

## Vessel Crowd Movement Pattern Mining for Maritime Traffic Management

Rong Wen<sup>1\*</sup> and Wenjing Yan<sup>1</sup>

<sup>1</sup>*Planning and Operations Group, Singapore Institute of Manufacturing Technology, Singapore, Singapore; Email: wenr@simtech.a-star.edu.sg*

**\*Corresponding Author:** Rong Wen

**Abstract:** The goal of maritime traffic management is to provide a safe and efficient maritime environment for different type of vessels facilitating port logistics and supply chain business. However, current maritime traffic management mainly relies on the massive individual vessel's data for decision making. Lack of macro-level understanding of vessel crowd movement around port challenges maritime safety and traffic efficiency. In this paper, we describe a spatio-temporal data mining method to discover crowd movement patterns of vessels from their short-term history data. The method first captures vessels' crowd movement features by building vessels' tracklets with their speed and location. A movement vector clustering algorithm is developed to find different travel behaviors for different group of vessels. With nonparametric regression on the classified vessel movement vectors which represent the crowd travel behaviors, an overall vessel movement pattern can then be discovered. In this research, we tested real trajectory data of vessels near Singapore ports. Comparing with the actual massive vessel movement data, we found that this method was able to extract vessels' crowd movement information. The hotspots on risk area in terms of vessel traffic and speed can be identified. The method can be used to provide decision-making support for maritime traffic management.

**Keywords:** Spatio-temporal data mining, travel behavior mining, vector clustering, nonparametric regression, maritime traffic planning and management

### 1. Introduction

In maritime industry, maritime traffic management is a key to prevent development of dangerous maritime traffic situations, and improve efficiency of maritime logistics and transport operations. Maritime traffic management, usually conducted by staff of vessel traffic service (VTS), has been challenged by massive vessel data monitoring, smooth coordination between ports and the vessels sailing around ports, lack of manpower support and intelligent data analytics. With development and application of various vessel tracking and positioning sensors in the maritime industry, e.g.,

Automatic Identification System (AIS) and Vessel Monitoring Systems (VMS), abundant spatial and temporal data sets related to vessel voyage are available for vessel motion data analysis. Recent development of machine learning methods provided intelligent data-driven solutions to logistics and transportation problems, as well as traffic management for land and sea transportation [1,2]. Frequent route mining with AIS data has been widely studied [3-5] to find route patterns of vessels between ports and harbours. Trajectory data partitioning and re-grouping methods [6,7] could be used to explore massive trajectory segment of vessels. The final route patterns can be generated by density-based route segment clusters. Spatial outlier detection has been used to detect and remove the noise data for vessel route data processing [8,9]. Spatio-temporal sequential pattern mining was used to explore vessel's travel behaviors and sequence patterns of vessel movement between different location [10]. To find insightful knowledge of vessels' movement pattern, most current methods focus on mining individual or certain type of vessels' trajectory data from which trajectory segments or similar vectors are assembled to a general trajectory. However, to rapidly identify a risk area and improve efficiency of maritime traffic flow, maritime traffic management highly requires a macro-level understanding of the recent movement patterns of different types of vessels, which is crowd movement pattern.

In this research, a comprehensive spatio-temporal data mining method is proposed and developed to discover crowd movement patterns of vessels from their short-term history data. To establish a vessel crowd motion pattern for decision making support for maritime traffic management, the method applies a three-stage pipeline: (1) Vessel movement pattern detection and extract crowd movement features, (2) movement vector clustering, and (3) crowd trajectory pattern is generated by nonparametric spatial regression. By concatenating crowd movement pattern across different vessel types and different time periods, an overall vessel movement pattern can then be discovered.

## **2. Crowd Movement Feature Mining**

This section describes data processing pipeline for vessel movement feature extraction and normalization. Spatio-temporal units, nodes, at location (latitude, longitude) and timestamp are created for construction of vector and tracklets which represent the crowd movement features. The crowd movement can be then described by tracklets which contain crowd location and speed information.

### **2.1 Spatio-temporal Tracklet Extraction**

In order to find movement patterns from the massive spatio-temporal data sets, topological graph methods [11,12] are employed. Topological graph theories describe spatial embedding of graphs,

which have been commonly used in real-world applications including social network data analysis, transportation planning and text mining with graphs of word recurrences. Graph embedding is used in this method to reduce complexity of massive spatio-temporal data analysis. To describe vessel movement pattern, a stochastic proximity embedding (SPE) [11] is used to convert massive vessel data to motion vectors represented by tracklets.

