design a cooperative objective for calibration that meets the following requirements:

1. Visual signals are sensitive and easy to accurately detect the position of fixed points;
2. The radar echo signals are strong and easy to accurately detect the position of the fixed point;

Therefore, we creatively designed a radar and vision cooperative target, as shown in Figure 2. The cooperative target consists of a set of concentric circles and a centrally located radar reflector. In order to obtain the sub-pixel coordinates of the center projection point, a recursive algorithm based on harmonic relation was proposed. Then, the radar points and image points were fitted to obtain the linear correspondence to accurately obtain the homographic matrix of radar coordinates to image coordinates.
Millimeter-wave radar is the main component of the surface target information acquisition system. The millimeter-wave radar used in this paper is the ARS300 series radar provided by Continental, Germany, which operated at 77 GHz and can detect up to 40 targets simultaneously, and is equipped with a special controller. This millimeter-wave radar has the characteristics of small size, strong anti-interference ability, and stable detection. Its performance indicators are shown in Table 1.

<table>
<thead>
<tr>
<th>Indicators</th>
<th>Detection Distance</th>
<th>Working Frequency</th>
<th>Range Accuracy</th>
<th>Speed Range</th>
<th>Detection Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance</td>
<td>2 km</td>
<td>77 GHz</td>
<td>0.5 m</td>
<td>265 km/h</td>
<td>near 60° far 20°</td>
</tr>
</tbody>
</table>

The millimeter wave radar detection is divided into short-range wave and long-range wave ranges. The long-range wave mainly captures distant targets and improves detection distance; The short-range wave mainly expands the radar perspective and reduces the dead detection zone. The millimeter-wave radar detection range is shown in Figure 3.

Assuming that the homogeneous coordinate of a point on the radar detection plane is $p_r = (x_r, y_r, 1)^T$, and the homogeneous coordinate of a point on the image plane is $p_i = (u, v, 1)^T$, Equation (1) indicates the conversion relationship from the radar coordinate system to the image coordinate system.

$$sp_i = Hp_r$$

where, $s$ is a scalar and $H$ is a $3 \times 3$ reversible homography matrix with 8 degrees of freedom. The purpose of calibration is to estimate the homography matrix $H$. After moving linearly in the common field of vision of the camera and radar during calibration, a series of radar and image points can be captured, and then the image sequence was processed to extract the coordinates of the concentric circles in each image. It should be emphasized here that the concentric circle center coordinates corresponded to the center of the image coordinates of the radar reflector. Finally, the least-squares method was used to fit the
image point sequence and the radar point sequence. It was assumed that the connection line of the continuous image center point is \( l \) and that of the radar detection point is \( L_r \), which has the following relationship:

\[
l^T p_i = 0, \quad L^T r = 0
\]  

(2)

Combined with Equations (1) and (2), the relationship between the straight lines in the image plane and the radar plane is shown in Equation (3):

\[
sL = HTl
\]  

(3)

Solving Equation (3) requires at least four sets of corresponding lines. We found that there was no need to align the time stamp of the radar data and the single frame of the video series, and the linear motion of the calibrated target can be easily captured by the radar due to the motion prediction algorithm adopted when the radar tracks the target. Therefore, the homographic matrix \( H \) can be calculated by using the linear-based homographic estimation method.

Figure 4 is the result of calibration using the method proposed in this section, where the blue dot represents the radar detected target and the red dot represents the pixel coordinates, which are consistent with the position of the ship detected in the image to achieve the radar and visual fusion calibration.

3.2. Object Detection

The purpose of image preprocessing is to capture every frame of the video and detect the moving ships after removing the random noise from the images. At present, the common object detection methods include the frame difference method [21], the motion modeling method [22], and the deep learning method [23]. These methods are simple and fast. However, in contrast, the frame difference and motion modeling methods are susceptible to environmental changes such as light changes and noise.

Through experiments, we found that the You Only Look Once (YOLO) series methods based on deep learning had the best detection results. Therefore, in the following paper, we employed a YOLO v5 network to extract moving ships in images [24]. YOLO v5 detection network is a typical object detection network, which has been widely used in many detection tasks and can meet the requirements of real-time detection of moving ships. Although this method is faced with the challenge of long training times, we can train the network offline in the actual ship detection task, so the training time cost is acceptable for the ship detection task.

