THE JOURNAL OF NAVIGATION (2018), **71**, 1195–1209. © The Royal Institute of Navigation 2018
doi:10.1017/S0373463318000085

Visual Analytics Approach to Vessel Behaviour Analysis

Liang Jin[1*], Zhengyi Luo[2*] and Shu Gao[1]

[1](*School of Computer Science and Technology, Wuhan University of Technology, Wuhan 430063, China*)

[2](*School of Engineering and Applied Science, University of Pennsylvania, Philadelphia 19104 USA*)

(E-mail: gshu@whut.edu.cn)

Vessel behaviour analysis plays an important role in maritime situational awareness. However, available technology still provides only limited approaches to vessel behaviour analysis. In this paper, we propose a visual analytics framework to interactively explore the characteristics of vessel behaviour by means of integrating visualisation with data mining and a human-computer interaction controlling model, which combines human insight with the enormous storage and processing capacities of computers to gain insight into vessel behaviour. In addition, we provide multiple views for visually analysing vessel trajectories, densities and speeds. Case studies with 15 days' AIS data collected from the middle Hankou channel to Yangluo channel in the Yangtze River demonstrate the effectiveness of our approach.

KEYWORDS

1. Automatic Identification System (AIS). 2. Vessel behaviour. 3. Vessel trajectory. 4. Visual Analytics.

Submitted: 17 October 2017. Accepted: 10 February 2018. First published online: 2 April 2018.

1. INTRODUCTION. Waterway traffic monitoring is important for marine security and safety. Recently, in efforts to improve vessel safety, many efforts have been made in vessel behaviour analysis. The Automatic Identification System (AIS) is an automatic tracking system for identifying and locating vessels by exchanging data with other nearby vessels and AIS base stations. The data provided by AIS can reflect vessels' behaviour in real-time. Due to the high information density nature of this data, the analysis of vessel behaviour based on AIS data has become a popular research topic. The characteristics of AIS data, such as its time-space attributes, large volumes, and real-time features also bring challenges to current data analysis and user comprehension methods.

Today, data mining is widely employed in analysing vessel behaviour. However, it only displays the final results to users, so it cannot interact for further analysis of the data.

Visual analytics (Thomas and Cook, 2006) is a multidisciplinary field of scientific/information visualisation, human-computer interaction, cognitive science and data mining.

* Liang Jin and Zhengyi Luo contributed equally to the work and should be considered co-first authors.

https://doi.org/10.1017/S0373463318000085 Published online by Cambridge University Press

It couples the powerful cognitive ability of the human brain with the analytical ability of computers. Based on data mining and other technologies, especially Human and Computer Interaction techniques (HCI), it enables information and knowledge from data to be discovered and gained more directly and effectively.

This work aims to explore the characteristics of vessel behaviour by means of integrating data mining with visual analytics and building an HCI control model which allows Vessel Traffic Services (VTS) supervisors and operators to combine their human experience with the enormous storage and processing capacities of computers to gain insight into vessel behaviour. Using advanced visual interfaces, VTS supervisors and operators will be able to directly interact with the data, allowing them to make well-informed decisions in complex navigation situations.

This paper is organised as follows. In Section 2, related work is discussed. Section 3 presents a framework of visual analytics for vessel behaviour analysis. Our visual analysis approaches to vessel trajectory, vessel density distribution estimation and vessel speed are proposed in Sections 4, 5 and 6, respectively. Section 7 introduces a prototype system and Section 8 shows an experiment using real-world data. In the final section, conclusions and suggestions for future work are provided.

2. RELATED WORK. Object movement analysis in general is a hot topic in many fields of research, including Geographic Information Science, data mining and visual analytics. Wang et al. (2013) presented an interactive system for visual analysis of urban traffic congestion based on Global Positioning System (GPS) trajectories. Traffic jam information is extracted and derived from these trajectories. Spatially and temporally related events are concatenated in traffic jam propagation graphs. These graphs form a high-level description of a traffic jam and its propagation in time and space. Ferreira et al. (2013) described a system that is able to visually explore big origin-destination and spatio-temporal data. The system supports interactive response times, makes use of an adaptive level-of-detail rendering strategy to generate clutter-free visualisation for large results, and shows hidden details to the users in a summarised report through the use of overlay heat maps. Tominski et al. (2012) proposed an approach to visualising trajectory attribute data which covers space, time and attribute values. A Two-Dimensional (2D) map serves as a reference for the spatial context, and the trajectories are visualised as stacked Three-Dimensional (3D) trajectory bands along which attribute values are encoded by colour. In addition, time is integrated through appropriate ordering of bands and through a dynamic query mechanism that feeds temporally aggregated information to a circular time display.

Some research efforts have focused on visualisation in maritime navigation. Willems (2011) introduced three approaches to visualising vessel trajectories: vessel density, density maps and composite density maps. The first approach shows two important features for vessel traffic, anchor zones (stopping areas) and sea lanes (common routes). The second approach is a generalisation of vessel density, which enables users to combine the values of the density fields with density aggregation and make a density map from this single density field. The third approach allows the user to define the way density fields are computed with a number of predefined blocks based on experience and domain knowledge. Scheepens et al. (2016) used a density map of the trajectories combined with moving particles to visualise locations of AIS data and design a toolkit to select AIS data through user specified movement directions. Cazzanti et al. (2016) described a product, the Maritime Patterns-of-Life Information Service (MPoLIS), which provides the maritime industry, governments,

https://doi.org/10.1017/S0373463318000085 Published online by Cambridge University Press