Assume nodes,  $N = \{n_1, n_2, n_3, \dots, n_n\}$ , are the massive historical vessel data at locations  $(lat_i, long_i)$  with all the time stamp  $T = \{t_1, t_2, t_3, \dots, t_n\}$ . SPE performs multiple rounds of calculation to find node  $n_i$  and  $n_j$  within proximity at time  $t_i$ . An average moving distance between the nodes,  $d$ , is used to sort the massive historical vessel locations in pairs by filtering and adjusting their distance. In our method, an improved SPE [12] is employed to generate the node pairs and generate the tracklets as follows. Firstly, nodes  $[n_i(t_i), n_j(t_{i+1})]$  are randomly selected from the node space. The average moving distance  $d_s$  can be derived from vessels crowd moving speed. Secondly, the nodes  $[n_i(t_i), n_j(t_{i+1})]$  are iteratively adjusted by following rules,

$$\text{If } D_i [n_i(t_i), n_j(t_{i+1})] < d_s, n_j(t_{i+1}) = n_j(t_{i+1}) + k_s \quad (1)$$

$$\text{If } D_i [n_i(t_i), n_j(t_{i+1})] > d_s, n_j(t_{i+1}) = n_j(t_{i+1}) - k_s \quad (2)$$

to meet the condition,

$$D_i [n_i(t_i), n_j(t_{i+1})] - d_s * h \geq 0 \quad (3)$$

where:  $D_i$  is the function calculating the distance between the node  $n_i(t_i)$  and  $n_j(t_{i+1})$ .

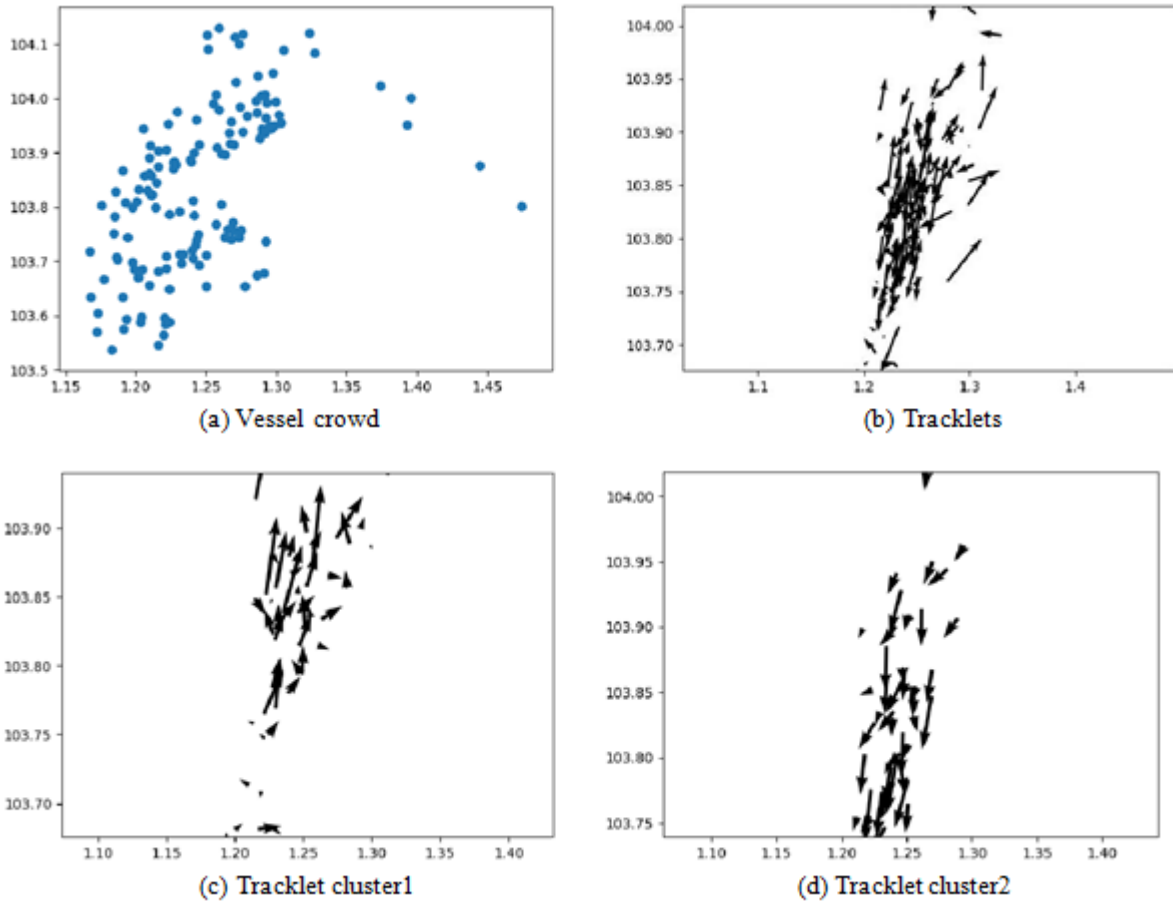
The expected movement distance,  $d_s$ , is acquired from the historical average speed of different vessel types. The increment/decrement,  $k_s$ , could be stochastically optimized, in order to efficiently finding the clusters for tracklet calculation. The hopping parameter,  $h$ ,  $h=2, 3$  and  $5$ , are used in this study for different vessel crowd. Crowd of different type of vessels normally may have different hop number in different time periods. Proper selection of hopping parameter plays a key role in generating accurate tracklets cluster. Based on the node location, a tracklet is defined as,