This study uses a method based on supervised learning to detect the ship. Specifically, the ship detector based on YOLO is built, and a foreground recognition module is inserted into the detector to ensure that for each detected ship object, the detector would output its
specific position in the image coordinate system (usually represented by a rectangular box). In the training process, the YOLO detector model was firstly pre-trained on a public object detection dataset [20], which provided ship-bounding boxes (1000 training samples in total). The parameters of the detector model were fine-tuned by our collected ship dataset (manually annotating ship location) to adapt to the specific scene of intersecting waters.

As shown in Figure 5, the trained detector can successfully detect the moving ship from the images collected by the shore-based camera and give the specific detection bounding box.

![Figure 5. Moving ships’ detection results.](image)

3.3. Visual Measurement Model of Ship Attitude

The position and speed of ships are physical quantities used to describe a ship’s motion state. At the same time, combined with the relationship between static traffic environment and ship position as well as the change of speed, the ship’s future motion intention can be effectively predicted. The posture and motion features of ships can be extracted from vision sensors and radar sensors. Compared with the two-dimensional motion features, images often contain more information. Considering that shore-based sensors collect the data in this study, we only extracted the attitude characteristics of the ship from the images collected by a monocular camera. The measurement model of the ship’s motion characteristics is shown in Figure 6.

![Figure 6. Measurement model of ship motion characteristics. (a) Three-dimensional model of ship image coordinate system to world coordinate system transformation. (b) The measurement model of ship motion parameters.](image)

Assuming the ship is a cuboid, we built a visual measurement model on this basis. Considering that a cuboid has three visible sides, we defined three coordinate systems, namely a physical coordinate system $oxyz$, a camera coordinate system $o'x'y'z'$, and an image coordinate system $XOY$. Suppose $(x, y, z)$, $(x', y', z')$ and $(x, y)$ are the coordinates...
of a point in the object coordinate system, camera coordinate system and image point respectively, then:
\[
\begin{bmatrix}
X_H \\
Y_H \\
H
\end{bmatrix} =
\begin{bmatrix}
F & 0 & u & 0 \\
0 & F & v & 0 \\
0 & 0 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
x' \\
y' \\
z'
\end{bmatrix}
\]
(4)

Suppose that the coordinates of the origin of the camera coordinate system in the physical coordinate system are \((X_c, Y_c, Z_c)\), then the transformation of coordinate points from the physical coordinate system to the camera coordinate system is:
\[
(x', y', z', 1)^T = A(x y z 1)^T
\]
(5)
where \(T\) indicates transpose,
\[
A = R_x(\theta)R_y(\varnothing)R_z(\varphi)D =
\begin{bmatrix}
a & b & c & p \\
d & e & f & q \\
g & h & i & r \\
0 & 0 & 0 & 1
\end{bmatrix}
\]
(6)
\(R_x(\theta), R_y(\varnothing), \) and \(R_z(\varphi)\) are matrices for rotation around the \(x, y,\) and \(z\) axes, respectively. While \(\theta, \varnothing,\) and \(\varphi\) are corresponding rotations, \(D\) is a translation matrix with a translation of \(-(x_c, y_c, z_c)\) and meets the following requirements:
\[-\pi < \theta \leq \pi, -\pi/2 \leq \varnothing \leq \pi/2, 0 \leq \varphi < 2\pi\]
(7)

Substituting Equation (5) into Equation (4), the following relation can be obtained:
\[
(X_H Y_H H)^T = T(x y z 1)^T
\]
(8)
where,  
\[
T =
\begin{bmatrix}
aF + gu & bF + hu & cF + iu & pF + ru & \ldots & \ldots & \ldots \\
dF + gv & eF + hv & fF + iw & qF + rv & \ldots & \ldots & \ldots \\
& g & h & i & r & \ldots & \ldots & \ldots \\
\end{bmatrix} =
\begin{bmatrix}
t_{11} & t_{12} & t_{13} & t_{14} & \ldots & \ldots & \ldots \\
t_{21} & t_{22} & t_{23} & t_{24} & \ldots & \ldots & \ldots \\
& t_{31} & t_{32} & t_{33} & t_{34} & \ldots & \ldots & \ldots \\
\end{bmatrix}
\]
(9)

In short, the problem to be solved was given a group of known physical coordinates of the object to obtain the camera coordinate parameters: \(X_C, Y_C, Z_C, U, V,\) and \(F,\) and the coordinates of the object were converted from the physical coordinate system to the image coordinate system. Assuming that the height of the ship remained constant, then \(z = 0, \theta = 0, \varnothing = 0.\)

3.4. Static Environmental Parameter Measurement Based on Environmental Message

In this study, ships, channel obstacles, and intersection corners were labeled with rectangular boxes to show each object’s relative and absolute positions.