and international organisations with visual analytics on vessel traffic in seaports. An intuitive and interactive interface for Subject Matter Experts (SMEs) in the maritime domain was also given by this approach. Wang et al. (2017) proposed a trajectory visual analytics tool, TraSeer. It summarises vessel movement trends, detects anomalous movement paths and allows the user to explore vessel movement data interactively. Lu and Gao (2016) combined a data clustering method and parallel coordinates to explore the relationship between the multiple factors of ship accidents in inland rivers. This work also uses human-computer interaction technology and the "multiple views" visualisation method to analyse the key factors that cause ship accidents in inland rivers. Sidibe and Gao (2017) presented a systematic review of state-of-the-art automatic anomalous maritime vessel behaviour detection techniques based on AIS movement data, including statistical analysis, machine learning, visualisation and data mining. Riveiro and Falkman (2009) developed an interactive visualisation system, based on a combination of Self-Organising Maps (SOM) and Gaussian Mixture Model (GMM), which represents both normal behavioural statistical models and expert knowledge-based rules, used in the detection of anomalous behaviour in maritime traffic data.

In contrast to the visual analysis approaches mentioned above, which mainly focus on the visual representation aspects, our work focuses on combining data mining with visualisation technology to build interactive models for vessel behaviour analysis. In addition, our work factors in the multi-dimensionality of vessel behaviour data, including vessel trajectories, densities and speeds. Our contributions are to offer new strategies and solutions for the visual representation of maritime navigation information and to highlight the importance of understanding all the interactions and processes that happen in maritime navigation.

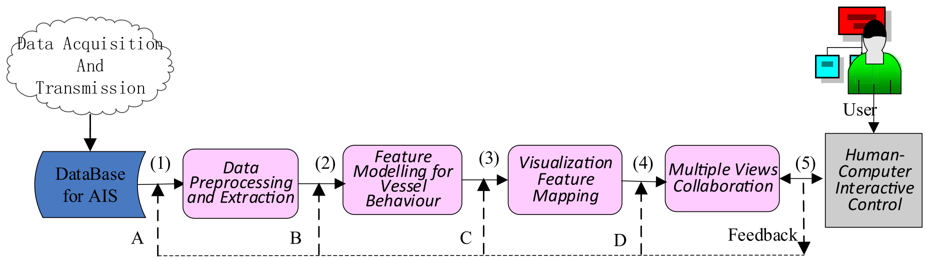
3. VISUAL ANALYTICS FRAMEWORK FOR VESSEL BEHAVIOUR ANALYSIS.

The visual analytics framework for analysing vessel behaviour consists of five modules: *Data Pre-processing and Extraction, Feature Modelling for Vessel Behaviour, Visual Feature Mapping, Multiple Views Collaboration*, and *Human-Computer Interaction Control*, as shown in Figure 1.

In Figure 1, the subsections are as follows:

(1) *Data Pre-processing and Extraction*. This includes data cleansing, data subset extraction and related visualisation attribute processing.
(2) *Feature Modelling of Vessel Behaviour*. This constructs three feature models, each based on vessel trajectories, density and speeds.
(3) *Visual Feature Mapping*. This transforms vessel behaviour characteristics data into geometry data and establishes a mapping relation between image elements.
(4) *Multiple Views Collaboration*. This provides users with multiple views and multiple interactive methods, which enable users to track, detect and analyse vessel behaviour more intuitively.
(5) *Human-Computer Interactive Control*. This receives requests from users and sends interactive instructions to the *Multiple Views Collaboration* module and then acquires feedback information from the *Multiple Views Collaboration* module and dispatches the information (A, B, C and D, as shown in Figure 1) to different modules. Then the modules process the feedback information and respond interactively.

https://doi.org/10.1017/S0373463318000085 Published online by Cambridge University Press



Where : (1) Extraction (2)Data Mining (3)Transformation (4) Mapping and Rendering (5)Interaction
A:Data Extraction B:Model Reconstruction C:Mapping Configuration D:Parameter Adjustment

Figure 1. Visual Analytics Framework for Vessel Behaviour.

As shown in Figure 1, *A*, *B*, *C* and *D* represent different interactive feedback tasks for different modules. *A* indicates interactive re-extraction of the data set. *B* denotes model correction and reconstruction, such as setting cluster threshold value for mining vessel features. *C* represents the modification of the data mapping, such as setting speed level and location of the trajectory cluster subset. *D* is interaction operations, such as zooming in, zooming out and refreshing views and setting the size of the water area grids.

4. VISUAL ANALYTICS APPROACH TO VESSEL TRAJECTORY ANALYSIS BASED ON CLUSTERING FEATURES.

4.1. *Feature model of vessel trajectory.* In order to build the vessel trajectory feature model for visual analysis, a trajectory clustering algorithm is applied. It is based on the tree clusters method proposed by Piciarelli et al. (2005). There are two advantages of this method. Firstly, we improve the Euclidean distance measurement (when a match between a trajectory and a cluster is found, the cluster must be updated). This way each cluster is a dynamic approximation of the mean and variance of trajectories that matched it, with an exponentially decreasing weight for the older trajectories. Secondly, we use tree clusters, which eliminates the trajectory division processes without loss of local cluster features. The clusters are dynamically built in real-time as the trajectory data is acquired, saving the trouble of off-line processing.