$$\mathbf{Tr}(i) = n_i(t_i) + kv \quad (4)$$

to describe the movement from node  $n_i$  to  $n_j$  in time sequence. In this way, crowd's tracklets can be calculated with the massive node location data  $[n_i(t_i), n_j(t_{i+1})]$ . At this stage, the tracklets extracted from vessels' spatio-temporal data sets indicate local movement features. Fig. 1 shows an example of local movement features, represented by tracklets, are extracted from the vessel crowd data, represent by,  $(lat_i, long_i)$ , spanning a certain time period.

## 2.2 Crowd Movement Feature Generation

In order to discover the macro-level movement of the vessel crowd, spatial and temporal clustering of the tracklets are conducted. Crowd movement feature generation aims to mine cohesive motion pattern from the local features (tracklets).



**Fig. 1** Vessel crowd movement feature extraction. Source: authors

In this method, we explore spatial and temporal clustering to generate the crowd movement features. The tracklets generated in Section 2.1 provide elemental movement features of vessels. With location and movement vector of the tracklets, spatial feature of the crowd can be discovered based on density and distance of the tracklets. The temporal clustering groups the spatially clustered tracklets into time sequence based clusters and extract their movement features in terms of location and moving direction.

Spatial clustering is performed as describe by the Equation (5) and (6),

$$P_2 = (1/Nxh) \sum_n K(u), \text{ with } u = \frac{Tr(i)-Tr(j)}{h}, \text{ and } K(u) = \frac{1}{\sqrt{2\tau}} \exp\left(-\frac{1}{2}u^2\right) \quad (5)$$

$$\text{If } P_2 > H, Tr(S) = \sum_m Tr(1), Tr(2), Tr(3), \dots, Tr(m))/m, Tr(i) \in S_m \quad (6)$$

where  $P_2$  is tracklets' kernel density calculation in the region  $S$ , comparing with a threshold value  $H$ . The tracklets,  $\mathbf{Tr}(i)$  belongs to the region  $S$ . The crowd spatial feature is produced by generating representative featured tracklets. In traditional density estimations, the kernel functions are used without considering directional and time information. With a tilted trigonometric bell [13], the density of dynamic points can be acquired by incorporating information of time and position changes of points. In addition, the kernel's bandwidth is determined by the points at time  $t_i$  and  $t_{i+1}$ . To perform the density calculation in tracklet space, normalization of the tracklet in distance is required. In the region  $S$ ,  $\mathbf{Tr}(i)$  can be normalized by parameter,  $b = \sum_n \text{dist}[(\mathbf{Tr}(i), \mathbf{Tr}(i))]/n$ . Based on the density value, a featured tracklet can be then generated by its neighboring similar tracklets. In this way, the location of center of the  $\mathbf{Tr}(S)$  can then be acquired by location data of  $\mathbf{Tr}(i)$  in  $S$ . With featured tracklets located in different regions, the crowd spatial features are discovered.