The information of the traffic environment layer affects and restricts a ship’s path choice, which can be divided into the static environment and dynamic environment features. The static environment features refer to the relative position relationship between ships and static environment space, such as channel structure and static obstacles on the water surface, which mainly affect the long-term trajectory planning of ships. The short-term behavior of ships is also affected by the behavior of dynamic objects in the traffic environment, such as other ships, other moving obstacles, etc., which are called dynamic environment features. Both static environment characteristics and dynamic environment characteristics affect the behavior of ships. When identifying the intention of the ship in the intersection area, it is considered that the static environment feature that affects the intention mainly refers to whether the ship arrives at the intersection area. Only when arriving at the intersection area.
area, the possibility of turning will occur. The minimum relative longitudinal distance between the ship and other ships was selected for a dynamic environment feature, which is used to judge whether the ship is in danger and whether it will choose to turn at this moment. In addition, the relationship between the ship and other static environment features, such as channel structure and static obstacles on the channel, will be taken into account in predicting the ship’s future path.

In order to obtain the information from a forward channel structure, ships mainly rely on the electronic chart and an automatic ship identification system to locate. As the research scene selected in this study is a fixed water area, the electronic chart can be used to obtain map information. The shape of the intersection area can be described by map information, and then the relative position relationship between the ship and the intersection area can be obtained by combining the position relationship between the ship and each corner of the intersection area marked by millimeter-wave radar data. That is, the ship’s position in the global coordinate system with the intersection center as the origin was determined, and the 2 km range of the channel at the intersection is the intersection area. Therefore, the longitudinal distance between the ships entering the intersection area from different directions and the red dotted line in the corresponding intersection area can be used as an observation variable to represent the relative position relationship between the ships and the intersection area. In addition, the areas on the water that are not allowed to pass by ships, such as stationary ships or obstacles in the channel, can also be determined in the established 2D map model.

In this study, only ships were considered as traffic participants in the traffic environment. Therefore, the relationship between the ship and the dynamic traffic environment can be considered as the relative relationship between the target ship and other ships. Here, we considered the minimum distance between the target ship and other ships, that is, the minimum distance between the target ship and other ships continuing at their current speed. According to the relative position relationship between the target ship and other ships, and the speed information of the ship collected by millimeter-wave radar, it was used to represent the danger degree of the current situation, that is, whether there was a potential collision risk between the target ship and other ships.

3.5. Ship Intention Identification Based on Static and Dynamic Parameters

In Section 3.1, we extracted the multiple features of ships and determined the factors that affect the intentions of ships. On this basis, this paper proposes a basic idea to build an intention recognition model framework from the three levels (dynamic environment, static environment, and object factors), constructs a dynamic Bayesian network, and describes the intention inference algorithm in detail, which can be used for the intention recognition of ships in confluence waters of inland rivers.

The Bayesian Network (BN) is also called the Belief Network or Directed Acyclic Graph Model (DAGM). Since ship intention needs to be inferred by combining the factors related to ship intention, each node in the Bayesian network corresponds to intent-related factors, observation quantity, and ship intention, respectively. The probability distribution of each variable was inferred by establishing the probability relationship between nodes to realize the ship intention identification.

Figure 7 shows a dynamic Bayesian network where the rectangular nodes represent discrete variables and hidden variables. The shaded rectangular nodes represent intention variables, and the unshaded rectangular nodes represent intent-related factor variables. The circular nodes are continuous variables, i.e., observed quantities. In this network, the probability distribution values of variables were updated by receiving observations at each moment, and the conditional probability relationship between observation nodes and intent-related factor nodes was obtained by prior knowledge and sample data training.
Figure 7. Dynamic Bayesian networks for intention recognition.

Variables in the dynamic Bayesian network mainly included the conditional variable set H and observation variable set O. Conditional variables are all discrete variables that satisfy the Markov hypothesis, and the same node has transition probability relationships at adjacent moments.