However, in order to gain better trajectory clustering results from AIS data and increase clustering efficiency, the following improvements are implemented. Firstly, to reduce invalid matches, a method for trajectory pre-classification is proposed: we select the starting points of all trajectories clustered by a Density-Based Spatial Clustering of Applications with a Noise (DBSCAN) algorithm (Ali et al., 2010) and classify the trajectories based on their starting points. The classification can be used for future trajectories' clustering and selection. This method can be regarded as building indices for trajectory clustering, which will significantly reduce the matching failure rate. Secondly, a new method of updating clustering is presented to support the trajectory contribution rate: adding a support ct_j to each feature point j of the cluster. When the cluster C only consists of one trajectory, the ct_j of each point is 1. Each time a feature point is updated, we add 1 to ct_j. This support can also yield the overall arithmetic mean of the whole pool of parameters when the parameters of every point are updated.

https://doi.org/10.1017/S0373463318000085 Published online by Cambridge University Press

The trajectory is defined as follows:

$$T_i = \{t_{i1}, \ldots, t_{in}\}, \text{ where } t_{ij} = \{x_{ij}, y_{ij}\} \tag{1}$$

Each trajectory T_i is represented by a list of vectors t_{ij} which gives the spatial position of the vessel i at time j.

The improved cluster is defined as Equation (2).

$$C_i = \{c_{i1}, \ldots c_{in}\} \text{ where } c_{ij} = \{x_{ij}, y_{ij}, \sigma_{ij}, ct_{ij}\} \tag{2}$$

where C_i denotes cluster i, and c_{ij} is the feature point of the cluster. Every feature point comprises the trajectory point (x_{ij}, y_{ij}), local approximation σ_{ij} of the cluster variance (Piciarelli et al., 2005) and support ct_{ij} at time j.

In order to check if a trajectory fits a given cluster, a distance measure must be defined. The distance between a trajectory $T = \{t_1 \ldots t_n\}$ and a cluster $C = \{c_1 \ldots c_n\}$ is defined as Equation (3), which is as in Piciarelli et al. (2005) and is an improved Euclidean distance measure.

$$D(T, C) = \frac{1}{n} \sum_{i=1}^{n} d(t_i, C) \tag{3}$$

where

$$\begin{gathered} d(t_i, C) = \min_j \left(\frac{dist(t_i, c_j)}{\sqrt{\sigma_j}} \right) \\ j \in \{\lfloor (1-\delta)i \rfloor \ldots \lceil (1+\delta)i \rceil\} \end{gathered} \tag{4}$$

$dist(t_i, c_j)$ is the usual Euclidean distance. The distance of a trajectory from a cluster is thus the mean of the normalised distances of every trajectory element t_i from the nearest cluster element c_j found inside a temporal window centred in i and with a variable size, in order to permit matching under accumulating temporal differences. The window is clipped in the range $[1 \ldots m]$ and, if it completely falls outside this range, the distance D is set to ∞.

However, a different method is used to update the cluster when a match between a trajectory and a cluster is found. The improved method uses arithmetic mean instead of a fixed parameter α (Piciarelli et al., 2005). If elements $t_i = \{x_i, y_i\}$ and $c_j = \{x_j, y_j, \sigma_j, ct_j\}$ match (c_j is the cluster element that is cluster feature point, nearest to t_i inside the temporal window), then c_j is updated as follows:

$$\begin{cases} x_j = \left(\dfrac{ct_j}{ct_j + 1}\right) x_j + \left(\dfrac{1}{ct_j + 1}\right) x_i \\ y_j = \left(\dfrac{ct_j}{ct_j + 1}\right) y_j + \left(\dfrac{1}{ct_j + 1}\right) y_i \\ \sigma_i = \left(\dfrac{ct_j}{ct_j + 1}\right) \sigma_j + \left(\dfrac{1}{ct_j + 1}\right) (dist(t_i, c_j))^2 \\ ct_j = ct_j + 1 \end{cases} \tag{5}$$

where $dist(t_i, c_j)$ is the same as in Equation (4).

https://doi.org/10.1017/S0373463318000085 Published online by Cambridge University Press

Table 1. Description of TrajVisualise.

```
Algorithm Name: TrajVisualise
Input: Vessel trajectory clustering result set SSR = {StR1, StR2, ..., StRm}
Output: Visualisation view of trajectory clustering
TrajVisualise(SSR){
/*plotting every cluster subset*/
   Foreach (StR in SSR){
     Foreach (tree in StR)
        PaintTree(tree);        //plotting a tree cluster
   }
   plotting trajectory cluster view; //Algorithm is over
}
/*The function for plotting tree cluster*/
PaintTree (Cluster tree)    {
   /*Plotting the trajectory cluster segments based on the location coordinates,
   Support and local approximation of each feature point */
   Foreach (point in tree.TrajPoint[])
        paintSeg(point.location, point.ct, point.radius);
   /*plotting each child node in this cluster */
   Foreach ( child in tree.child[]) {
     PaintTree(child);
   }
}
```

4.2. *Visual analysis of vessel trajectory.* Since vessel behaviour is a study on navigation features of vessel groups, not on those of an individual vessel, a visual analysis approach based on clustering vessel trajectories (Clustering Visualisation for short) can be better for showing the vessel trajectory distribution and vessel navigation trends.

Clustering Visualisation is a method for visualising the results of clustering analysis of the vessel trajectories, which is described in Section 4.1. The result is a set $SSR = \{StR_1, StR_2, \ldots, StR_m\}$ which consists of multiple subsets; each subset is a forest of clusters. Visualising the result means plotting every cluster in result set SSR on a map. However, it is necessary to distinguish the trajectory support and the degree of clustering during the plotting. "Cluster" is similar to trajectory and can be represented by a series of point coordinates, but it also contains the support ct_{ij} and local approximation σ_{ij} for each point, as shown in Equation (2). Support ct_{ij} is the number of trajectories in the cluster while σ_{ij} represents the average variance of the trajectories around the cluster centre. Therefore, support ct_{ij} reflects the frequency of the trajectories, and local approximation σ_{ij} the clustering degree of the cluster, can also be regarded as the extent of the trajectory close to the cluster centre. In this paper, clustering degree is distinguished by different trajectory widths, and the frequency is distinguished by colour depths. The description of algorithm of Clustering Visualisation (TrajVisualise for short) is shown in Table 1.