Temporal clustering aims to combine local spatial features and group them into clusters of movement patterns in time sequence. Based on the crowd spatial features extracted above, featured tracklets,  $\mathbf{Tf}(i)$ , can be observed in time,  $t_i$ . In this research, we focus on the strong temporal relation between the featured tracklets,  $\mathbf{Tf}(T_1)$  and  $\mathbf{Tf}(T_2)$ , where the mean of  $t$  in tracklets of  $\mathbf{Tf}(T_1)$  precedes  $\mathbf{Tf}(T_2)$ . The relation of  $\mathbf{Tf}(T_1)$  and  $\mathbf{Tf}(T_2)$  can be described by with time indexes of the tracklets  $\mathbf{Tr}(i)$ . Assume  $\mathbf{Tf}(S) = \{ \mathbf{Tr}(t_1), \mathbf{Tr}(t_2), \dots, \mathbf{Tr}(n) \mid t_i \in (T_1, T_2) \}$ . For the cluster,  $\mathbf{Tf}(T_1) = \{ \mathbf{Tr}(t_1), \mathbf{Tr}(t_2), \dots, \mathbf{Tr}(t_j) \mid t_i \in T_1 \}$ ,  $\mathbf{Tf}(T_2) = \{ \mathbf{Tr}(t_1), \mathbf{Tr}(t_2), \dots, \mathbf{Tr}(t_k) \mid t_i \in T_2 \}$ ,

If  $T_2 > T_1$ ,  $\mathbf{Tf}(T_1)$  precedes  $\mathbf{Tf}(T_2)$ . To build cluster of  $\mathbf{Tf}(T_1)$  and  $\mathbf{Tf}(T_2)$ , time stamp distribution of cluster of tracklets of the spatial feature should be considered. An elbow view of the time stamp distribution could be used to decide threshold value for the temporal clusters and number of clusters. An alternative way to decide time boundary for the temporal cluster is to calculate the strength of the temporal relation,  $\mathbf{Tf}(T_1)$  precedes  $\mathbf{Tf}(T_2)$ .

### 3. Vessel Crowd Movement Pattern Mining

In the maritime traffic management, vessel crowd's travel behavior include vessels' past and current trajectories, sailing speed and directions are the key information for VTS officers making decisions. With spatio-temporal features generated for the vessel crowd, the crowd's movement pattern and travel behavior could be explored next. In this section, vessel crowd is further regrouped, based on their movement features, for vessel crowd movement pattern mining as well as their sailing behavior analysis.

### 3.1 Movement Feature Mining for Sub Crowd

Vessels' sailing direction is considered as one of the most key movement feature for the crowd travel behavior analysis. Since tracklets generated in the Section 3 contain vessels' local movement vectors, we develop a vector clustering method to group vessels into different sub crowd.

Firstly, vector data preprocessing is required. From the tracklet vectors between time  $t_1$  and  $t_2$ , the tracklet vector,  $\mathbf{Tr}(i) = \{Tr(kv_i)|(v_{xi}, v_{yi})\}$ , is examined by anomaly isolation to identify outliers of sailing directions in the crowd. We select only the major direction that have high impact on the crowd movement trend and behaviors. To extract the major movement direction, we exploit tracklet vectors,  $(v_{xi}, v_{yi})$ , and find major movement vectors of the sub crowd. However, current classification-based methods [7] normally has high computational cost in mining the tracklets' direction. An iForest isolation method is developed to efficiently find major directions from the high volume vector data. The iForest is built on the basis of iTrees [14] which construct an ensemble of binary search trees. Given a vector data set, average path length of search cost is described by,

$$c(n) = 2\mathbf{H}(n - 1) - (2(n - 1)/n), \quad (7)$$

where:  $n$  is the number of instance of the tracklet vectors.

$\mathbf{H}(i)$  can be estimated by  $\ln(i) + 0.5772156649$  (Euler's constant). The anomaly score  $s$  for an anomaly tracklet vector can be calculated by,

$$s(\mathbf{x}, n) = 2 \frac{W(h(\mathbf{x}))}{c(n)}, \quad (8)$$

where:  $E(h(\mathbf{x}))$  is the average of  $h(\mathbf{x})$  from a collection of isolation trees.