The transition probability relation of all nodes in the set of conditional variables can be expressed as:

\[
P(H_t | H_{t-1}) = P(H_t^{dyn} | H_{t-1}^{dyn}) \ast P(H_t^{stat} | H_{t-1}^{stat}) \ast P(H_t^{actc} | H_{t-1}^{actc})
\] (10)

According to the fitting results of the sample data, the probability distribution of the minimum distance \(D_{min}\) between the target ship and other ships conformed to the gamma distribution under the condition of dynamic environmental state \(H^{dyn}\). This is shown in Figure 8 in a static environment state.

H^{stat} condition, the probability distribution of the longitudinal distance between the target ship and the intersection area conformed to a Gaussian distribution. Under the condition of continuous object state \(H^{stat}\), the orientation of the target ship \(\theta\) conformed to the Weibull distribution.

The dynamic Bayesian network can be regarded as a forward-filtering process, and the probability distribution of each variable can be updated when the observed variables are received so as to realize the process of intention inference. In the process of intention inference, the assumed density filtering is adopted as the inference tool. The process of intention inference can be divided into prediction and update.
1. Prediction

The prediction process is to predict the prior distribution of the current moment through the posterior distribution of the previous moment and the fixed transition probability. Based on the joint posterior distribution of the previous moment, the prior joint distribution of the current moment can be calculated according to the transfer probability, and the edge distribution can be obtained by adding, as shown in Equations (11) and (12).

\[
P(G_t, G_{t-1}, H_t, H_{t-1}) = P(G_t | G_{t-1}, H_t) \ast P(H_t | H_{t-1}) \ast \hat{P}_{t-1}(G_{t-1}, H_{t-1})
\]

\[
P(G_t, H_t) = \sum_{G_{t-1}} \sum_{H_{t-1}} P(G_t, G_{t-1}, H_t, H_{t-1})
\]

2. Update

The updating process is to update the posterior distribution of the current moment according to the observed variables of the current moment. Based on the joint prior distribution obtained in the prediction step, the joint posterior distribution at the current moment can be calculated and added according to the observation variables and conditional probability relationship. The posterior distribution of the nodes obtained is shown in Equations (13) and (14).

\[
\hat{P}(G_t, H_t) \propto P_t(O_t | H_t) \ast \hat{P}_t(G_t, H_t)
\]

\[
\hat{P}_t(G_t) = \sum_{H_t} \hat{P}(G_t, H_t)
\]

4. Experiment and Result

This section includes the following parts: (1) experimental scenario design; (2) data set introduction; (3) simulation results; (4) real ships experiments; (5) model evaluation.

4.1. Experimental Scenario Design

In the experiment, the ship traffic scene at the intersection of the Hanjiang River and Yangtze River in Wuhan was selected as the experiment scene. This region covers longitudes 114°26′–114°30′1″ E and latitudes 30°53′8″–30°59′0″ N, as shown in Figure 9.

![Figure 9. Schematic diagram of experimental data acquisition environment.](image-url)

In the above inland river intersection scenario, the behaviors of ships were divided into three categories: straight, turn left, and turn right. In addition, combined with the intersection channel structure, it can be divided into up-straight, up-right turn, down-
straight, down-left, left-up, and left-down, respectively, corresponding to the situation when ships enter the intersection area from the three different directions of the intersection. According to the above classification method, the collected data can be divided into six scenarios, and the schematic diagram of each scenario is shown in Figure 10. The specific definition of each scenario is as follows:

![Schematic diagram of ship navigation traffic scene](image)

**Figure 10.** Schematic diagram of ship navigation traffic scene. Scenario (a): The ship enters the intersection area from the lower reaches of the Yangtze River, and the ship goes straight up. Scenario (b): The ship enters the intersection area from the lower reaches of the Yangtze River, and the ship turns right. Scenario (c): The ship enters the intersection area from the upper reaches of the Yangtze River, and the ship goes straight down. Scenario (d): The ship enters the intersection area from the upper reaches of the Yangtze River, and the ship turns left. Scenario (e): The ship enters the intersection area from the Han River, and the ship turns right, across the river, berthing. Scenario (f): The ship enters the intersection area from the Han River, and the ship turns right.