4.3. *Interactive visualisation model for analysing vessel trajectory.* Visualisation technology cannot provide users with dynamic analysis, so human-computer interaction methods are also added. Based on our algorithm TrajVisualise, an interactive visualisation model for analysing vessel trajectory is defined as follows.

(1) *Positioning starting point for trajectory cluster subsets*. By pre-classifying, trajectories are divided into different subsets. Users can forecast the trajectory

https://doi.org/10.1017/S0373463318000085 Published online by Cambridge University Press

direction by fixing the position of the start point of a trajectory cluster subset using this interactive visual analysis approach.

(2) *Positioning ending point for trajectory cluster subsets*. Trajectories which have the same starting point may move towards different directions, which forms a tree structure of trajectory clusters. However, the trajectories which have the same end point usually belong to the same course. Analysis of a trajectory cluster with a particular end point is helpful for the analysis of a particular course.

(3) *Setting threshold in the clustering algorithm*. Setting the threshold for trajectory similarity is the key to the clustering effect. Through human-computer interaction to change the threshold, users can more intuitively and quickly observe the clustering effect under different thresholds and enable the selection of a more suitable threshold.

5. VISUAL ANALYTICS APPROACH TO VESSEL DENSITY DISTRIBUTION ANALYSIS BASED ON KERNEL DENSITY ESTIMATION.

5.1. *Feature model of vessel density distribution.* Vessel density refers to the number of vessels per unit water area, such as per square kilometres or per 100 square kilometres. However, vessel density distribution is different from vessel density, which represents changes of vessel density across different locations. Vessel density distribution is very important in analysing waterway condition and maritime accidents.

In this paper, kernel density estimation is used to model vessel density distribution. Vessel density distribution can be treated as the distribution of vessels in 2D. The kernel density estimation function is defined as Equation (6).

$$\hat{f}_{2D}(x) = \frac{1}{nh_{1,1}h_{2,2}} \sum_{i=1}^{n} K\left(\frac{(x - x_i)}{h_{1,1}}, \frac{(y - y_i)}{h_{2,2}}\right) \tag{6}$$

where n is the number of points whose distance from point (x, y) is less than $h_{1,1}$ along the x-axis and less than $h_{2,2}$ along the y-axis. $h_{1,1}$ and $h_{1,2}$ represent the attenuation threshold along the x-axes and y-axes respectively. $K(\mathrm{x}, \mathrm{y})$ is a two-dimensional Gaussian kernel.

Based on the feature of kernel density estimation, this paper focuses on the following problems when modelling vessel density distribution.

(1) A kernel density estimation matrix requires too much computation power: to reduce the memory and Graphics Processing Unit (GPU) capacity consumption during plotting, grid partitioning is adopted. Grid partitioning divides the water area into equal-sized grids, and separately calculates a kernel density estimation matrix for each grid.

(2) How to obtain instantaneous density information: drawing a heat map for vessel density distribution requires overlaying vessel density information from each time entry. However, the frequency of AIS message transmissions depends on vessel speed, so different vessels' position information is not synchronised. Thus, in the dynamic AIS database, position information for a vessel at time t is queried by combining the data from a range of predefined time zones centred at time t. Then the closest position to t is chosen as the instantaneous position information. Since approximately 5 minutes is the maximum gap between AIS reports, a range of 300 seconds is chosen.

(3) Removing the repeated samplings for vessel position information: the data for each vessel position is inserted into a dynamic AIS data table as soon as it is obtained. The

https://doi.org/10.1017/S0373463318000085 Published online by Cambridge University Press

number of the grid in which the vessel is located at time t can be acquired by searching the records which have the same Maritime Mobile Service Identity (MMSI) and are closest to time t in the table. Thus, using this method we can avoid the problem that a vessel passes through several grids during the same time quantum.

5.2. *Visual analysis of vessel density distribution.* A heat map is the most common and effective way to express density distribution. In this paper, a density distribution heat map is plotted using kernel density estimation which has the advantage that users can interact with a heat map in more depth by changing the attenuation threshold of kernel density estimation.

The procedure of the visual algorithm for vessel density distribution analysis based on kernel density estimation is as follows:

(1) Mesh the water area, and obtain grids of equal size;
(2) Calculate the kernel density estimation matrix for each grid in turn;
(3) Take the estimated value of the kernel density estimation matrix at each point and map it with the heat map colour;
(4) Mark the points from the kernel density estimation matrix in the map;
(5) Plot a heat map for each grid, and then combine them to form a heat map for the overall water area.

5.3. *Interactive visualisation model for analysing vessel density.* Since kernel density estimation is used to model the features of vessel density distribution, an interactive analysis model is built by means of setting the attenuation threshold. The only parameter h, the attenuation threshold, in kernel density estimation, influences the degree of detail in the heat map. The bigger the attenuation threshold h, the smaller the difference of density estimation from neighbouring points. The smaller the attenuation threshold, the bigger the difference of the density estimation from the neighbouring points. Therefore, changing the attenuation threshold can effectively change view precision and deliver detail information of different vessel densities.