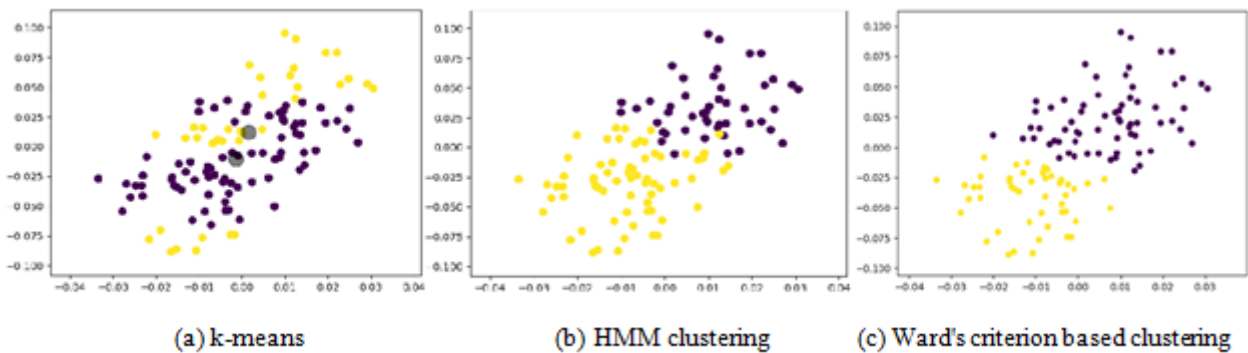
Given a tracklet vector, the anomalies are considered to have short average path to be reached. In this way, no distance or density measurement is required in searching the anomalies, so that computational cost could be significantly reduced. In addition, utilization of the iForest enables span of searching process to quickly converge with a small number of trees, which means it could process large volume of vessel data sets in real time and provide online decision-making support. Apart from the generation of major direction, normalization of the spatial features of the tracklets,  $\mathbf{Tf}(i)$ , is required in the stage of movement vector processing. The Fig. 1 (c) and (d) show the major direction of the vector extracted by iForest.

### 3.2 Vector clustering

Vector clustering aims to find major direction of the sub crowd. Due to less physical traffic constraints on vessels, e.g., lanes, curbs and traffic signs, vessels normally have high travel flexibility comparing with vehicles. However, higher flexibility means more complexity for

movement behavior analysis. To describe vessel crowd's movement pattern, we develop a vector clustering method to automatically group the movement features of the sub crowds.

To properly cluster the tracklets' vectors, Ward's minimum variance criterion [15] is applied to the two-dimensional vector data sets,  $\mathbf{Tr}(i) = \{\mathbf{Tr}(k_{vi})|(v_{xi}, v_{yi})\}$ . A random pair of vectors are selected to calculate vectors' distance at each iterative step, so that the pairs that lead to minimum distance increment between clusters can be found. The distance increment could be considered as weighted squared distance or Euclidean distance. In order to properly determine the number of the clusters to be generated, user-defined parameters or elbow approach can be used. The tracklets with similar moving direction are grouped representing the local movement pattern in a certain time period. The Fig. 2 shows a comparison of performance of clustering of vessels'  $\mathbf{Tr}(i)$  vectors with k-means, Hidden Markov Model (HMM) and Ward's criterion based clustering. Normally,  $\mathbf{Tr}(i)$  vector clustering is challenging due to irregularity and flexibility of vessels' trajectories. As shown in Fig.2, we can see that Ward's criterion based clustering can produce a more satisfying result.



**Fig. 2** Temporal clustering to find crowd movement sequences. Source: authors

### 3.3 Crowd Route Pattern Mining

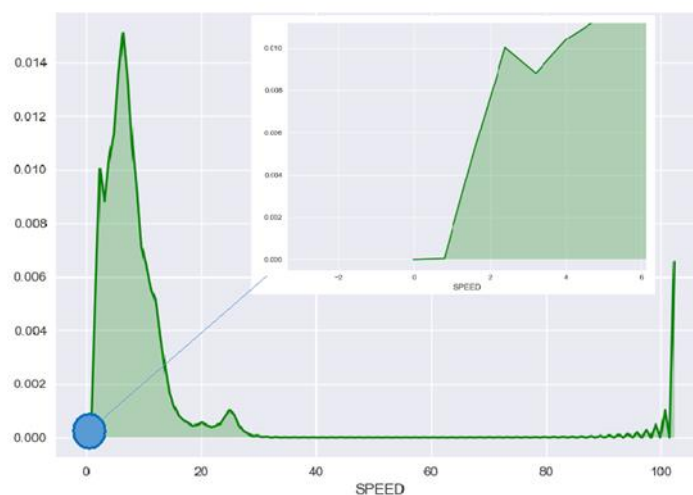
To find the crowd route pattern, nonparametric regression of the tracklets is used to generate macro-level crowd's trajectories. In this method, an algorithm for shipping route extraction [1] that we developed for frequent route mining is employed. We applied locally weighted smoothing to tracklets' start and end points. A squared distance based measurement is used to evaluate deviation of the neighbouring points with respect to each local route segment. In this way, an iterative local polynomial fitting process can be then implemented to refine the local regression weight to generate proper route points. To overcome the poor performance of the route generation for none one-to-one correspondence, we use affine coordinate transformation to generate a proper route in an inversed vessels' location data space.