### 4.2. Data Set Introduction

Currently, the available ship datasets contain fewer scenarios, and the sampling frequency is usually low. The data set of this study is from the surveillance video of the intersection between the Yangtze River and the Han River, and the video data of 86 groups of ships were established. Each set of data recorded a ship that was close to the intersection area and intended to cross the intersection. The longest time of each set of data series was 5 min, and the shortest time was 1 min. After the data was processed by the frame difference method, there were a total of 800 images with a resolution of 1280 × 720. We randomly selected 600 of them as the training set and the remaining 200 as the test set. In the training set, we labeled the ships in it. We used this labeled training set to train the YOLO model and optimize the parameters of the network. When the video was fed into our detection system, the image was first obtained and preprocessed at a certain frame rate. Then, these images were inputted into the YOLO object detection network. Through this detection network, we can extract the target ship from each frame and obtain the position and motion parameters. Thus, the results of the detection and location of the ship in the video were obtained.

### 4.3. Simulation Results

In order to verify the feasibility of the algorithm proposed in this paper, two scenarios of the ship turning right and the ship going straight were simulated before the actual ship experiment. The simulation results are shown in Figures 11–13.
Figure 11. Simulation intention prediction results of rightward ship. (a) Simulation of ship movement trajectory. (b) Ship speed and direction.

Figure 12. Simulation intention prediction results of direct ship. (a) Simulation of ship movement trajectory. (b) Ship speed and direction.

Figure 13. Simulation intention prediction results of direct ship. (a) The turning probability of ship1. (b) The turning probability of ship3.
Assuming that the speed of the ship remains stable when passing through the intersection area, Figure 11a shows that the ship traveled along a straight line. At this time, the included angle between the ship and the observation point gradually decreased, so the radial speed of the ship in Figure 11b gradually increased. Figure 12a shows that the ship turned right, and the included angle between the ship and the observation point first gradually decreased and then remained unchanged. Therefore, the radial velocity of the ship in Figure 12b first gradually increased and then changed smoothly. As shown in Figure 13, the conclusion obtained by calculating the turning probability of the ship is that the turning probability of the target in Figure 11 is 0.34, and it was determined that the turning will not occur. The turning probability of the target in Figure 12 is 0.81 and it was determined that the turning would occur. Therefore, from the simulation results, the ship intention recognition results were in line with the scene designed by the simulation experiment.

4.4. Real Ship Experiments

In order to more accurately reflect the actual ship navigation, several sets of real ship tests were carried out to evaluate the accuracy of the proposed model for predicting ship intentions in cross channels. We obtained radar and video data collected by monitoring equipment at the Yangtze River and Han River intersections from September 2020 to October 2020. The intersection and dense traffic flow make the study area a highly complex cross-channel, and after data preprocessing and sampling, 500 pieces of data containing moving ships were obtained. We illustrate the prediction results of ship sailing intention in typical scenarios.

The testing process of each trajectory is divided into the following three steps.

1. Target detection and tracking. The ships in the 0–2 km region from the intersection area are detected, and the pixel coordinates and radar coordinates of the ships are output.
2. Track segmentation. The 0–2 km area is divided into 10 track segments, and the length of each track segment is 0.2 km.
3. Predicted intent. The observed trajectory sequences are input to the HMM, LSTM, and the Ours models, and the predicted intent labels and probabilities for each intent class are determined. Return to step 2.

In Figure 14, the black line represents that the ship does not change speed and direction during navigation at the intersection, the green line represents the ship is turning right, and the red line represents the ship is turning left. Next, we analyzed the ship movement process in the scenario. We found that Ship 1 would turn right when entering the monitoring area, and the speed and direction would remain unchanged after turning, while Ship 2 and Ship 3 maintained their original speed and direction. The predicted results of our algorithm are consistent with the actual results (we can obtain the actual sailing results from the surveillance videos in advance by eye).

Figure 15 shows the movement process in this scenario, found that Ship 1 and Ship 2 kept their original speed and direction to move. Ship 3 first maintained speed and direction after entering the monitoring area and then turned right after arriving at the intersection area. After turning right, Ship 3 maintained its speed and direction for the rest of the voyage. The actual results are also consistent with the predicted results of the algorithm.

4.5. Model Evaluation

We conducted numerical experiments on HMM, LSTM, and the Ours ship intention identification method to predict the intent classes of the 500 tested trajectories. To quantitatively evaluate the prediction performance of the three models, we used accuracy and mean square error (MSE) as measures. The mean and variance of the accuracy and MSE at different distances to the precautionary area are listed in Table 2.
Figure 14. Prediction results of ship navigational intention in typical scenarios of intersection channel. (a) Ships detection results. (b) Intention identification probability of Ship 1. (c) Intention identification probability of Ship 2. (d) Intention identification probability of Ship 3.