6. VISUAL ANALYTICS APPROACH TO VESSEL SPEED ANALYSIS BASED ON GRID.

6.1. *Feature model of vessel speed distribution.*

6.1.1. *Definition of vessel speed distribution.* In a narrower sense, vessel speed distribution is the spatial distribution of average speed of all vessels in a water area. The speed distribution shows the speed of the vessel at different positions. However, in a broader sense, vessel speed distribution also includes the probability distribution of vessel speed at different speeds, which is called the vessel speed probability distribution. Researching vessel speed probability distribution can express the gradient variation of vessel speeds and find the relationships governing vessel group speed. Therefore, vessel speed distribution in this paper contains the vessel speed spatial distribution and the vessel speed probability distribution.

6.1.2. *Feature model of vessel speed spatial distribution.* First of all, the water area is meshed. Then the model of vessel speed distribution based on this grid is built as follows:

$$\bar{v} = \frac{\bar{v}_1 + \bar{v}_2 + \bar{v}_3 + \cdots + \bar{v}_n}{n} \tag{7}$$

https://doi.org/10.1017/S0373463318000085 Published online by Cambridge University Press

where $\bar{v}_1$, $\bar{v}_2$, $\bar{v}_3, \ldots, \bar{v}_n$ are the average speeds of single vessels, n represents the total number of vessels in the current water area and $\bar{v}$ represents the average speed.

According to the Equation (7) and the dynamic AIS data table, Equations (8) and (9) can be used to calculate the average vessel speed in the gridded water area.

$$\textit{Average speed of a vessel in a grid} = \frac{\sum \textit{speed of the vessel's records in the grid}}{\textit{The number of records}} \quad (8)$$

$$\textit{Average speed of vessels in a grid} = \frac{\sum \textit{average speed of every vessel in the grid}}{\textit{The number of vessels}} \quad (9)$$

6.1.3. *Feature model of vessel speed probability distribution.* A view is provided, which can express speed probability distribution in each grid to further explore vessel speed distribution. Vessel speeds are divided into different intervals to obtain speed levels and the view shows the probability distribution of vessel speed of every grid at different speed intervals using a histogram. Thus, the gradient variation of speed in every grid is presented.

Suppose vessel speed is divided into m intervals, and $p_1, p_2, \cdots, p_m$ are the probabilities of speeds falling into different intervals. Two requirements must also be met: $\sum_{i=1}^{m} p_i = 1$ and $p_i = \frac{n_i}{N}$, where n_i is the number of vessels whose average speed is in the speed interval i and N is the number of all vessels. In order to plot the view, the water areas are divided into grids, and the model of speed probability distribution for each grid is then built.

6.2. *Visual analysis of vessel speed distribution.*

6.2.1. *Visual analysis of vessel speed spatial distribution.* In order to provide the spatial distribution information of vessel average speed and a convenient interaction with users, this paper proposes a meshed speed distribution map which divides the water area into grids and calculates the average vessel speed in a single grid. Different speed levels can be represented by colour coding.

The key points of the approach are as follows:

(1) Grid division of water area: because it is difficult for a user to calculate the exact latitude and longitude of every grid when the water area is meshed, we use the metre as the metric unit of division and calculate the accurate latitude and longitude range by coordinate projection transformation.
(2) Colour coding of the speed: the grids with different speed levels have differing colours according to graded colour coding. Therefore, the hidden information in the data can be conveyed intuitively to the users by colour.

The steps of the visualisation approach are as follows:

(a) Select an appropriate grid width wf, hf (wf, hf represent the width and height of a single grid horizontally and vertically in metres), and divide the water area into grids;
(b) Grade the speeds and code the speed colours to construct the mapping relation of speed level and colour coding;
(c) Calculate the average speed of vessels in each grid using Equation (7) in Section 6.1.2, and obtain the list of speeds, called SpeedList, and acquire related speed levels according to the average speed of vessels in every grid in the SpeedList in order to build the mapping relation of grid and colour coding;
(d) Fill the grids with colours by means of the above mapping relation.

https://doi.org/10.1017/S0373463318000085 Published online by Cambridge University Press

6.2.2. *Visual analysis of vessel speed probability distribution.* For each grid, a histogram of the speed probability distribution will be plotted. The histograms provide the probabilities of vessel speed at different speed intervals. They provide more statistical information than a speed spatial distribution map.

6.3. *Interactive visualisation model for analysing vessel speed.* In this section we aim at visually analysing vessel speed spatial distribution and vessel speed probability distribution. The interactive visualisation model is defined as follows:

(1) *Setting the grid parameters*. In our model of vessel speed spatial distribution which is based on grid-enabled colour coding, grid height *hf* and grid width *wf* influence the accuracy of the visualisation result and processing efficiency. When *wf* and *hf* are too large, the visualisation result does not reflect the gradient variation of vessel speed in different areas accurately, and when *wf* and *hf* are too small, the time consumed in dividing the grid and computing average vessel speed is too high. Therefore, users need to adjust the grid scale to achieve different views of speed distribution.
(2) *Setting the speed level*. Expressing speed gradient variation by different colours is helpful for users in analysing speed spatial and probability distribution. Users can set the colour coding *cl*, that is, the level of the speed level boundaries, based on their preference so that different visualisations can be shown.
(3) *Browsing grid details*. Each grid has its histogram of vessel speed probability distribution in the view. To ensure clearness of view, the view is only rendered by colour but does not show values by default. The speed probability distribution cannot be displayed in the same view as speed spatial distribution to avoid one data set obscuring the other. The details of vessel speed spatial distribution and vessel speed probability distribution can be shown by interactive display.