Affine coordinate transformation is a special case of homography transformation by which variable correspondences between longitude and latitude can be changed by scaling, rotating, reflection and shear mapping. The affine transformation enables variable correspondence to be constructed in terms of coordinates (longitude, latitude) to which Locally weighted scatterplot smoothing (LOWESS) can then be applied. After crowd route is acquired in the transformed coordinate system, an inverse mapping is implemented to map the routes generated in the transformed coordinate system back to the original system.

#### 4. Experimental Results and Discussion

Malacca-Singapore Strait is one of the busiest water way in the world, connecting maritime transportation from Asia to Middle East and Europe. The vessel traffic in the Strait keep growing with number of vessels consistently increasing. In the experiment, we analyzed two-week vessels' AIS data recording movement of different types of vessels, including cargos, tankers, tugs and passenger ships, near Singapore.

The first step of movement spatial feature mining is vessels speed data processing. Due to regular wave motion, vessel speed value record by AIS normally cannot reflect vessels' actual movement status, especially when vessels anchor on the sea. As shown in the Fig. 3, most of vessel sailing at a low speed near Singapore. The vessel speed above 0.76 knot was considered as in active status which may contribute the pattern generation.

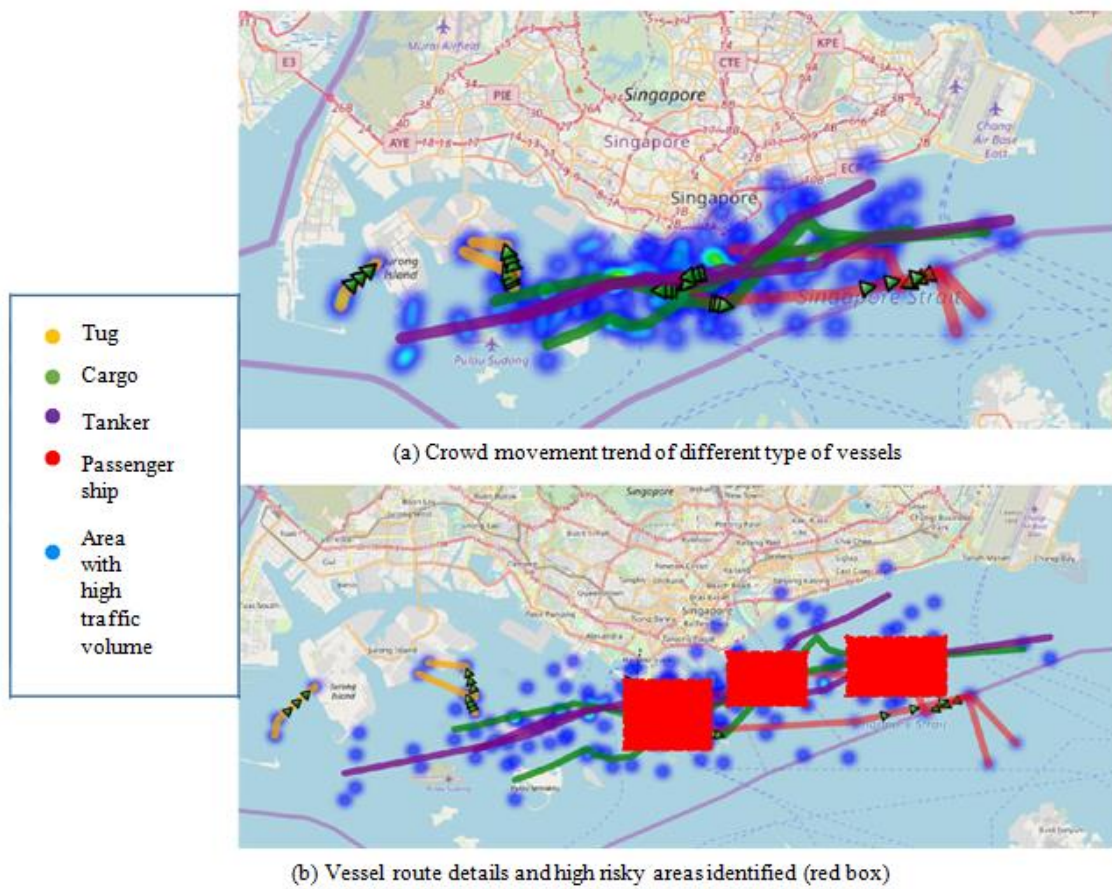