Figure 15. Prediction results of ship navigational intention in typical scenarios of the intersection channel. (a) Ships detection results. (b) Intention identification probability of Ship 1. (c) Intention identification probability of Ship 2. (d) Intention identification probability of Ship 3.

Table 2 shows the accuracy of the intention prediction of the three models. Each test takes into account three different distances (0, 1, and 2 km) from the intersection area. The proposed model is better than the LSTM model and HMM model. For example, when $d = 0$ km, the accuracy of our model is about 30% higher than HMM model and 10% higher than the LSTM model. At $d = 1$ km, the accuracy of our model is about 12% higher than
HMM and 14% higher than the LSTM model. When the distance from the intersection area increased to 2 km, the preparation rate of the three models decreased. This is an expected result because early prediction is more challenging than later prediction. According to the regulations of the People’s Republic of China on river collision prevention, the safety distance of ships with a length of more than 30 m is 2 km, and that of ships with a length of less than 30 m is not less than 1 km. Therefore, our algorithm can effectively avoid collision by predicting the ship’s intention within 2 km.

Table 2. Mean (variance) accuracy and MSE of intent prediction models at different distances to precautionary area.

<table>
<thead>
<tr>
<th>Model</th>
<th>Measure</th>
<th>d = 2.0 km</th>
<th>d = 1.0 km</th>
<th>d = 0.0 km</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMM</td>
<td>Accuracy</td>
<td>0.522</td>
<td>0.619</td>
<td>0.612</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.592</td>
<td>0.445</td>
<td>0.421</td>
</tr>
<tr>
<td>LSTM</td>
<td>Accuracy</td>
<td>0.446</td>
<td>0.594</td>
<td>0.814</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.636</td>
<td>0.508</td>
<td>0.206</td>
</tr>
<tr>
<td>Ours</td>
<td>Accuracy</td>
<td>0.652</td>
<td>0.734</td>
<td>0.912</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.541</td>
<td>0.343</td>
<td>0.016</td>
</tr>
</tbody>
</table>

5. Conclusions

The intention prediction of ships at intersections can effectively reduce the occurrence of ship collisions. However, the prediction accuracy of ship intention is easily affected by the validity and real-time data. In this paper, we propose an intention prediction model based on the fusion of video and radar data by using the Bayesian framework, and the model is verified on the real channel data at the intersection of the Yangtze River and the Han River. It was found that the ship motion intention is highly correlated with ship motion parameters and environmental factors. In order to effectively utilize this finding, we introduced a Bayesian framework and finally calculated the probability of ship motion intention by reasonably assuming the probability distribution of different factors. Due to the high acquisition frequency of an image and radar monitoring data, ship motion can be stably tracked, which accurately predicts ship intention and improves the real-time decision-making, thus effectively solving the problem of poor real-time prediction of ship intention caused by the data delay of other sensors.

It is undeniable that there are still some shortcomings in the algorithm. For example, it is necessary to integrate the environmental data, video, and radar data of specific areas to identify the ship’s intention, which is difficult to be transplanted into the monitoring equipment of mobile ships. In future work, the cross-domain adaptive scene understanding method based on radar and video research will be considered. Then, the ship’s intention recognition can be based on the results of dynamic scene understanding, which can be transplanted to mobile ships without static environment data fusion. This will provide a decision-making basis for intelligent ship collision avoidance.

Author Contributions: Conceptualization, Q.C. and C.X.; methodology, Q.C. and C.X.; software, Q.C.; validation, Q.C.; formal analysis, Q.C.; investigation, Q.C.; resources, Y.W.; data curation, Q.C. and M.T.; writing—original draft preparation, Q.C.; writing—review and editing, Q.C.; visualization, Q.C. and M.T.; supervision, Y.W.; project administration, W.Z.; funding acquisition, C.X. and Y.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Natural Science Foundation of Shandong Province under Grant ZR2020KE029; by the National Natural Science Foundation of China under Grant 52001241; by the 111 Project(B21008); by the Zhejiang Key Research Program under Grant 2021C01010.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No public data sets were used in the study.
Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

5. Gao, M.; Shi, G.; Li, S. Online prediction of ship behavior with automatic identification system sensor data using bidirectional long short-term memory recurrent neural network. Sensors 2018, 18, 4211. [CrossRef]
17. Praczyk, T. Using evolutionary neural networks to predict spatial orientation of a ship. Neurocomputing 2015, 166, 229–243. [CrossRef]