7. PROTOTYPE SYSTEM.

7.1. *Universal model of interactive analysis.* In order to obtain better interaction with the visual analytics system, in addition to the interactive visualisation model for analysing the characteristics of vessel behaviour, a universal interactive model is constructed in the prototype system. The model includes *Data presetting, Switch of visual analysis view and Universal interactive operations*.

(1) *Data presetting*. Data presetting is to obtain the data sets filtered by the user that need to be analysed. Related operations include filtering the data by time interval, vessel type and geographic location.
(2) *Switch of visual analysis view*. Different views give an analysis of different characteristics of vessel behaviour. A user can switch between multiple views.
(3) *Universal interactive operations*. Basic interactive view operations are supported, including zooming in and out, refreshing, and dragging.

7.2. *Architecture of prototype system.* The prototype of our visual analysis system for vessel behaviour consists of three subsystems: *Visual analysis of vessel trajectory, Visual analysis of vessel density*, and *Visual analysis of vessel speed distribution*. These are built by feature model and interactive visualisation model, which are illustrated in Sections 4, 5 and 6 respectively. The architecture of the visual analysis of vessel trajectory subsystem is

https://doi.org/10.1017/S0373463318000085 Published online by Cambridge University Press

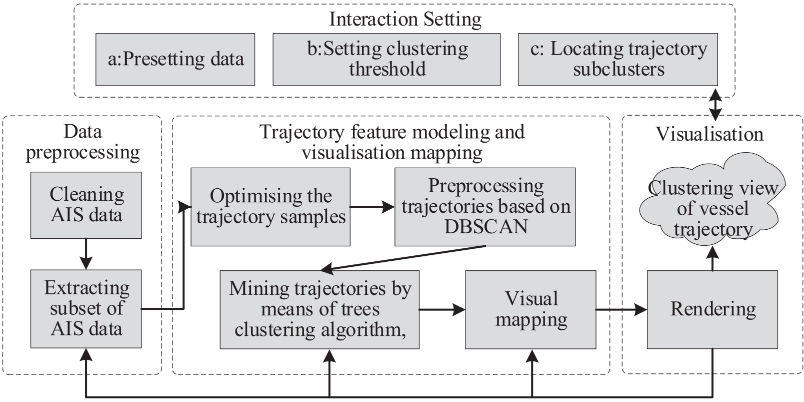


Figure 2. Architecture of *Visual analysis for vessel trajectory*.

shown in Figure 2. This subsystem consists of the following components:

(1) *Data preprocessing*: cleansing data and extracting subset of AIS data interactively;
(2) *Trajectory feature modelling and visualisation mapping*: pre-processing trajectories based on DBSCAN; mining trajectories by means of tree clustering algorithm, and visual mapping;
(3) *Visualisation*: rendering the results of trajectory clustering;
(4) *Interaction setting*: realising interactive analysis, including pre-setting data, setting clustering threshold, and locating trajectory sub-clusters.

8. EXPERIMENT AND ANALYSIS.

8.1. *Experimental environment and data.* A prototype system for visual analysis of vessel behaviour was realised according to Figure 1. For the experiment, we used a machine with the following configuration: a 2.80 GHz Intel® Xeon® CPU with 8 GB main memory and a 1TB hard disk. Operating System was Windows 7 x64. SQL Server 2008 R2 database was used to store AIS data and other processed data. To realise the related algorithms, we used the Java programming language. To implement the visual analysis of the vessel behaviour features we used the open source library Mapv of Baidu Map (http://huiyan.baidu.com/mapv/). The communication between front-end and back-end was ensured by a Javaweb framework and the application server was built using Apache-Tomcat-6.0.36.

As dataset for our experiment, we used historical AIS data collected from middle Hankou channel to Yangluo channel in the Yangtze River in 2015. The data was then cleaned and organised into three datasets D_1, D_2 and D_3 which are shown in Table 2. It should be noted that the AIS data for our experiment was collected in the framework of the project of Natural Science Foundation of China (No. 51479155) in 2015.

8.2. *Experimental results and analysis.* We present three experiment results, as follows.

8.2.1. *Setting cluster thresholds.* The subsystem *Visual analysis of vessel trajectory* requires the user to set two cluster thresholds, σ and ε, which are the key parameters for the vessel trajectory cluster. The definition of σ is given in Equation (2), and ε is the

https://doi.org/10.1017/S0373463318000085 Published online by Cambridge University Press

Table 2. Experimental Data Samples.

No.	Data	Size	AIS Records	Trajectories Included
D_1	2015-01-02	About 10MB	About 100 thousands	503
D_2	2015-01-03~2015-01-07	About 53MB	About 520 thousands	2561
D_3	2015-01-09~2015-01-15	About 67MB	About 660 thousands	3283