**Fig. 3** Vessel speed distribution. Source: authors

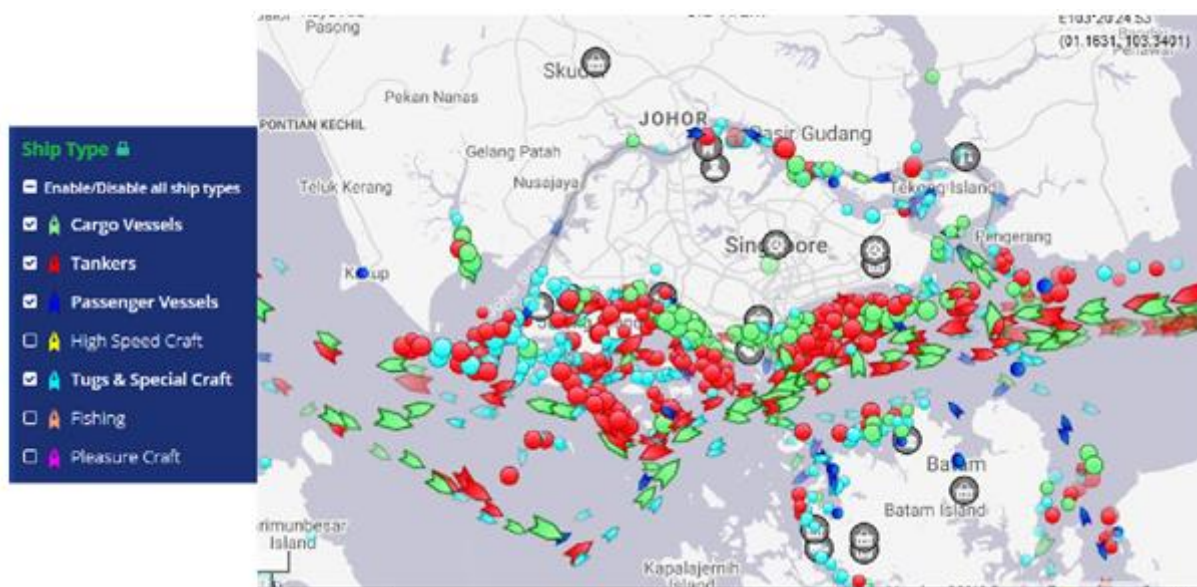
Fig.4 shows crowd movement trend of vessels of different types in a two-hour time period, from 8 am to 10 am, on October of 2015. The vessel types, cargo, tanker, tugs and passenger ship, were selected, because these types of vessels cover most number of vessels around the Singapore Strait. Fig. 5 is a screenshot captured from MarineTraffic (<https://www.marinetraffic.com/>) with the same vessel types. As shown in Fig.4 (a), vessel crowd trends represented by different routes in



different colors were generated, indicating a continuous crowd movement in the past two hours. We can see that tankers and cargo vessels have similar motion patterns during the two-hour journey. The Fig.4(b) shows the risky areas can be identified based on vessel crowds' movement trend.



**Fig. 4** Vessel crowd movement trend. Source: authors



**Fig. 5** Vessels' movement captured from MarineTraffic. Source: authors

## 5. Conclusion

Due to vessel traveling with more freedom comparing with the land transportation, understanding of vessels crowd movement in a given time period is required by vessel traffic offices, especially for the traffic management of busy waterway.

In this research, we developed a method that is able to explore vessels' crowd movement patterns to facilitate the maritime traffic management and risk prevention. With vessels' historical speed and location data, tracklets can be constructed as units to represent spatio-temporal features of vessel crowd movement.