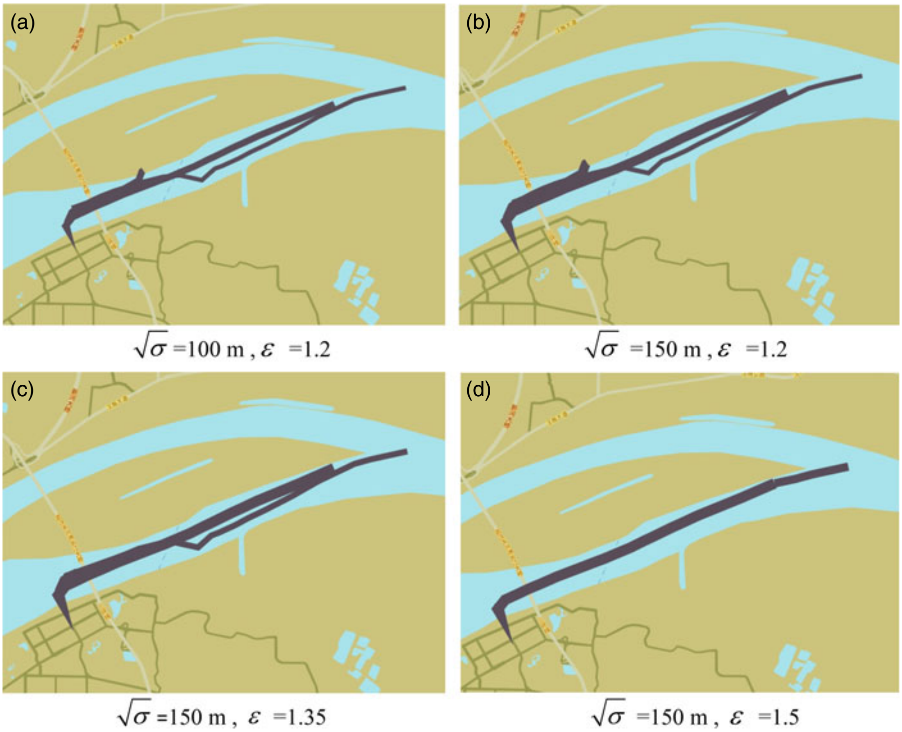


Figure 3. Interactive analysis comparison of setting cluster thresholds.

threshold for $D(T, C)$. The illustration of the effect of different values of these two thresholds on vessel trajectory clusters is depicted in Figure 3: (a) $\sqrt{\sigma} = 100\,\text{m}$, $\varepsilon = 1.2$; (b) $\sqrt{\sigma} = 150\,\text{m}$, $\varepsilon = 1.2$; (c) $\sqrt{\sigma} = 150\,\text{m}$, $\varepsilon = 1.35$; (d) $\sqrt{\sigma} = 150\,\text{m}$, $\varepsilon = 1.5$. It can be seen from Figure 3(a), 3(b) and 3(c) that the number of trajectories absorbed by the trajectory cluster will be more and their supports will be bigger as the cluster threshold $\sqrt{\sigma}$ and ε increase. However, trajectory clusters with different features will be together in a cluster and lose the discrimination between trajectories if the threshold ε is too large, just as the result shown in Figure 3(d). It can also be inferred that σ will become small so that those trajectories which should be in a cluster, will be clustered in different clusters if ε is too small, for example, if it is less than 1. Therefore, by setting cluster thresholds interactively, the user can change and analyse the result of trajectory clustering and choose suitable thresholds.

https://doi.org/10.1017/S0373463318000085 Published online by Cambridge University Press

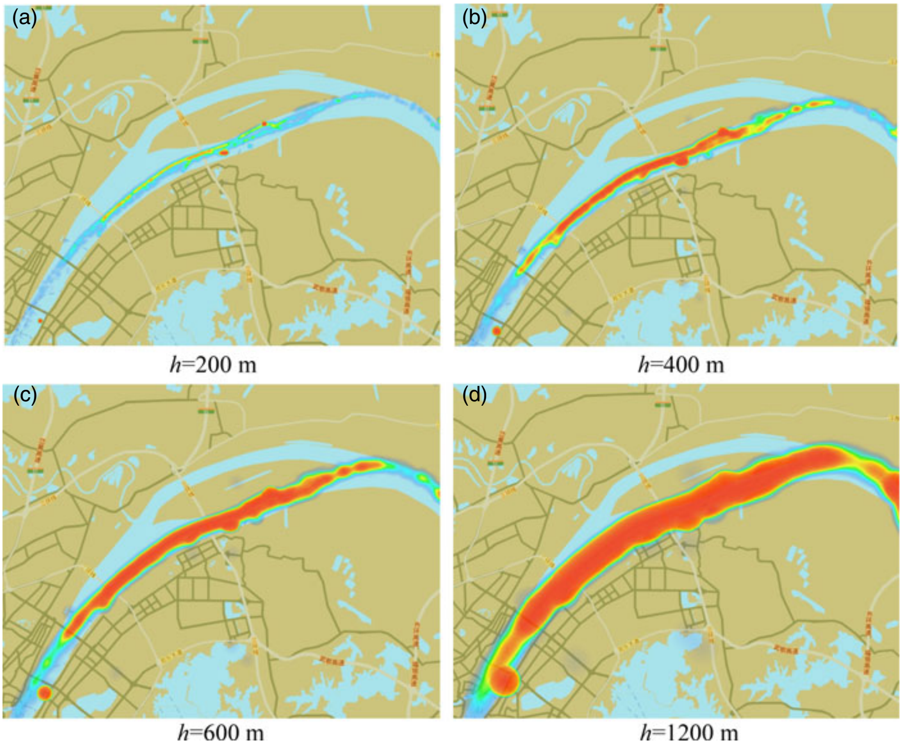


Figure 4. Comparison of attenuation threshold adjusting in a heat map.

8.2.2. *Setting the attenuation threshold of kernel density estimation.* In *Visual analysis of vessel density distribution*, the user can set the attenuation threshold h of kernel density estimation in the heat map.

As shown in Figure 4, the attenuation threshold h in the four figures are respectively 200 m, 400 m, 600 m and 1200 m. Through comparison and analysis, it can be seen that the effect of the heat map is obviously different because of the different attenuation thresholds. The smaller the threshold, the more detailed the density distribution of the heat map. The larger the threshold, the more the heat map can show the trend of density distribution of the whole channel.

8.2.3. *Setting the grid parameters.* In *visual analysis of vessel speed distribution*, users can set the grid parameters themselves. As shown in Figure 5, the meshed parameters $wf \times hf$ in (a) and (b) are 1000×1000 m and 400×400 m. From the two figures, we can obtain a view with finer meshing and more detailed speed distribution when the grid parameters are smaller. The view of vessel speed distribution with large grids provides more detailed information of local area speed distribution, while the view with small grids can give information of the trend of speed distribution in the water area in general.

In summary, through the comparison of the results before and after changing various parameters interactively, the user can obtain the information of vessel behaviour more comprehensively, clearly and intuitively.

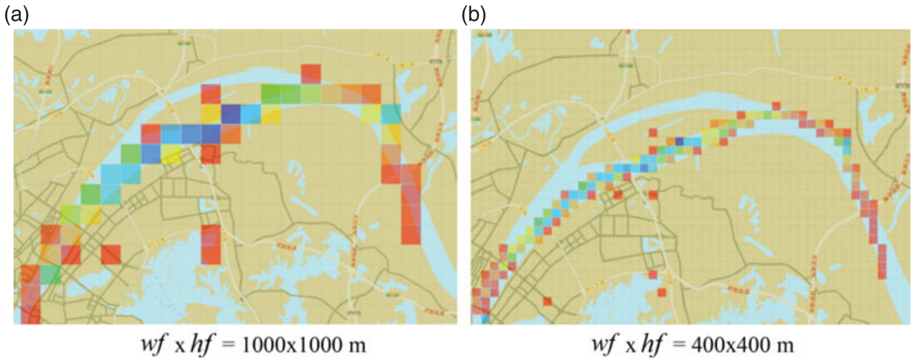


Figure 5. The result of changing the meshed parameter.

9. CONCLUSION AND FUTURE WORK. In this research, we present a novel visual analytics approach to vessel behaviour analysis. It preprocesses AIS data, constructs vessel behaviour models based on vessel trajectory, density and speed, transforms vessel behaviour characteristics data into geometry data and provides the user with multiple views and the ability to interactively control the view to track, detect and analyse vessel behaviour intuitively. Our contributions are to combine data mining with visualisation technology to build interactive models of vessel behaviour analysis and highlight the importance of understanding all the interactions and processes that happen in maritime navigation. The effectiveness of our solution has been demonstrated using real-world data.

There is potential for future research. We will improve the human-computer interactive control model for visual analysis of vessel behaviour and further investigate the parameters in the model, the interaction to set them and possibilities for new visual cues to render these parameters. We will also investigate whether the method can support analysis of vessel behaviour with big data sets in real time.

REFERENCES

Ali, T., Asghar, S. and Sajid, N.A. (2010). Critical analysis of DBSCAN variations. *Proceedings of 2010 International Conference on Information and Emerging Technologies*, 24–29.

Cazzanti, L., Davoli, A. and Millefiori, L.M. (2016). Automated Port Traffic Statistics: From Raw Data to Visualisation. *Proceedings of 2016 IEEE International Conference on Big Data*, 1569–1573

Ferreira, N., Poco, J., Vo, H.T., Freire, J. and Silva, C.T. (2013). Visual Exploration of Big Spatio-Temporal Urban Data: A Study of New York City Taxi Trips. *Transactions on Visualisation and Computer Graphics*, **19**(12), 2149–2158.

Lu, C. and Gao, S. (2016). Visual Analysis of Multi-factor Association on Inland Waterway Accident. *Proceedings of 2016 International Conference on Engineering and Advanced Technology*, 110–115.

Piciarelli, C., Foresti, G.L. and Snidaro, L. (2005). Trajectory clustering and its applications for video surveillance. *Advanced Video and Signal Based Surveillance*. *Proceedings of AVSS 2005. IEEE,* 40–45.

Riveiro, M. and Falkman, G. (2009). Interactive Visualisation of Normal Behavioural Models and Expert Rules for Maritime Anomaly Detection. *Proceedings of Sixth International Conference on Computer Graphics, Imaging and Visualisation*, 459–466.

Scheepens, R., Hurter, C., Van de Wetering, H. and Van Wijk, J.J. (2016). Visualisation, selection, and analysis of traffic flows. *IEEE Transactions on Visualisation and Computer Graphics,* **22**(1), 379–388.

Sidibé, A. and Gao, S. (2017). Study of Automatic Anomalous Behaviour Detection Techniques for Maritime Vessels. *Journal of Navigation*, **70**(4), 847–858.

https://doi.org/10.1017/S0373463318000085 Published online by Cambridge University Press

Thomas, J.J. and Cook, K.A. (2006). A visual analytics agenda. *Computer Graphics and Applications, IEEE*, **26**(1), 10–13.

Tominski, C., Schumann, H., Andrienko, G. and Andrienko, N. (2012). Stacking-Based Visualisation of Trajectory Attribute Data. *IEEE Transactions on Visualisation & Computer Graphics*, **18**(12), 2565–2574.

Wang, G.Z., Malik, A., Yau, C., Surakitbanharn, C. and Ebert, D.S. (2017). TraSeer: A Visual Analytics Tool for Vessel Movements in the Coastal Areas. *Proceedings of IEEE International Symposium on Technologies for Homeland Security*, 1–6

Wang, Z., Lu, M., Yuan, X.R., Zhang, J.P and Wetering, H.V.D. (2013). Visual Traffic Jam Analysis Based on Trajectory Data. *Transactions on Visualisation and Computer Graphics*, **19**(12), 2159–2168.

Willems, N. (2011). *Visualisation of vessel traffic*. PhD thesis, Eindhoven University of Technology. http://alexandria.tue.nl/extra2/719764.pdf.

https://doi.org/10.1017/S0373463318000085 Published online by Cambridge University Press