By spatial and temporal mining of tracklets' location and motion vectors, vessel crowd spatio-temporal movement features can be derived, revealing the crowd's continuous travel behavior in a given past time period. In addition, the crowds' trajectories generated from different types of vessels can be used to identify movement trend of maritime logistics and transportation, as well as the risky areas where potential vessel collision may occur.

## Acknowledgments

This work is partially supported under the Project, Massive Spatio-temporal Data Analytics to Support Logistics Planning (C17-O-013) and Ensemble Prediction for Double Seasonal Time Series in Supply Chain (U18-O-025SU).

## References

- [1] Kleiner, F.S., Mamiya, C.J. & Tansey, R.G. (2001). Gardner's art through the ages (11th ed.). Fort Worth, USA: Harcourt College Publishers.
- [2] Wen, R., Yan, W., Zhang, A. N., Chinh, N. Q. & Akcan, O. (2017). Spatio-temporal route mining and visualization for busy waterways. Paper presented at the 2016 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2016 - Conference Proceedings, 849-854. DOI:10.1109/SMC.2016.7844346
- [3] Wen, R., Yan, W. & Zhang, A. N. (2017). Adaptive spatio-temporal mining for route planning and travel time estimation. Paper presented at the Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017, 2018-January 3278-3284. DOI:10.1109/BigData.2017.8258311
- [4] Tu, E., Zhang, G., Rachmawati, L., Rajabally, E. & Huang, G. (2018). Exploiting AIS data for intelligent maritime navigation: A comprehensive survey from data to methodology. IEEE Transactions on Intelligent Transportation Systems, 19(5), 1559-1582. DOI:10.1109/TITS.2017.2724551

- [5] Yan, W., Wen, R., Zhang, A. N. & Yang, D. (2016). Vessel movement analysis and pattern discovery using density-based clustering approach. Paper presented at the Proceedings - 2016 IEEE International Conference on Big Data, Big Data 2016, 3798-3806. DOI:10.1109/BigData.2016.7841051
- [6] Wen, R., Yan, W. & Zhang, A. N. (2016). Weighted clustering of spatial pattern for optimal logistics hub deployment. Paper presented at the Proceedings - 2016 IEEE International Conference on Big Data, Big Data 2016, 3792-3797. DOI:10.1109/BigData.2016.7841050
- [7] Song, J., Wen, R. & Yan, W. (2016) Identification of traffic accident clusters using kulldorffs space-time scan statistics. Paper presented at the Proceedings - 2016 IEEE International Conference on Big Data, Big Data 2016, 3792-3797. DOI:10.1109/BigData.2016.7841050
- [8] Lee J. G., Han J. & Whang, K.-Y. (2007). Trajectory clustering: A partitionand-group framework. Proceedings of the ACM SIGMOD International Conference on Management of Data.
- [9] Wisdom M. J., Cimon N., Johnson B., Garton E. & Thomas J. (2004). Spatial partitioning by mule deer and elk in relation to traffic. Transactions, North American Wildlife and Natural Resource Conference, 69(01).
- [10] Zhen R., Jin Y., Hu Q., Shao Z. & Nikitakos N. (2017). Maritime anomaly detection within coastal waters based on vessel trajectory clustering and nave bayes classifier. Journal of Navigation, 70(3), 648-670.
- [11] Bermingham, L. & Lee, I. (2014). Spatio-temporal sequential pattern mining for tourism sciences. Paper presented at the Procedia Computer Science, vol. 29 379-389. DOI:10.1016/j.procs.2014.05.034
- [12] Agrafiotis, D. K. (2003). Stochastic proximity embedding. Journal of Computational Chemistry, 24(10), 1215-1221. DOI:10.1002/jcc.10234
- [13] Heldens, S., Litvak, N., & Van Steen, M. (2018). Scalable detection of crowd motion patterns. IEEE Transactions on Knowledge and Data Engineering, DOI:10.1109/TKDE.2018.2879079
- [14] Peters S. & Krisp J. M. (2010). Density calculation for moving points. in GIScience 2010.
- [15] Liu, F. T., Ting, K. M. & Zhou, Z. (2008). Isolation forest. Paper presented at the Proceedings - IEEE International Conference on Data Mining, ICDM, 413-422. DOI:10.1109/ICDM.2008.17
- [16] Kang, S. & Chien, W. K. (2016). A method to group reliability data by hierarchical clustering. Paper presented at the IEEE International Conference on Industrial Engineering and Engineering Management, 2016-December 345-349. DOI:10.1109/IEEM.2016.7797894